BAO in the projected cross-correlation function

Application to eBOSS quasars and photometric ELGs from the DESI Legacy Imaging Surveys

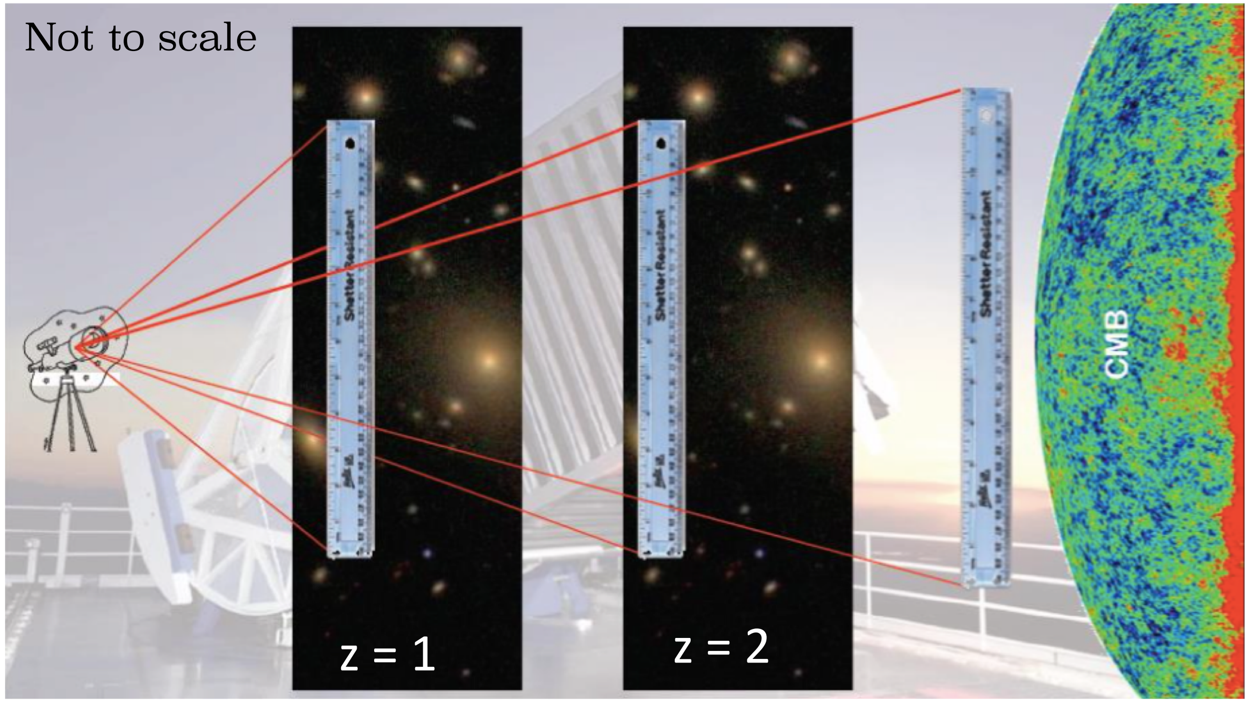
Durham University
ICC

Pauline Zarrouk

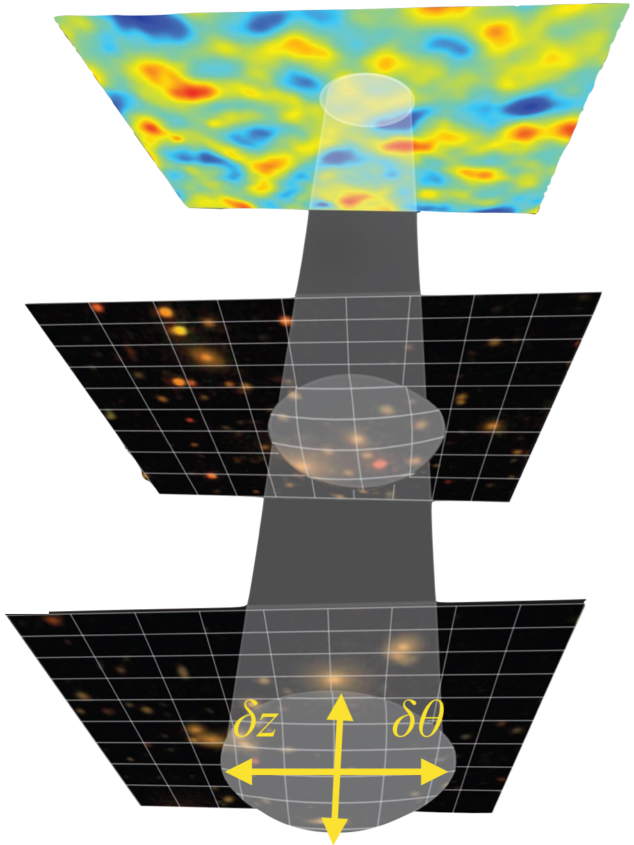
With: Mehdi Rezaie, Anand Raichoor, Ashley Ross, Shaun Cole, Daniel Eisenstein, Peder Norberg, Hee-Jong Seo

Ateliers Dark Energy– May 21th 2020

Baryon Acoustic Oscillations –



BAO: imprint on the galaxy clustering left by the baryon-photon plasma in the early universe which propagated as sound waves until decoupling
→ **Characteristic scale**: two galaxies are preferentially separated by 100 Mpc/h in comoving units



Using the Alcock-Paczynski test and the BAO as standard ruler

Anisotropic BAO parameters

$$\alpha_{\parallel} = \frac{H^{\rm fid}(z) r_{\rm drag}^{\rm fid}}{H(z) r_{\rm drag}}, \qquad \alpha_{\perp} = \frac{D_{\rm M}(z) r_{\rm drag}^{\rm fid}}{D_{\rm M}^{\rm fid}(z) r_{\rm drag}}$$

Isotropic BAO parameter with D_V spherically-averaged distance

$$\alpha_{\rm iso} = \frac{D_{\rm V}(z) r_{\rm drag}^{\rm fid}}{D_{\rm V}^{\rm fid}(z) r_{\rm drag}} \qquad D_{\rm V} = \left[(1+z)^2 cz \frac{D_{\rm M}^2}{H} \right]^{\frac{1}{3}}$$

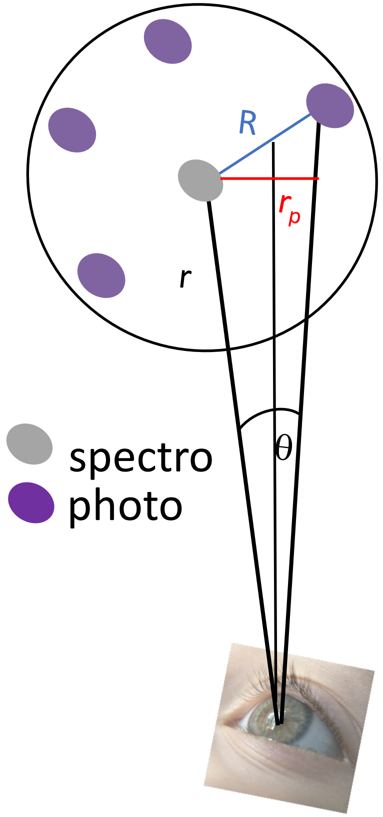
Projected cross-correlation function
Sparse spectroscopic and dense photometric samples

Calculate pair counts in **2D annular bins around spectroscopic objects**, assuming $z_{photo}=z_{spectro}$ in the $z_{spectro}$ bin

→ Angular correlation function binned in physical transverse separation R, defined by:

$$\mathbf{R} = D_{\mathrm{M}}(z_{\mathrm{s}}) \arccos(\boldsymbol{\gamma}_{\mathrm{s}} \cdot \boldsymbol{\gamma}_{\mathrm{p}}) \simeq D_{\mathrm{M}}(z_{\mathrm{s}}) |\boldsymbol{\theta}_{\mathrm{s}} - \boldsymbol{\theta}_{\mathrm{p}}|$$

where D_M is the comoving angular diameter distance



The projected cross-correlation function is thus defined as:

$$w_{\theta}(\mathbf{R}) = \langle \delta_{\mathrm{s}}(\mathbf{r}) \Delta_{\mathrm{p}}(\boldsymbol{\theta} + \mathbf{R}/r) \rangle$$

3D density field of the spectroscopic sample

projected density field of the photometric sample

→ w(R) preserves the physical scales inherent in LSS such as the BAO scale while w(θ) mixes different scales and thus smears the BAO peak

Padmanabhan et al. (2009)
Nishizawa, Oguri and Takada (2013)
Patej & Eisenstein (2017)

Projected cross-correlation function

Analytical prediction

Padmanabhan et al. (2009)

$$w_\theta(\mathbf{R}) = \langle f(b_s, n_s, b_p, n_p) \rangle w_\mathrm{p}(\mathbf{R})$$

Angular correlation function binned into transverse separation
→ Projected correlation function

Standard projected correlation function

where $$w_\mathrm{p}(\mathbf{R}) = \frac{1}{2\pi} \int d^2 k_\perp P(k_\perp) e^{i\mathbf{k}_\perp \cdot \mathbf{R}}$$

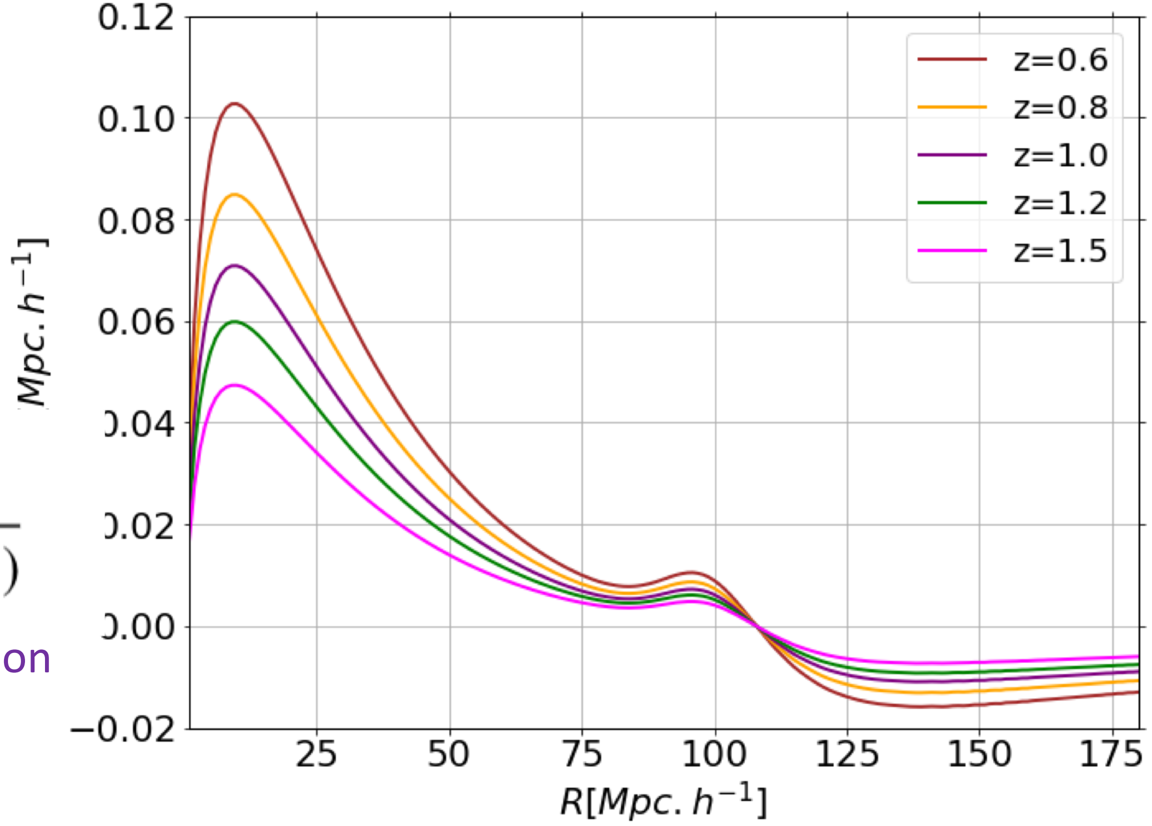
Patej & Eisenstein (2017)

$$\langle f(n_\mathrm{s}, n_\mathrm{p}) \rangle = \frac{b_\mathrm{s} b_\mathrm{p}}{2\pi} \frac{\int dr\, r^2 n_\mathrm{s}^2(r) W(r,\eta) n_\mathrm{p}(r)}{\int dr\, n_\mathrm{s}(r) W(r,\eta) \int dr'\, r'^2 n_\mathrm{p}(r')}$$

