Mitigation of observational effects for galaxy clustering with photometric surveys

Martín Rodríguez Monroy

IJCLab, A2C Seminars, 19th September 2022

IJCLab Irène Joliot-Curie
Laboratoire de Physique des 2 Infinis

THE DARK ENERGY SURVEY

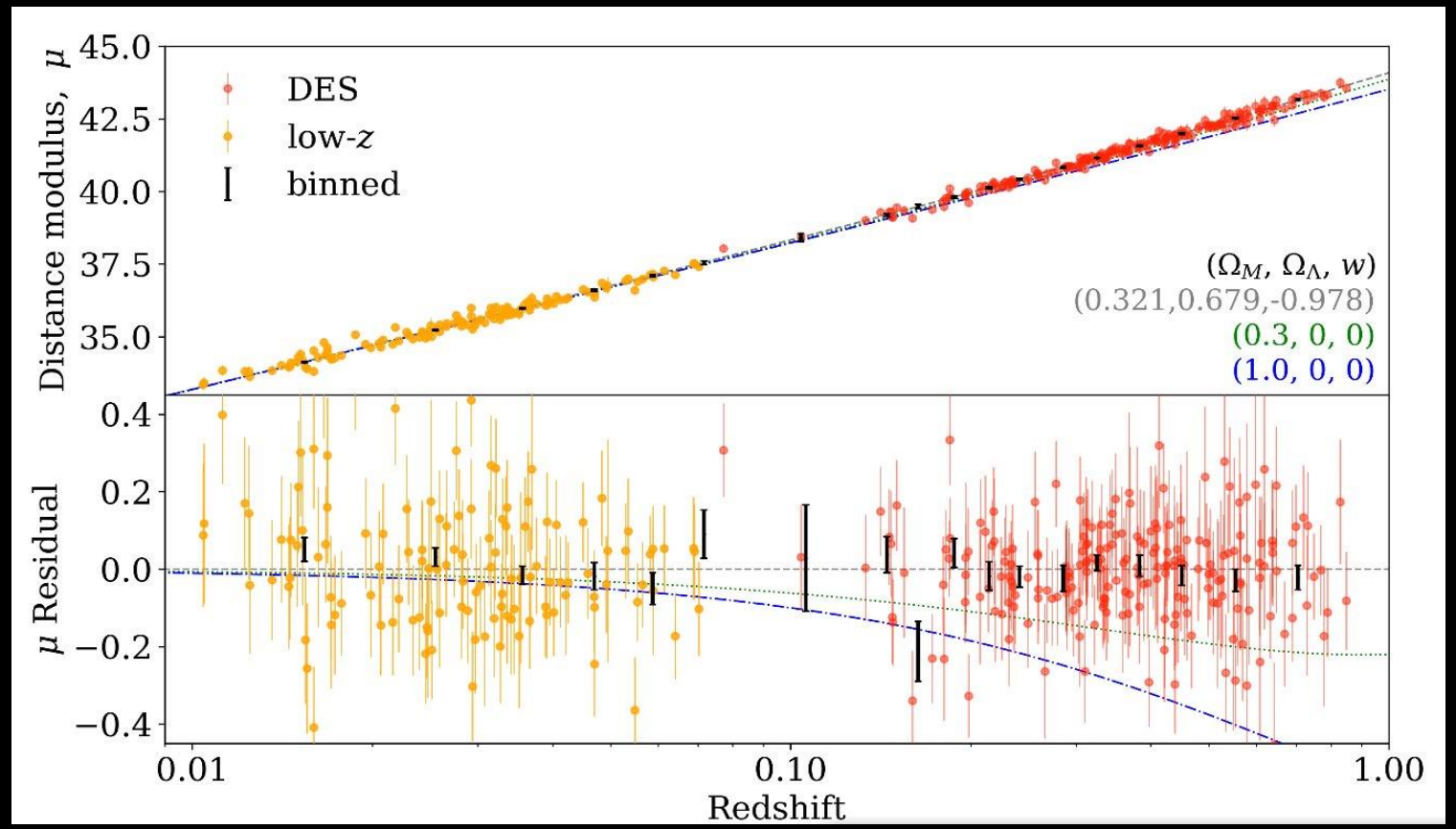
Outline

- Standard cosmological model, ΛCDM
- Large-scale structure and galaxy clustering
- The Dark Energy Survey
- DES Y3 results on galaxy clustering and cosmological parameters
- Systematic effects on galaxy clustering
- Methods for observational systematics mitigation
- Validation of the weights and robustness tests
- Lessons learned

A long time ago, everywhere….