Weighting function → Angular and radial selection

Assumptions:

- Weighting scheme independent of r and eta
- n(z) of the photometric sample unknown a priori
- Redshift evolution



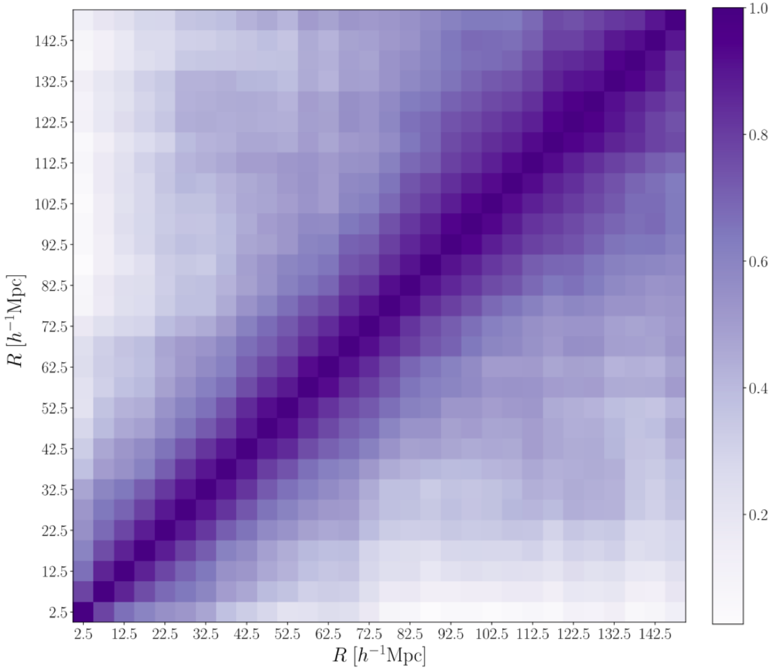
Redshift evolution of the projected cross-correlation function, bin width = 1 Mpc/h

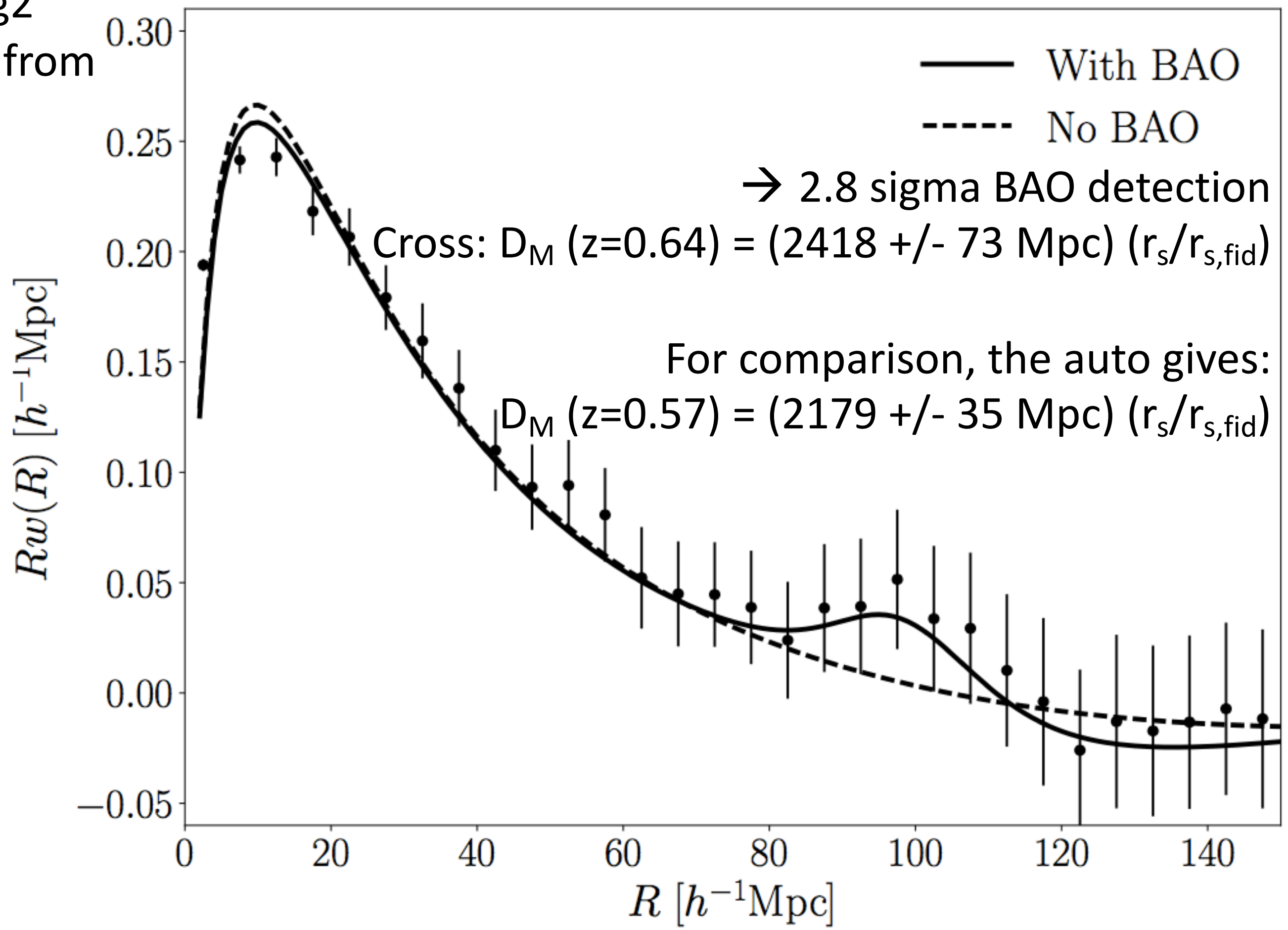
BAO in the projected cross-correlation function

Proof of concept: Patej & Eisenstein (2017)

Fitting conditions

- Spectroscopic sample: z > 0.6 BOSS CMASS, 30/deg2
- Photometric sample using SDSS DR9 and selection from Law-Smith & Eisenstein (2017), 1100/deg2
- Fitting range: 10-150 Mpc/h
- Covariance matrix from 150 jackknife realisations





→ 2.8 sigma BAO detection

Cross: D_M (z=0.64) = (2418 +/- 73 Mpc) ($r_s/r_{s,fid}$)

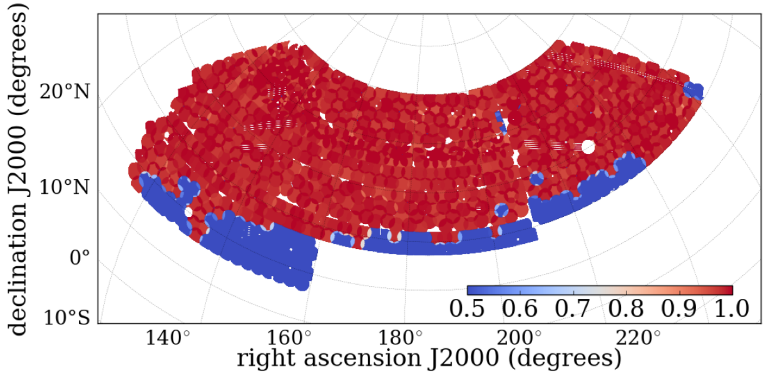
For comparison, the auto gives:

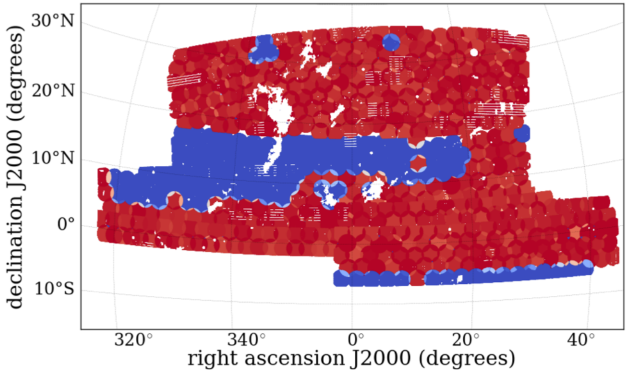
D_M (z=0.57) = (2179 +/- 35 Mpc) ($r_s/r_{s,fid}$)

Datasets: Spectroscopic sample
eBOSS DR16 quasars

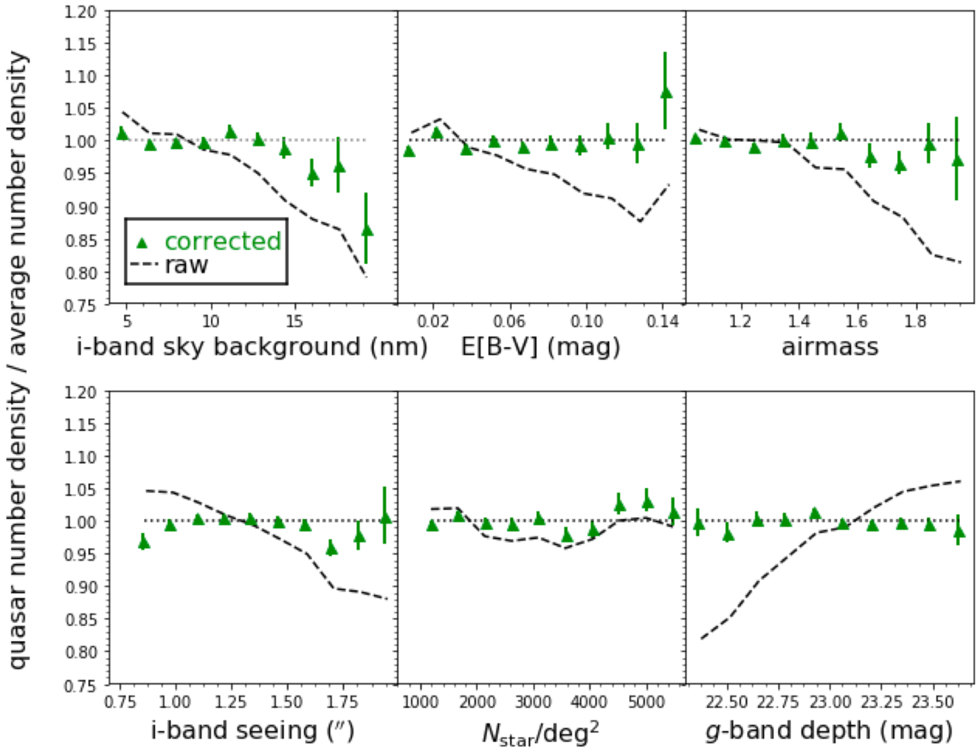
Official eBOSS DR16 LSSquasar catalogue (Ross et al. 2020)

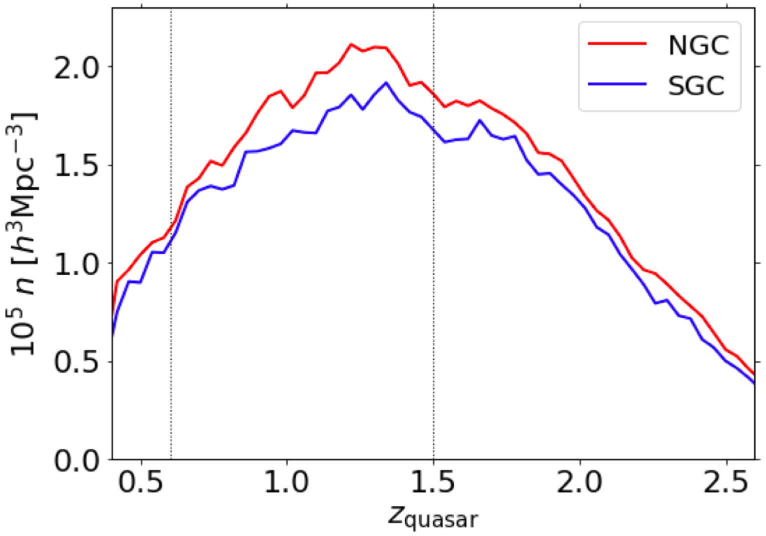
343708 quasars over 4699 deg^2 in 0.8 <= z <= 2.2





Variation of the quasar target density with photometric systematics. Correction based on 1) g-band depth, 2) Galactic extinction, 3) i-band sky background (not applied in the plot)

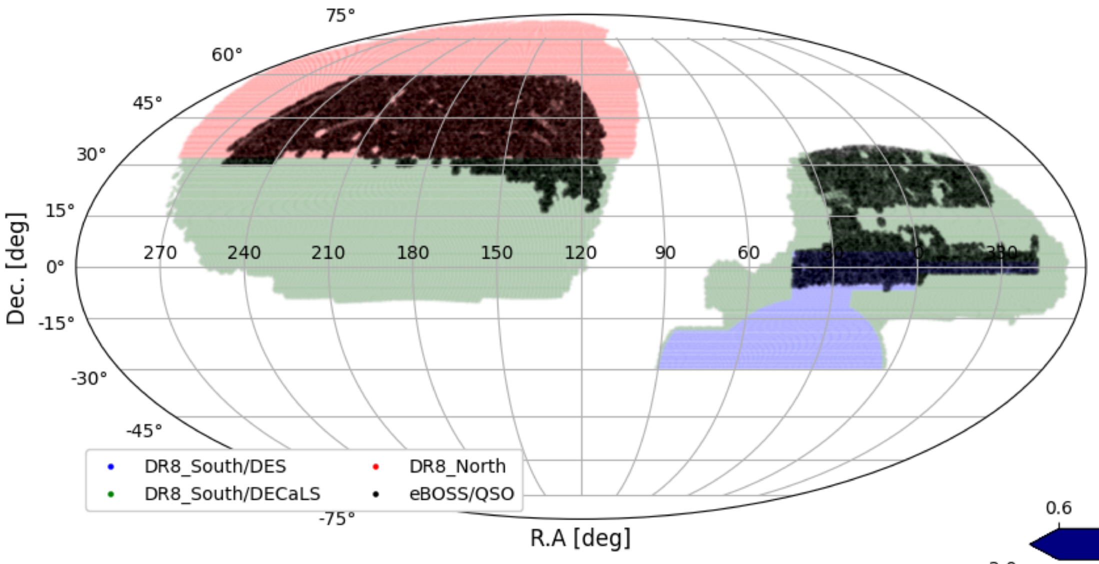




Redshift distribution of the eBOSS DR16 quasars.

This analysis uses another set of quasars in 0.6 <= z <= 1.5.

Datasets: Photometric sample
DESI Legacy Imaging Surveys



3 imaging surveys:
- BASS in dec > 32 deg: g and r-band
- MzLS in dec > 32 deg: z-band
- DECAM in dec < 32 deg : g, r and z-band (DECaLs and DES)

→ Treat each 3 region separately for systematics

Cut in NGC: dec > 32.315 deg → BASS/MzLS only

Cut in SGC: dec > 5 deg → DECaLS only

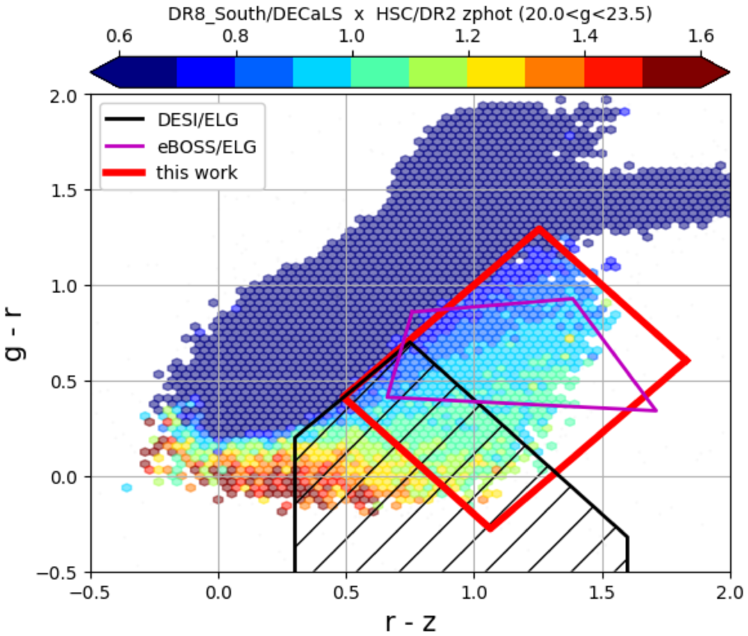
Selection criteria:

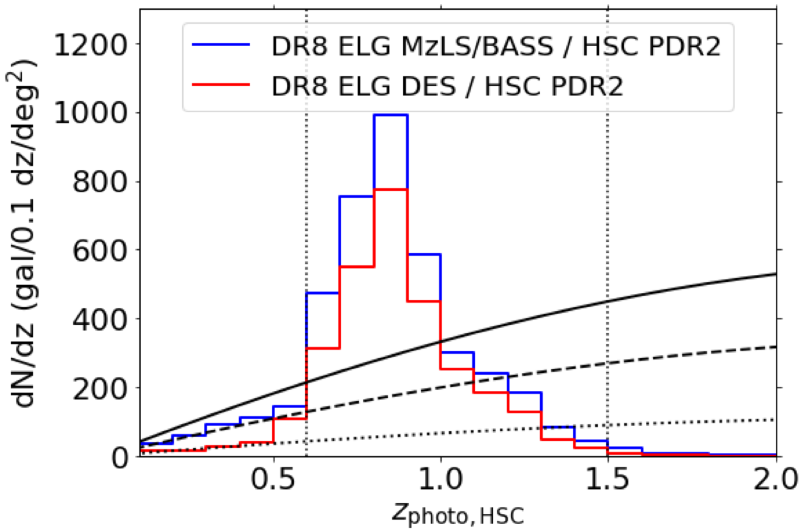
i) clean photometry

ii) high target density

iii) galaxies in 0.6 < z < 1.5

Criterion	DESI BASS/MzLS	DESI DECaLS
Clean photometry	$n_{\mathrm{obs},g/r/z} > 0$, GAIA stars masking and pixel masking	
Magnitude range	$20 < g < 23.6$	$20 < g < 23.5$
OII emitters	$g - r < -1.2 \cdot (r - z) + 2.8$	
Remove stars	$g - r < 1.15 \cdot (r - z) - 0.15$	
Redshift range	$1.15 \cdot (r - z) - 1.5 < g - r < 1.15 \cdot (r - z) - 0.15$	
	$-1.2 \cdot (r - z) + 1.0 < g - r < -1.2 \cdot (r - z) + 2.8$	





Datasets: Photometric sample
Neural Network (arXiv:1907.11355)

Observed galaxy density within a pixel i: $n_i^o(\mathbf{s}_i) = n_i \mathcal{F}(\mathbf{s}_i)$, where n_i true number of galaxies and F contamination model

→ **F unknow function which accounts for the systematic effects: output of the Neural Network**

In previous works, F = linear combination of the imaging attributes

Here we use the neural network approach to estimate F and: $\hat{n}_i = \frac{n_i^o}{\hat{\mathcal{F}}} = n_i^o \, wt_i^{\text{sys}}.$

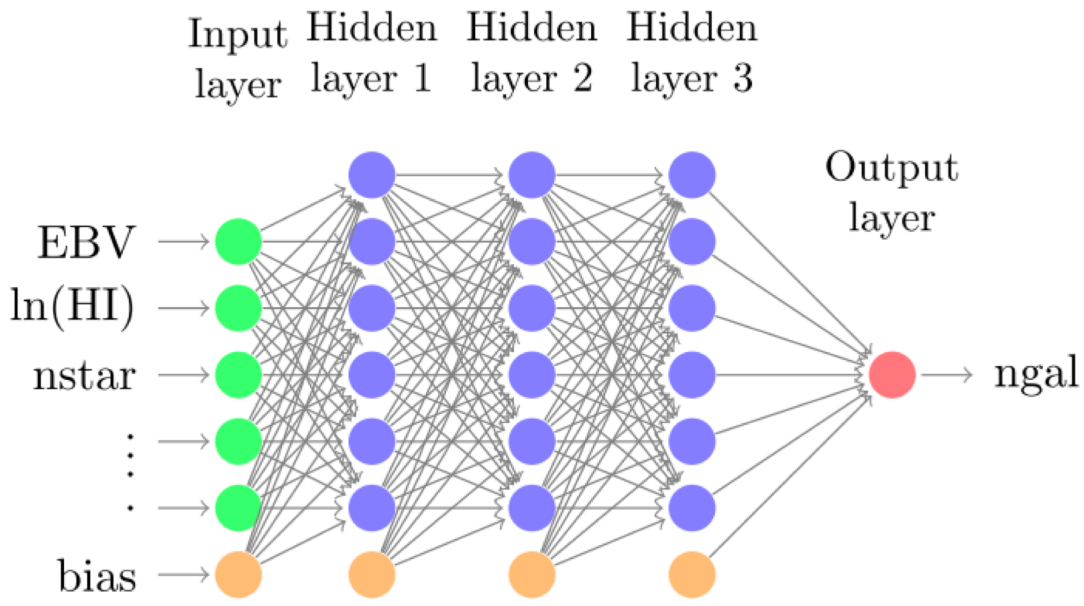
The network is trained to minimize the following cost function:

$$J = \frac{1}{N_{\text{batch}}} \sum_i^{N_{\text{batch}}} f_{\text{pix,i}} \, [t_i - \hat{\mathcal{F}}_i]^2 + \frac{\lambda}{2} \, ||w||^2,$$

Mean-Square Error

Regularization term to penalize higher weight magnitudes and larger number of neurons

N_{batch}: number of splits for the training set

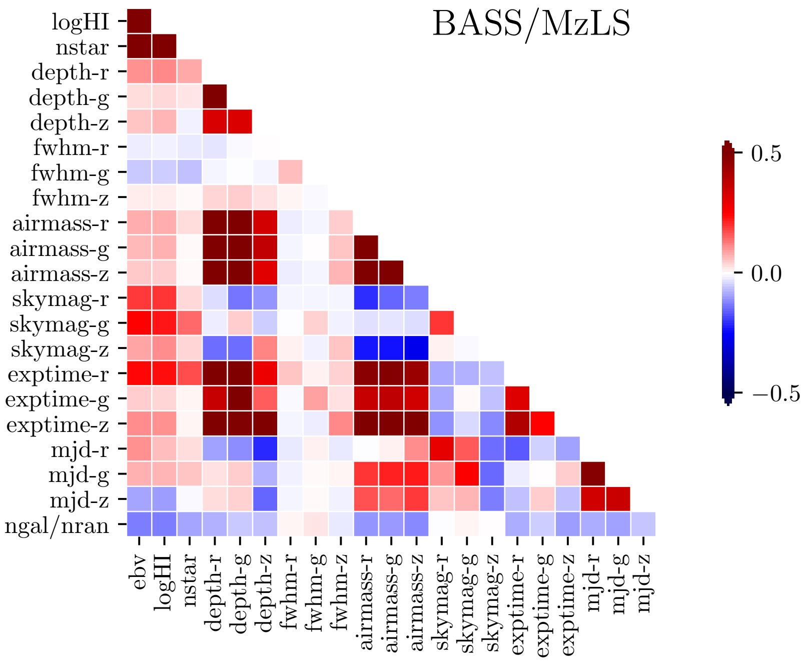


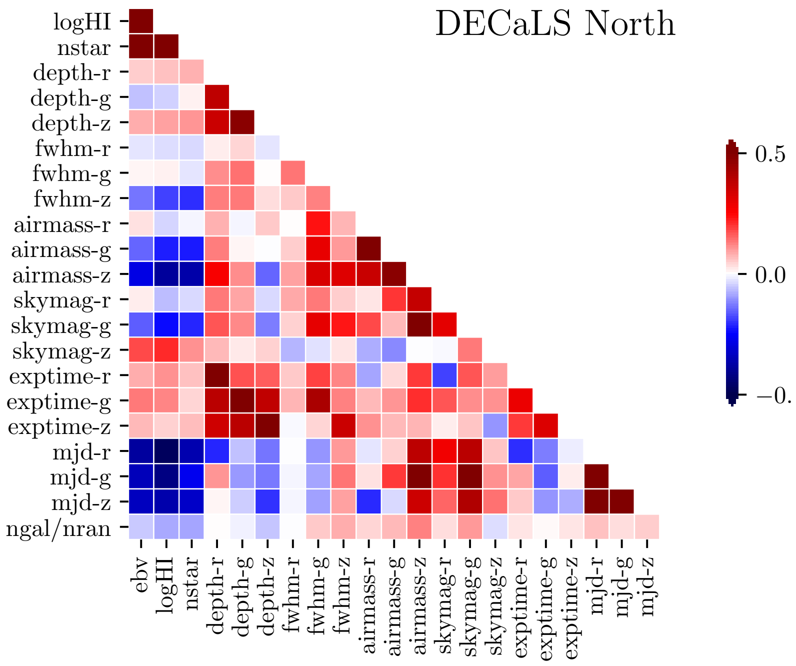
+ backward feature elimination (feature selection) to remove redundant input features and avoid overfitting