Accelerated expansion of the Universe: dark energy

- ★ The Universe came into existence with a rapid expansion from a state of high density and temperature → the Big Bang
- ★ The Universe expands at an accelerated rate



Accelerated expansion of the Universe: dark energy

- ★ The Universe came into existence with a rapid expansion from a state of high density and temperature → the Big Bang
- ★ The Universe expands at an accelerated rate
- ★ What is behind this acceleration?
  - Equation of state of the fluid: $p = w \cdot \rho$
  - Acceleration equation: $\frac{\ddot{a}}{a} = \frac{-4\pi G}{3}(\rho + 3p)$
  - Dark energy: $w < -1/3$

Accelerated expansion of the Universe: dark energy

- ★ The Universe came into existence with a rapid expansion from a state of high density and temperature → the Big Bang
- ★ The Universe expands at an accelerated rate
- ★ What is behind this acceleration?
- ★ Several candidates:
  - Quintessence: $-1 < w < -1/3$
  - Phantom dark energy: $w < -1$
- ★ Other explanations:
  - Incompleteness of General Relativity at large scales
  - Primordial magnetic fields

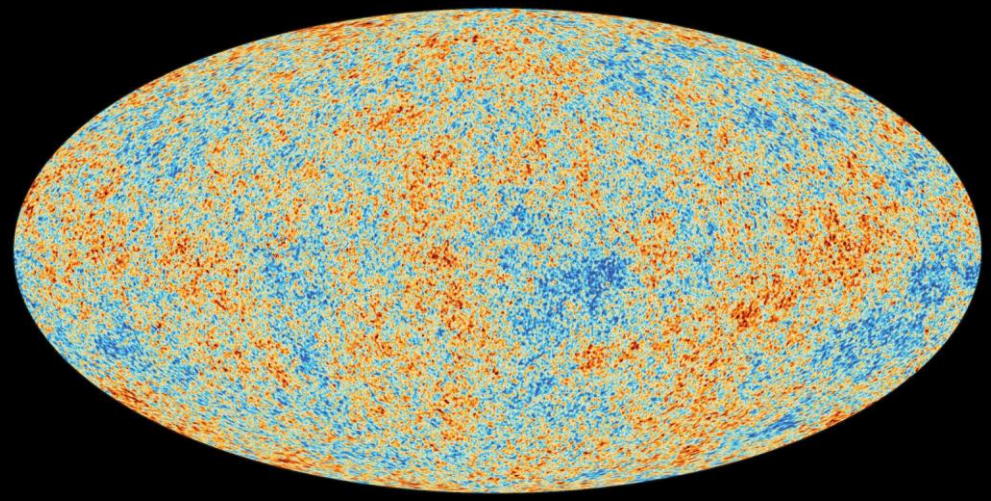
Accelerated expansion of the Universe: dark energy

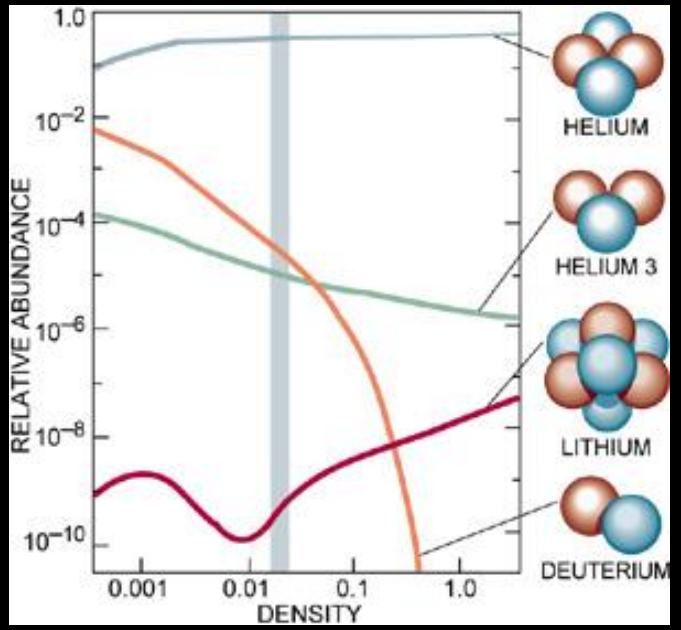
- ★ The Universe came into existence with a rapid expansion from a state of high density and temperature → the Big Bang
- ★ The Universe expands at an accelerated rate
- ★ What is behind this acceleration?
- ★ Several candidates:
  - Quintessence: $-1 < w < -1/3$
  - Phantom dark energy: $w < -1$
- ★ Other explanations:
  - Incompleteness of General Relativity at large scales
  - Primordial magnetic fields
- ★ All observations to date suggest that the dark energy is the cosmological constant, Λ:

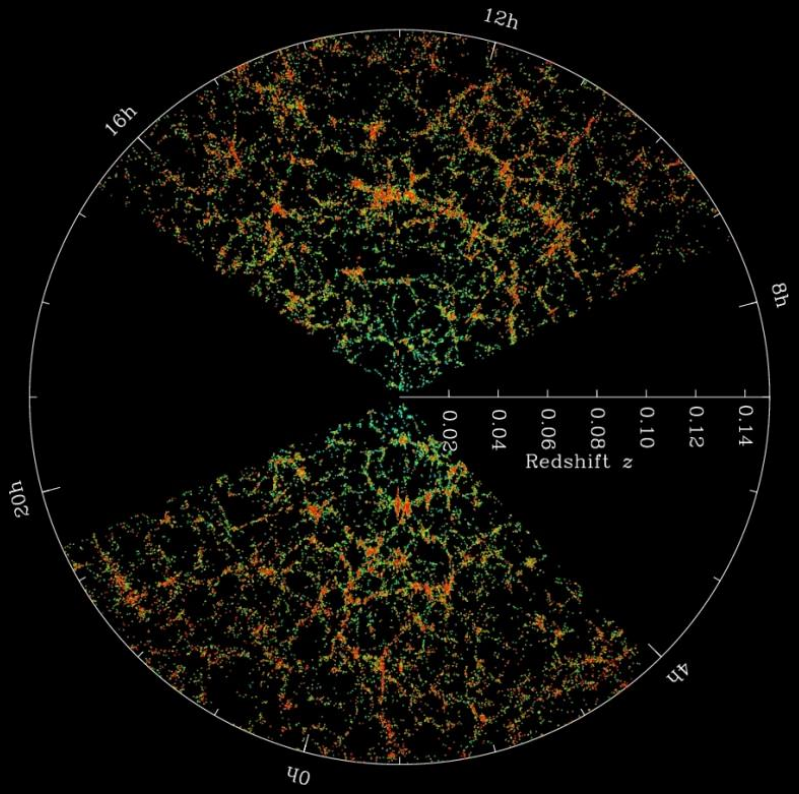
$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R + g_{\mu\nu}\Lambda = \frac{8\pi G}{c^4} T_{\mu\nu} \quad \begin{cases} w = -1 \\ \Lambda = (4.24 \pm 0.11) \cdot 10^{-66}\ \mathrm{eV}^2 \end{cases}$$

The standard cosmological model, ΛCDM

- ★ Observational basis:
  - Accelerated expansion
  - Cosmic microwave background
  - Abundances of light elements
  - Large-scale structure

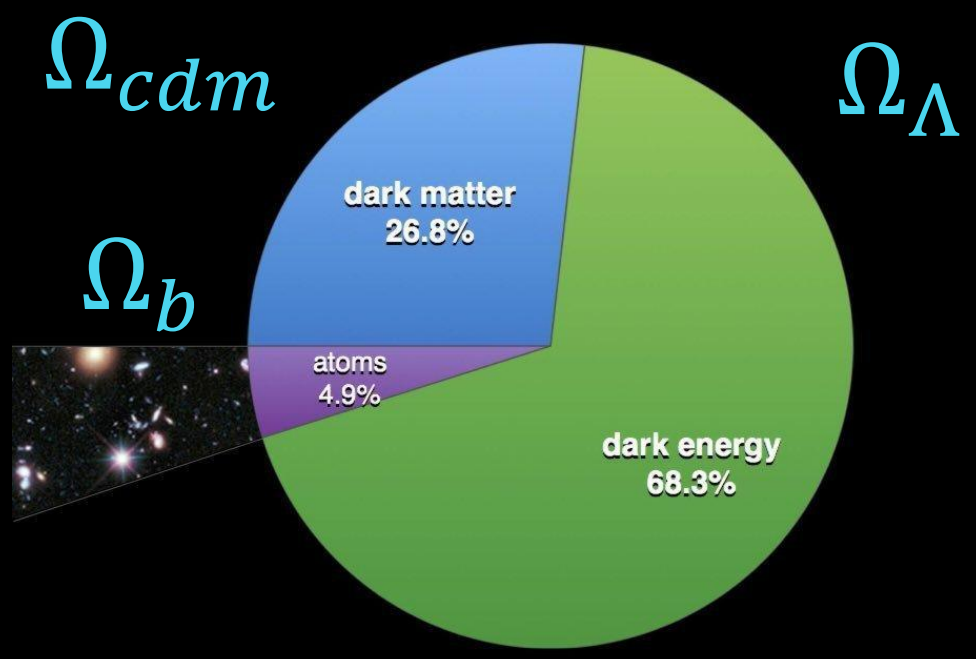






The standard cosmological model, ΛCDM

- ★ Observational basis:
  - Accelerated expansion
  - Cosmic microwave background
  - Abundances of light elements
  - Large-scale structure
- ★ Principal elements:
  - Cold dark matter (CDM)
  - Dark energy
  - Cosmological parameters:
    - Matter density, $\Omega_m = \Omega_b + \Omega_{cdm}$
    - Hubble parameter, h
    - Amplitude and scale dependence of primordial fluctuations, A_s and n_s
    - Optical depth at reionisation, τ