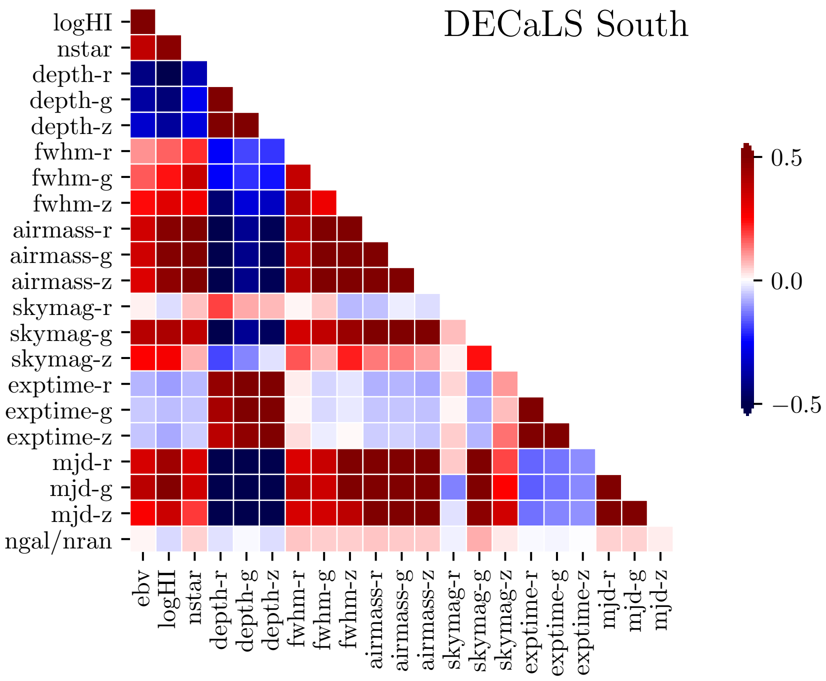
Purple: Non-linear activation functions

Datasets: Photometric sample
Neural Network to deal with imaging systematics

Imaging systematics from the CCDs file: Galactic extinction, Galaxy depth, Stellar density, Hydrogen atom column density, Sky brightness, Seeing, MJD, Exposure time
→ **Weights are derived by fitting separately BASS/MzLS, DECaLS-North and DECaLS-South**
+ comparison with linear and quadratic regression



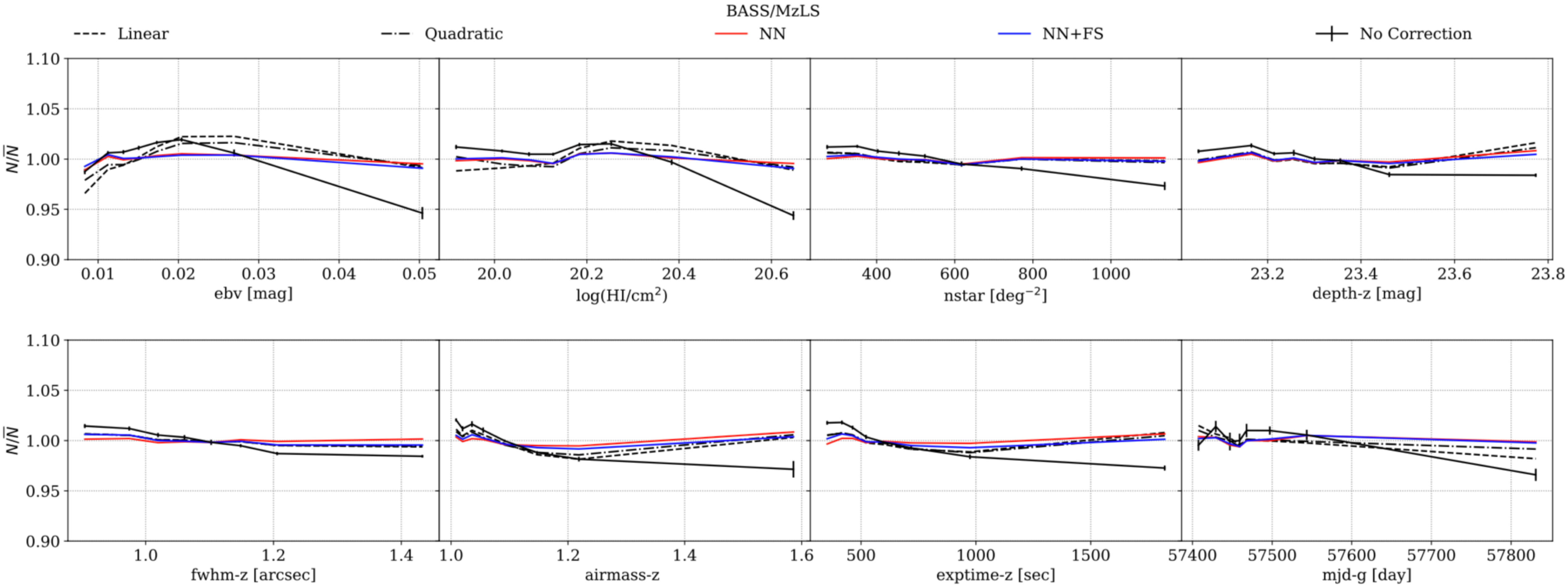




→ Complex dependencies with the imaging systematics which depend on the imaging surveys

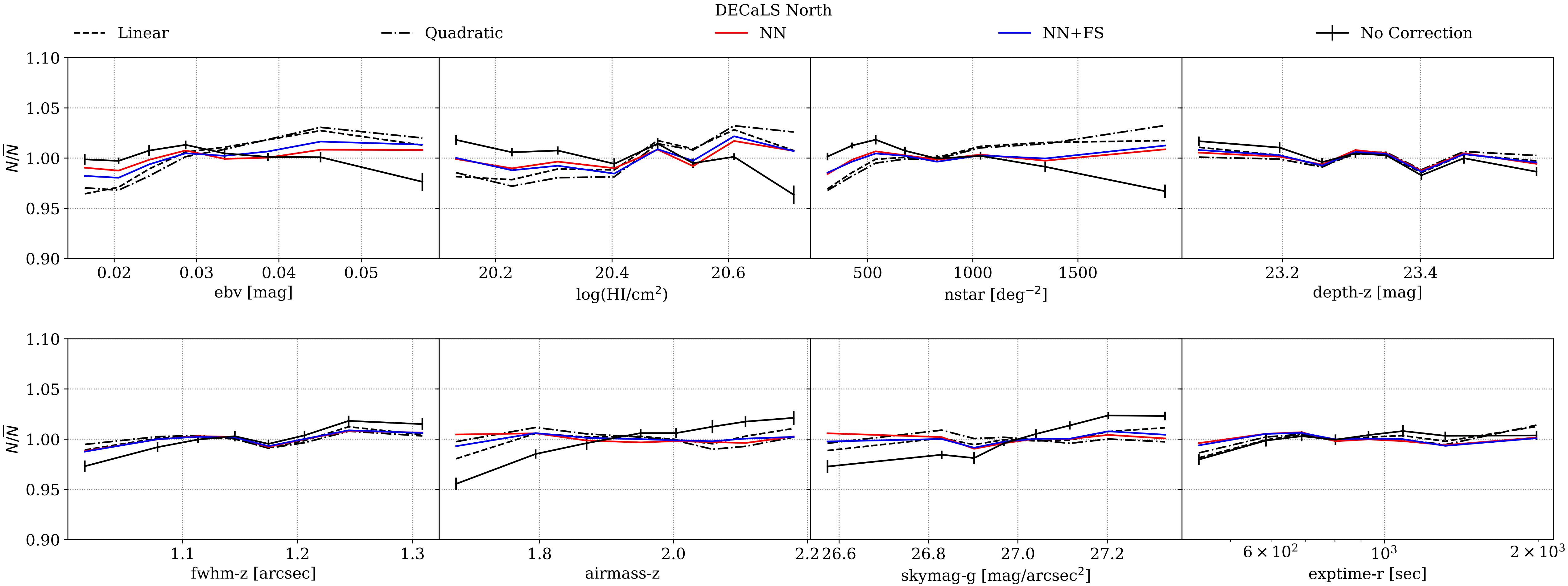
Datasets: Photometric sample

Target density variation with imaging systematics



The Neural Network (NN) mitigation technique removes most of the dependency with systematics

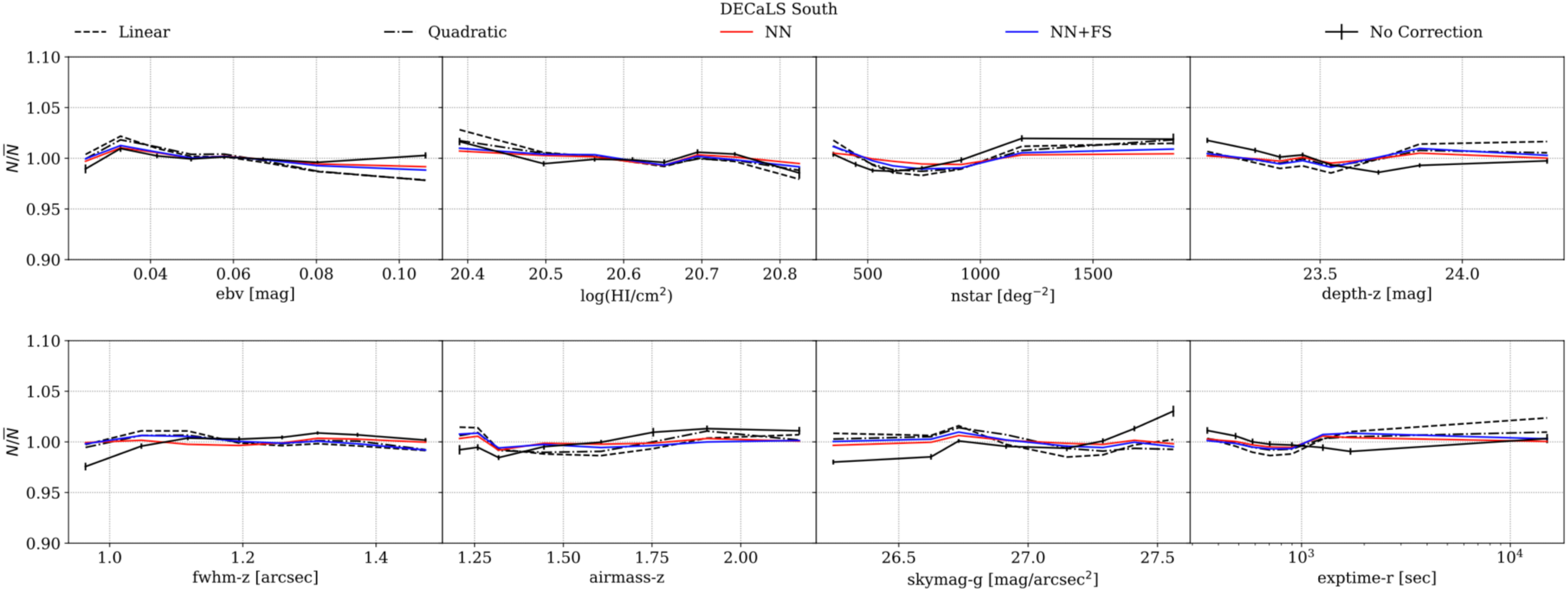
Datasets: Photometric sample
Neural Network to deal with imaging systematics



The Neural Network (NN) mitigation technique removes most of the dependency with systematics

Datasets: Photometric sample

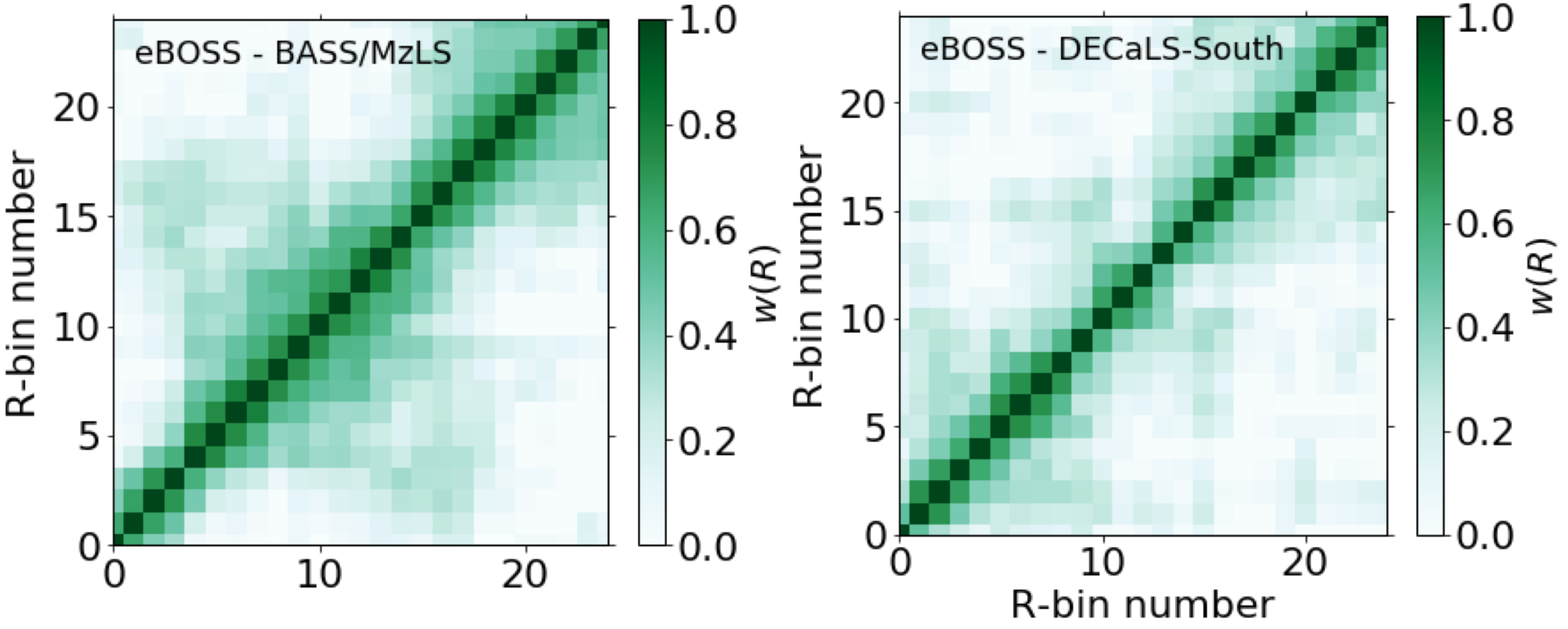
Neural Network to deal with imaging systematics



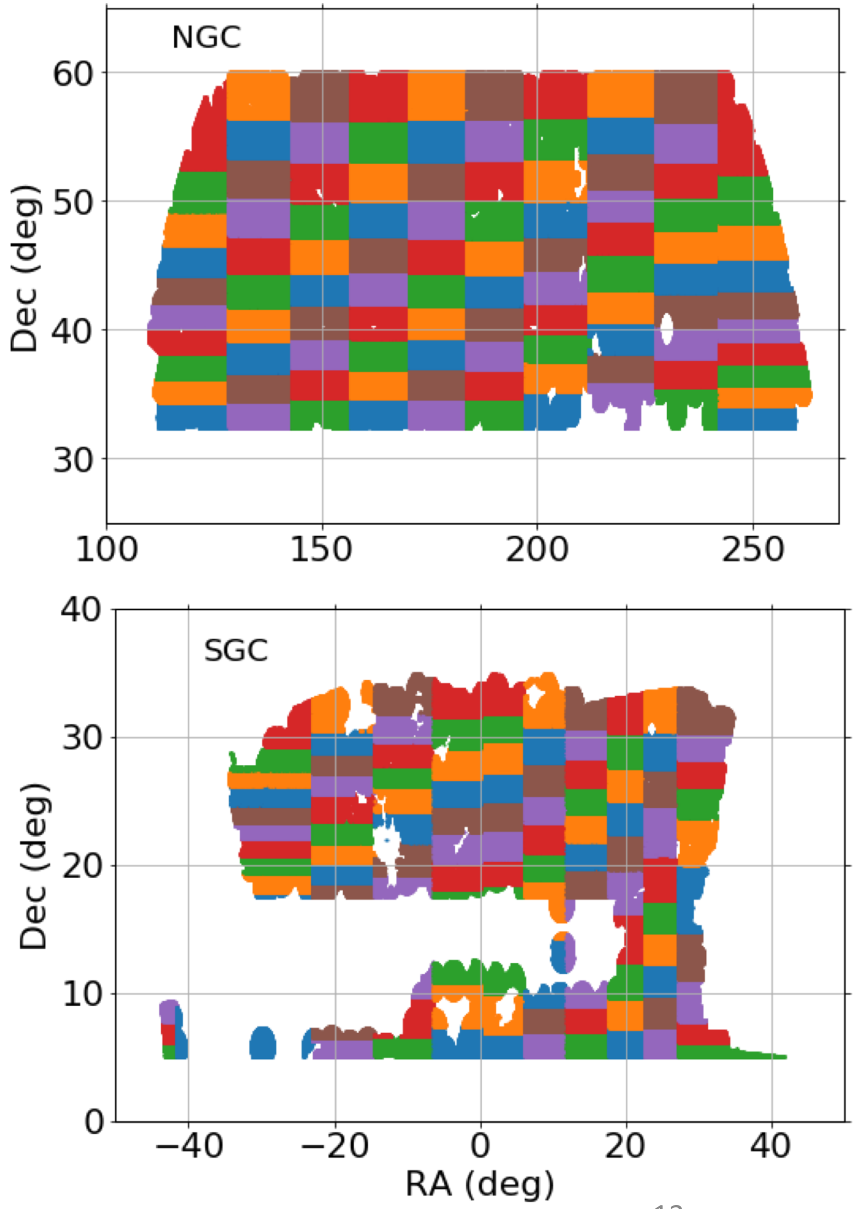
The Neural Network (NN) mitigation technique removes most of the dependency with systematics

Methodology
Covariance matrix

Covariance matrix from jackknife using a modified version of TWOPCF*
→ regions of similar area by splitting the survey with straight line cuts in RA and then Dec such that each region contains the same number of points in the random catalogue.



Covariance matrix of the projected cross-correlation function from the 100 equal-area jackknife realisations.



*https://github.com/lstothert/two_pcf

Methodology
BAO fitting procedure

BAO fitting procedure using a modified version of BAOfit (https://github.com/ashleyjross/BAOfit)
→ Same as in Ata et al. (2017) for eBOSS DR14 quasars
1. Generate a template with the BAO feature using $P_{lin}(k)$ from CAMB
2. Generate a template without the BAO feature, P_{nw} from the fitting formulae in Eisenstein & Hu (1998)

Same model for the auto- and projected cross-correlation function:

$$\xi^{mod}(s) = B_0 \xi_{temp}(\alpha, s) + A_1 + A_2/s + A_3/s^2$$

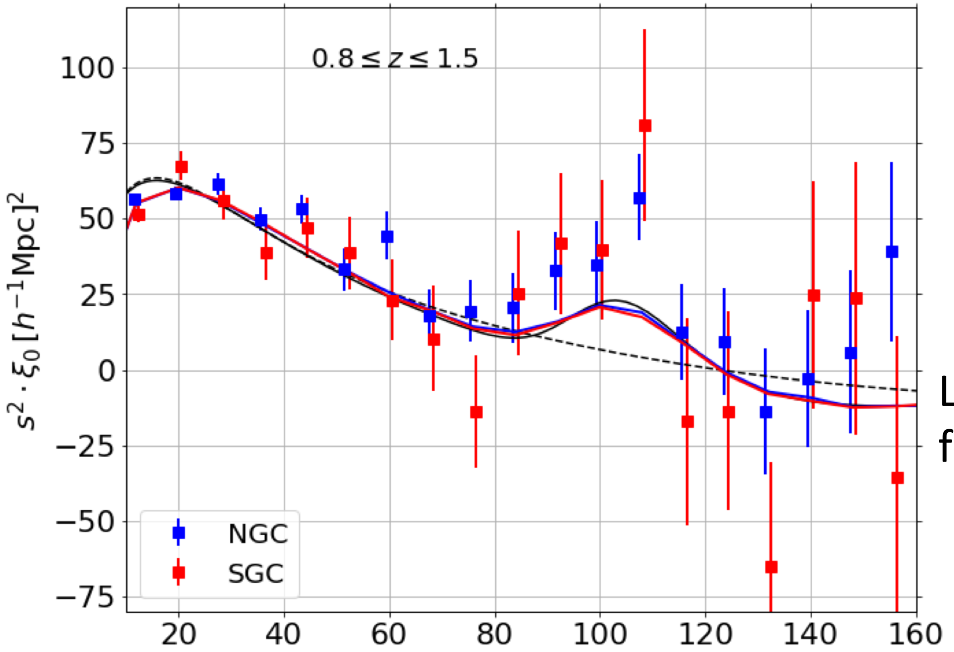
→ 1 cosmological parameter (α_{iso} or α_{perp}), 1 normalisation parameter B_0, 3 broad-band parameters ($A_{1,2,3}$)
Template for the projected correlation function with analytical prediction from slide 3

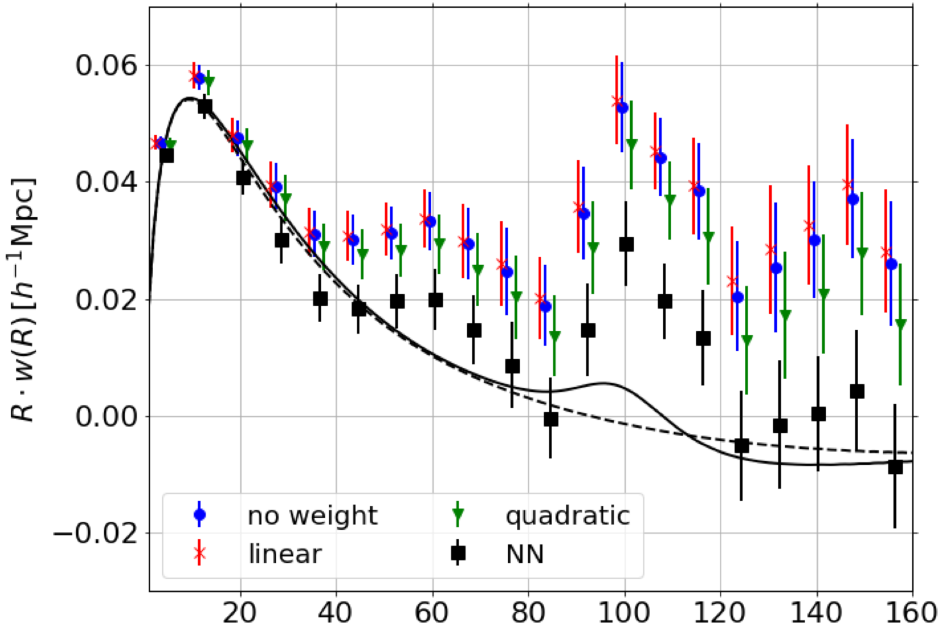
Template for the auto-correlation function:

$$\xi_{temp}(s) = \int \frac{k^2 dk}{2\pi^2} P_{temp}(k) j_0(ks) e^{-k^2 a^2} \quad \text{where} \quad P_{temp}(k) = P_{nw}(k)\left[1 + \left(\frac{P_{lin}(k)}{P_{nw}(k)} - 1\right) e^{\frac{1}{2}k^2\Sigma_{nl}^2}\right]$$

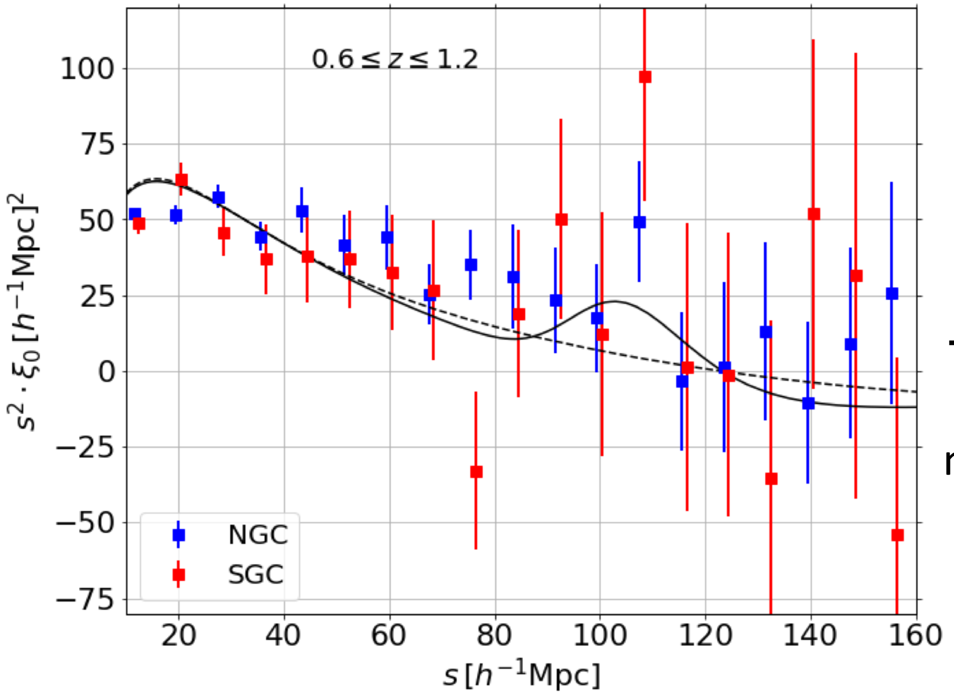
→ Damping term determined from tests using simulations

Results
Clustering



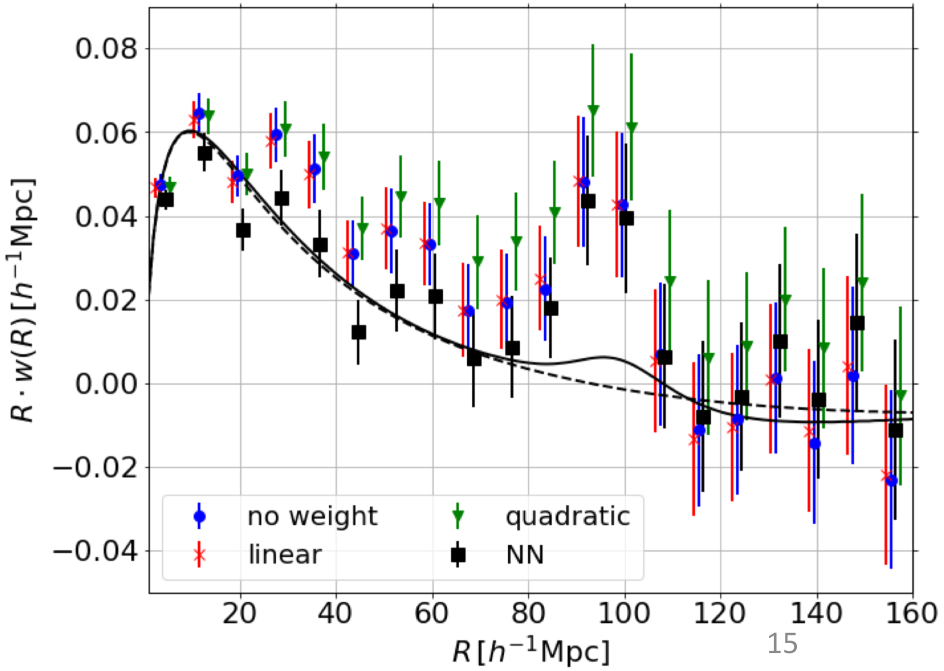


Left: auto-correlation function



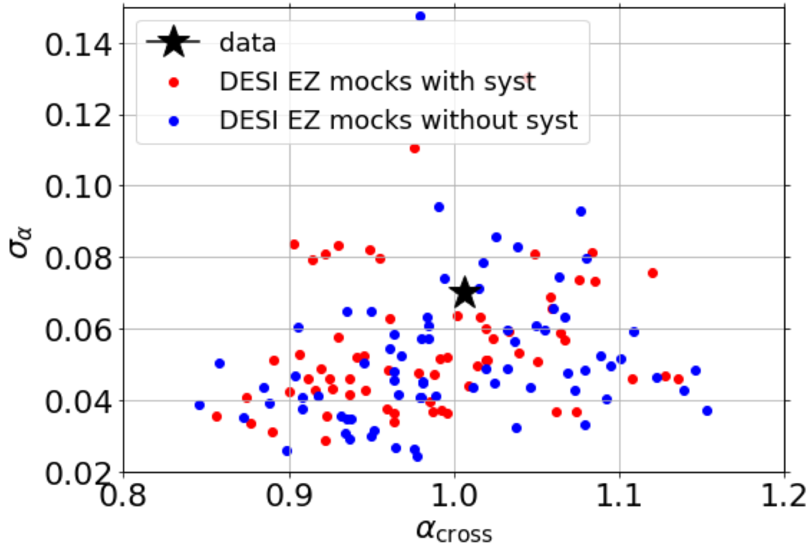
Right: projected cross-correlation function for different imaging systematics mitigation technique

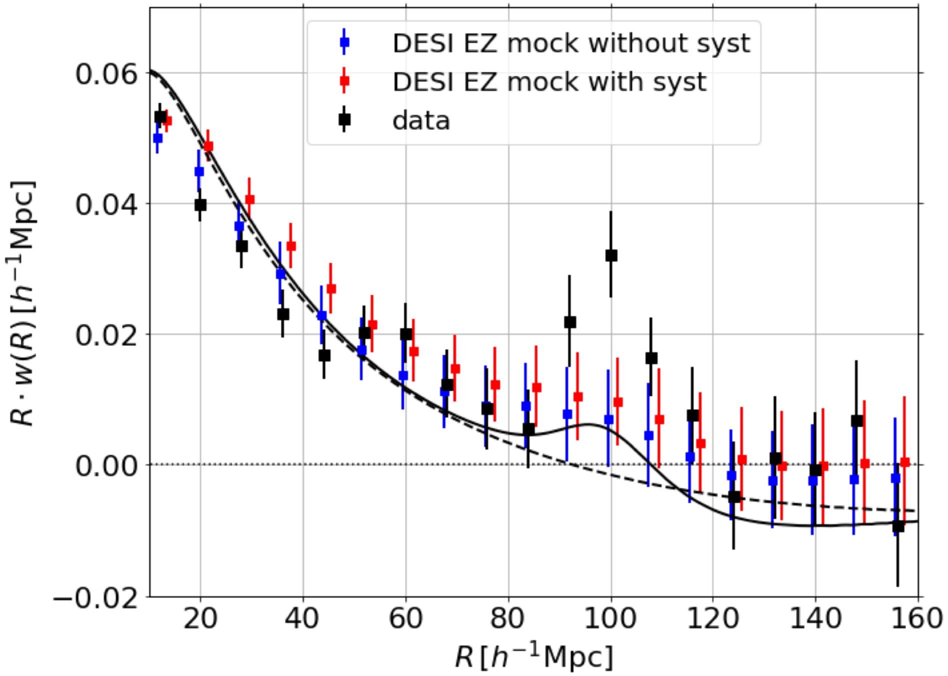
→ The NN approach provides the best correction, but mismatch of the amplitude of the BAO peak



Results
Testing the pipeline with mock catalogues

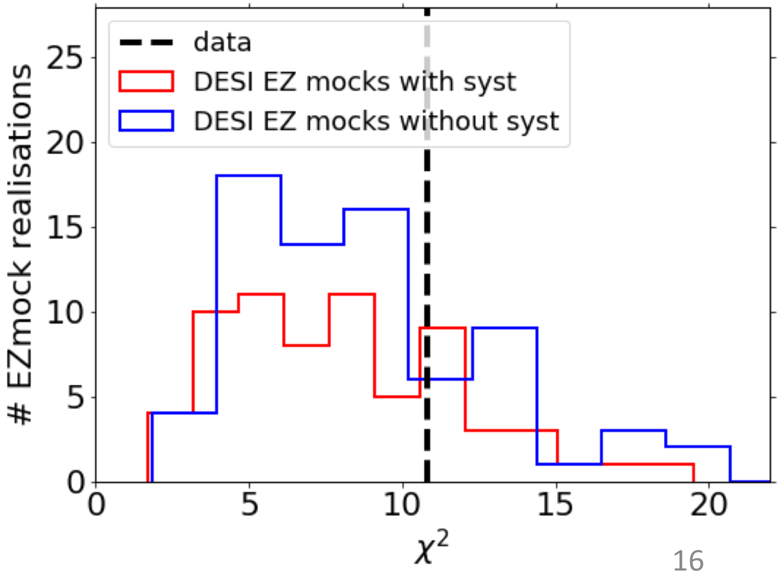
100 DESI EZ mocks (methodology of the EZ mocks in Chuang et al. 2015)
- Need both samples in each realisation: the spectroscopic sample is built by randomly downsampling the halos in 0.6 <= z <= 1.2 to match the density
- Imprint of the Legacy Imaging Surveys in the mock using HSC photometry and then photometric errors from the Legacy Surveys are added as a function of position in the sky.





→ BAO fit in the data is consistent with the statistics of the mocks

→ Amplitude in the template agrees with the one in the mocks



Results
Robustness tests

Baseline:
20 < s < 140 Mpc/h
Bin size of 8 Mpc/h

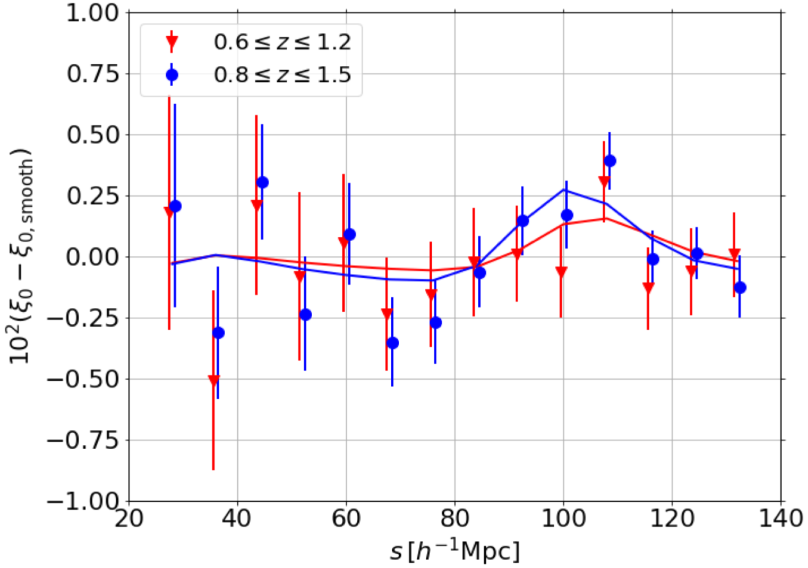
	Configuration	$\alpha_\perp$	χ^2/d.o.f.
	$0.8 \leqslant z \leqslant 1.5$		
	Fiducial	0.994 ± 0.051	9.1/9
	jackknife	0.994 ± 0.051	9.3/9
Binning	$\Delta s = 5h^{-1}$Mpc	1.005 ± 0.047	17.3/19
Fitting range	$20 < s < 150h^{-1}$Mpc	0.997 ± 0.049	9.4/11
	$10 < s < 140h^{-1}$Mpc	0.998 ± 0.052	10.8/11
Effective redshift*	$z_{\mathrm{eff},2}$	0.994 ± 0.051	9.1/9
Photometric weights	no w_{sys}	0.995 ± 0.044	8.6/9
	$w_{\mathrm{sys-lin}}$	0.989 ± 0.046	7.5/9
	$w_{\mathrm{sys-quad}}$	0.988 ± 0.046	8.4/9
	$w_{\mathrm{sys-nn-fs}}$	0.993 ± 0.046	9.4/9
Consistency NGC/SGC	NGC	0.970 ± 0.066	7.8/9
	SGC	1.025 ± 0.109	5.9/9

Configuration	$\alpha_\perp$	χ^2/d.o.f.
$0.6 \leqslant z \leqslant 1.2$		
Fiducial	0.999 ± 0.059	13.1/9
jackknife	0.999 ± 0.059	13.3/9
$\Delta_s = 5h^{-1}$Mpc	1.014 ± 0.058	20.2/19
$20 < s < 150h^{-1}$Mpc	0.991 ± 0.059	14.8/11
$10 < s < 140h^{-1}$Mpc	1.003 ± 0.061	13.5/11
no w_{sys}	1.004 ± 0.057	11.3/9
$w_{\mathrm{sys-lin}}$	0.997 ± 0.059	12.9/9
$w_{\mathrm{sys-quad}}$	0.994 ± 0.058	13.7/9
$w_{\mathrm{sys-nn-fs}}$	0.999 ± 0.056	13.4/9
NGC	0.965 ± 0.078	16.2/11
SGC	1.045 ± 0.112	5.3/11

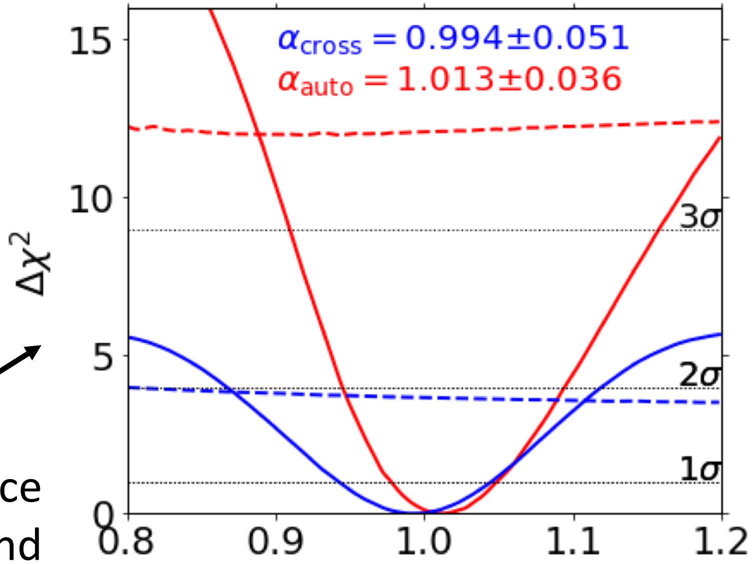
*Effective redshift
- Same definition as for the auto-correlation function (baseline)
- Weighted by the S/N of the cross-correlation signal using dz=0.1

Results

BAO measurements

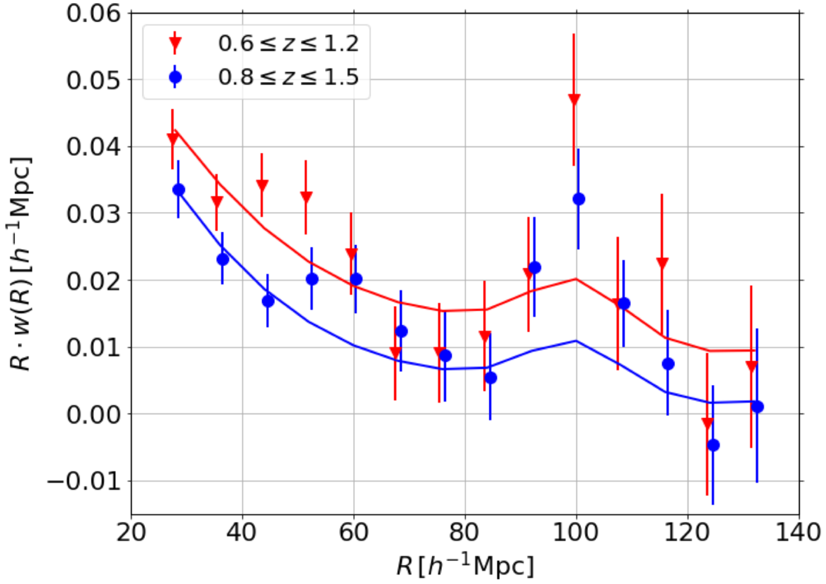


From the auto-correlation function

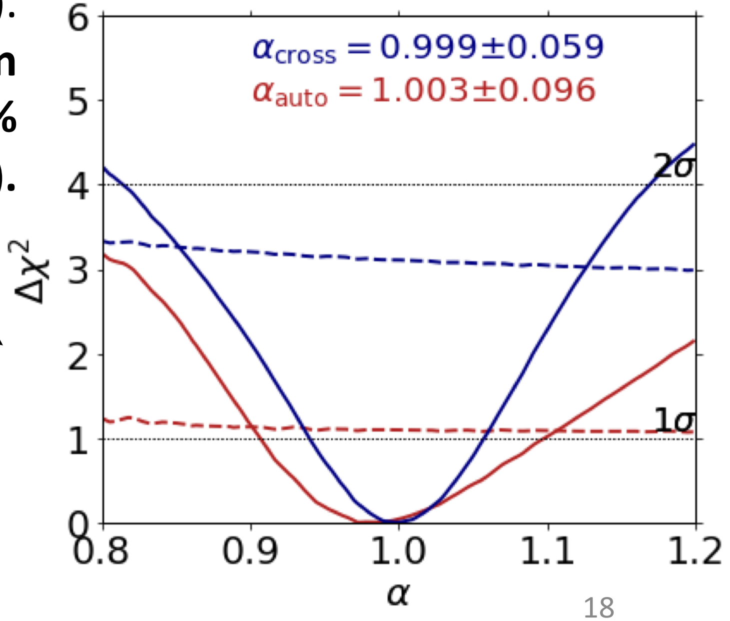


We can quantify the BAO peak significance by comparing the $\Delta\chi^2$ with BAO (solid) and the one without (dashed).

For 0.6 <= z <= 1.2, the cross-correlation technique gives more precise result (6% precision) than the auto (9% precision).



From the cross-correlation function



Results
BAO distance ladder

Not yet competitive with other measurements of D_M
But:

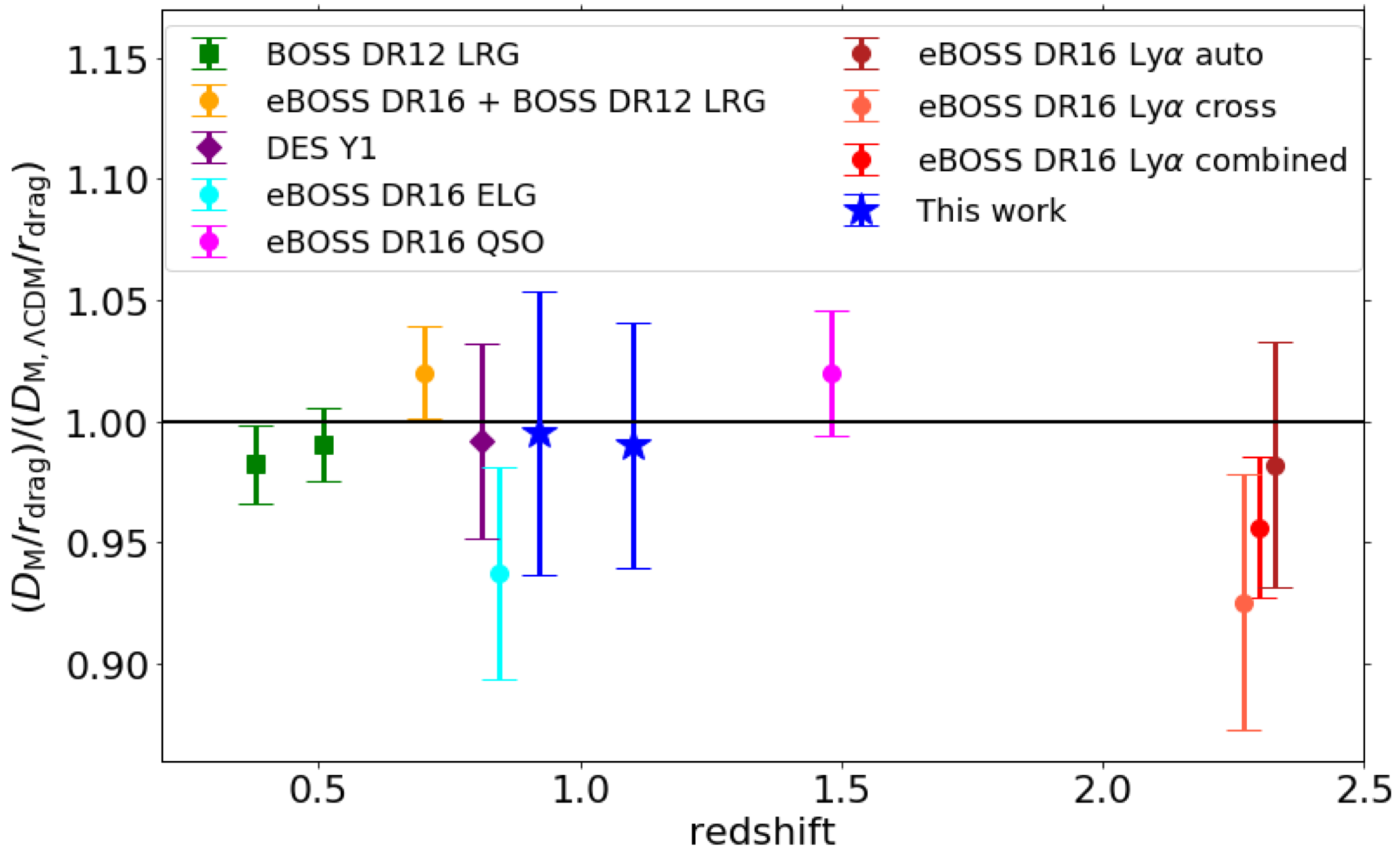
- The projected cross-correlation technique suffers limitation from the depth of the imaging surveys used in this analysis.

- The auto-correlation technique itself uses both the monopole and the quadrupole to get accurate constraints on D_M (and H).

- Could be combined with the result from the auto as done for the Ly-α

- Upcoming deeper photometric surveys with *Euclid* and LSST

→ In particular, using a sample of Hα (0.7 < z < 2) and [OIII] emission galaxies (2 < z < 2.7) selected from *Euclid* (Mehta et al. 2015)



Summary

- Previous works suggest that we can measure the BAO in the projected cross-correlation function between a sparse spectroscopic sample and a dense photometric sample

- **First detailed study of the projected cross-correlation technique using the latest data**: eBOSS DR16 quasars in 0.6 <= z <= 1.5 and photometric galaxies using the DESI Legacy Imaging Surveys with a tailored selection.

- **Main limitations in this analysis**:
  - Overlap in redshift between the photometric and spectroscopic samples
  - Density of the photometric sample → special colour selection
  - Purity of the photometric sample → imaging systematics mitigation technique using a Neural Network approach

Redshift range	area [deg^2]	spectro [deg^{-2}]	photo [deg^{-2}]	σ_{auto}	σ_{cross}
$0.8 \leqslant z \leqslant 1.5$	4000	35	2100	3.5%	5%
$0.6 \leqslant z \leqslant 1.2$	4000	20	2900	9%	6%
$0.6 < z < 0.8$	6000	35	1100	–	3%

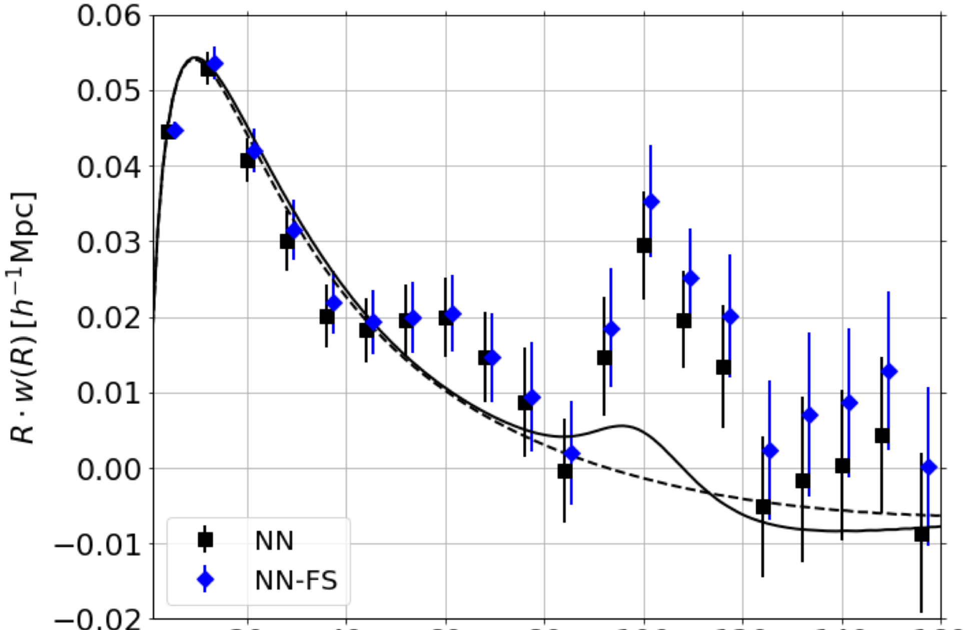
- **Future potential applications**:
  - DESI quasars (z < 2.2) x LSST/Euclid
  - For Euclid, sample of Hα (0.7 < z < 2) and [OIII] emission galaxies (2 < z < 2.7) as proposed in Mehta et al. 2015

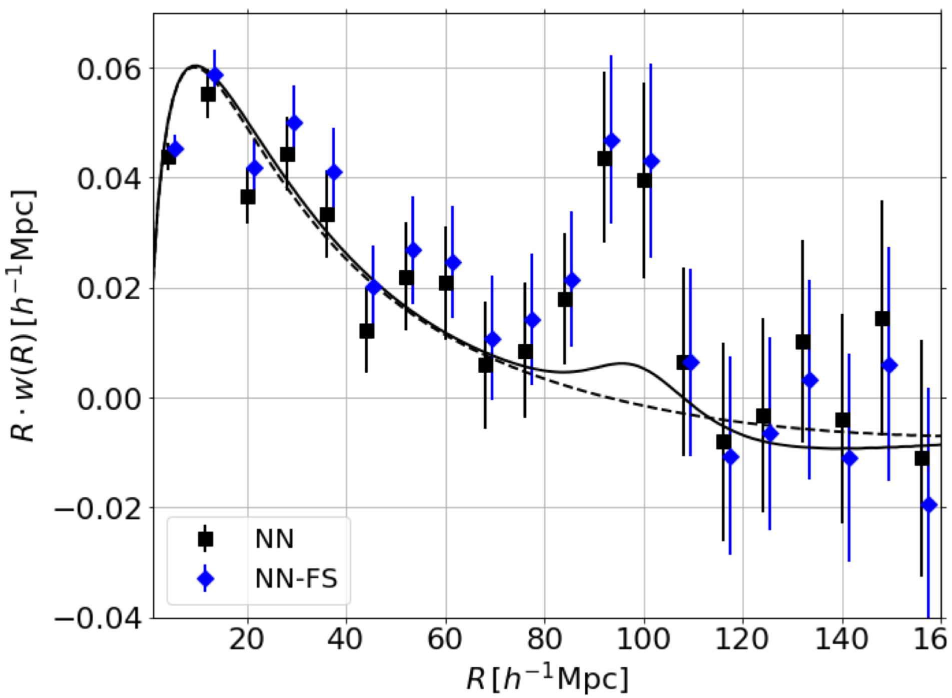
Back-up

Neural Network
Feature selection

Backward feature elimination (feature selection) to remove the redundant input features in order to reduce the noise in the prediction as well as to protect the cosmological information by avoiding too much freedom in modeling.

In Rezaie et al. 2019, they found that this step was essential to avoid overfitting.
In this analysis, adding this step has a small impact on the projected cross-correlation function.



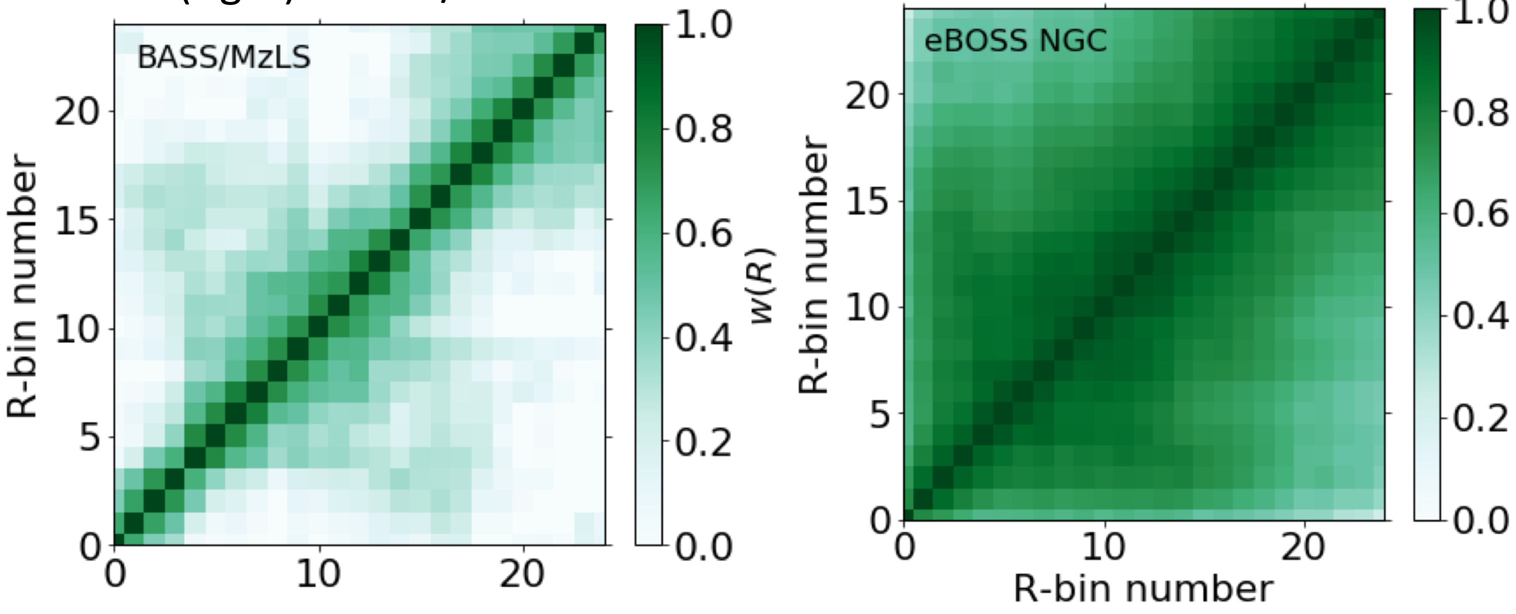


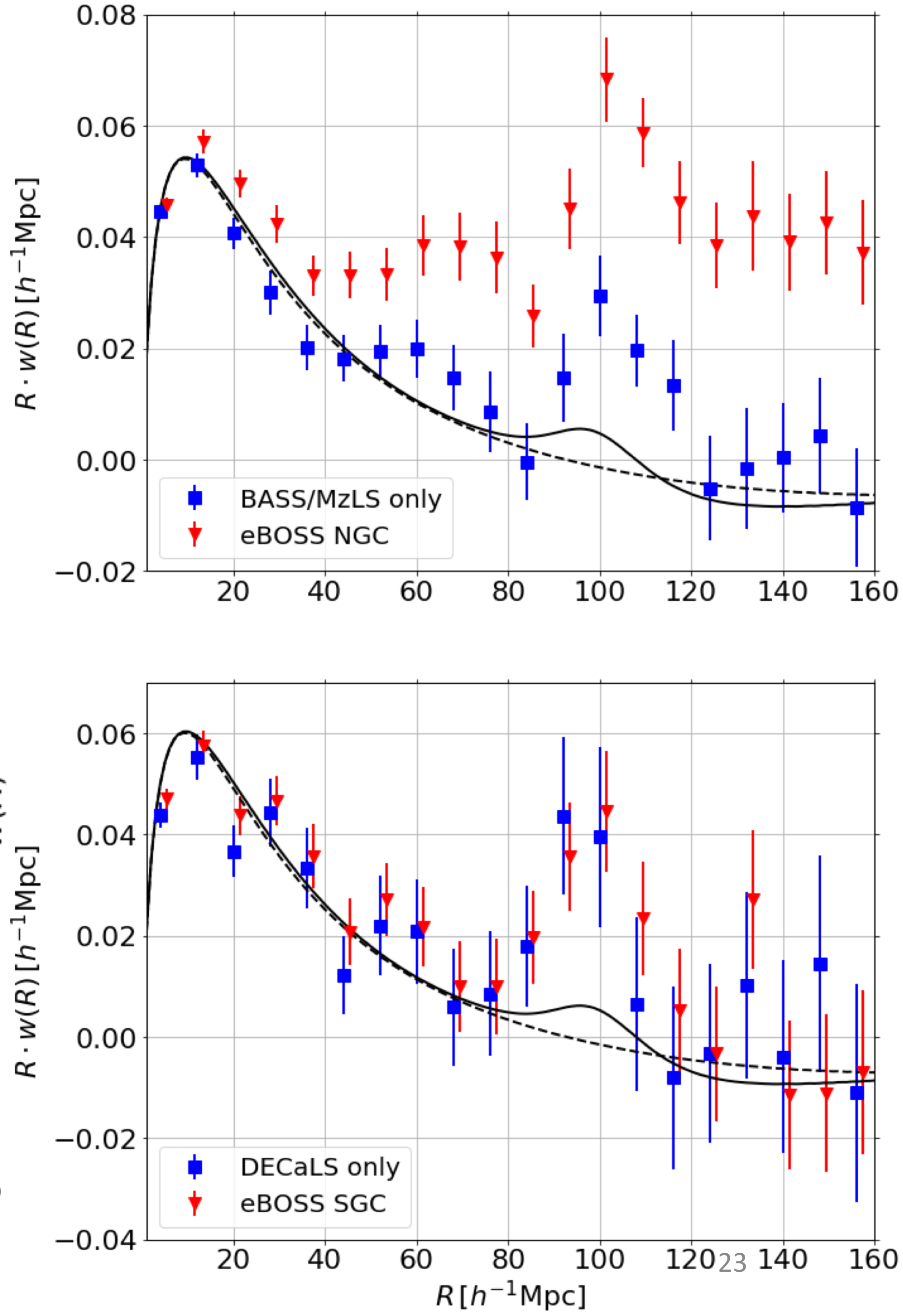
Dealing with different imaging surveys

- In the NGC we treat separately BASS/MzLS and DECaLS
- In the SGC, we treat separately DECaLS and DES

→ Because they correspond to imaging surveys with different depths and sensitivity to the systematics

Below: Covariance matrix in BASS/MzLS only (left) and the entire eBOSS NGC (right) = BASS/MzLS + DECaLS





Left: Cross-correlation function of BASS/MzLS (DECaLS) only on the top (bottom) compared to the entire NGC (SGC).