@AstroKatie/Planck13

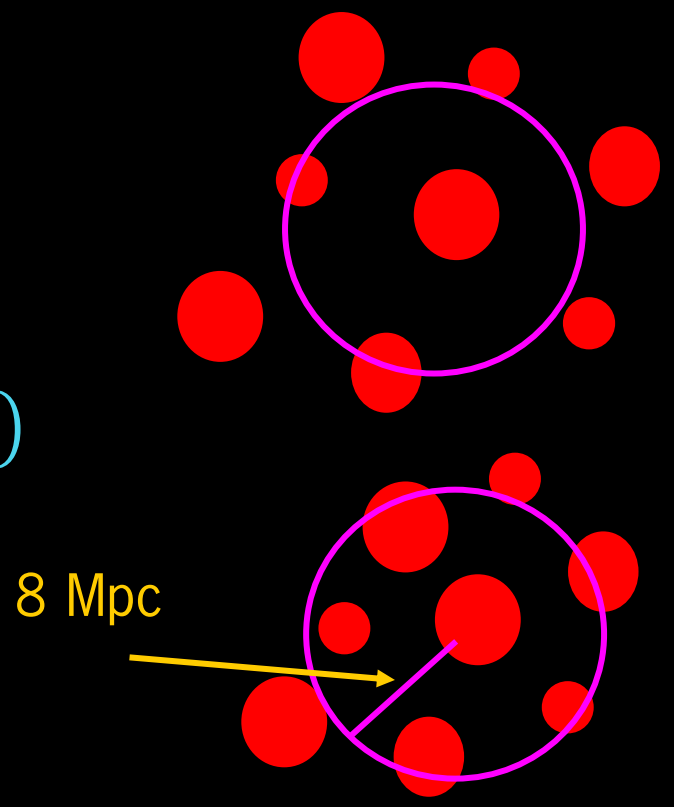
The standard cosmological model, ΛCDM

- ★ Observational basis:
  - Accelerated expansion
  - Cosmic microwave background
- ★ Principal elements:
  - Cold dark matter (CDM)
  - Dark energy
  - Cosmological parameters
- ★ Rest of the parameters are derived from them

$$H^2(a) = H_0^2\left(\Omega_{m,0}\, a^{-3} + \Omega_{r,0}\, a^{-4} + \Omega_{\Lambda,0} + \Omega_{k,0}\, a^{-2}\right)$$

- ★ Derived parameters: σ_8
  - How clumpy is the Universe?

Observational basis:
- Abundances of light elements
- Large-scale structure



Cosmological probes of dark energy

Dark energy and cosmological parameters can be probed in several ways:

- ★ Type Ia supernovae
- ★ Abundance of galaxy clusters
- ★ Baryon acoustic oscillations (BAO)
- ★ Galaxy clustering
- ★ Weak lensing
- ★ Strong lensing time delays
- ★ Multi-messenger GW astronomy

→

Two types of probes:

- ★ Geometrical
- ★ Evolution

Cosmological probes of dark energy

Galaxy clustering measurements, $w(\theta)$

- Correlates position – position of lens galaxies
- Sensitive to Ω_m, but degeneracies with galaxy bias and σ_8

Idea: combine with weak lensing measurements:

- Cosmic-shear, $\xi_\pm(\theta)$: correlates shapes – shapes of source galaxies
- Galaxy-galaxy lensing, $\gamma_t(\theta)$: correlates positions of lens galaxies – shapes of source galaxies

Combinations:

- $w(\theta) + \gamma_t(\theta) \rightarrow 2 \times 2$pt probe
- $w(\theta) + \gamma_t(\theta) + \xi_\pm(\theta) \rightarrow 3 \times 2$pt probe

The large-scale structure of the Universe: galaxy clustering....

Two-point angular correlation function

- Photometric surveys ⇒ uncertainty in radial distance → project in 2D bins
- Projected density contrast:

$$\tilde{\delta}_G\left(\vec{\theta}\right) = \int dz\, \phi^i(z)\, \delta_G\left(\vec{\theta}, z\right), \quad \delta(\vec{r}) \equiv \frac{\rho(\vec{r})}{\bar{\rho}} - 1$$

- Radial selection function:

$$\phi^i(z) = \frac{dN_g}{dz} \int dz_{survey}\, P\left(z \middle| z_{survey}\right) W^i(z_{survey}) = n_g^i(z)$$

Two-point angular correlation function

- Photometric surveys ⇒ uncertainty in radial distance → project in 2D bins
- Projected density contrast:

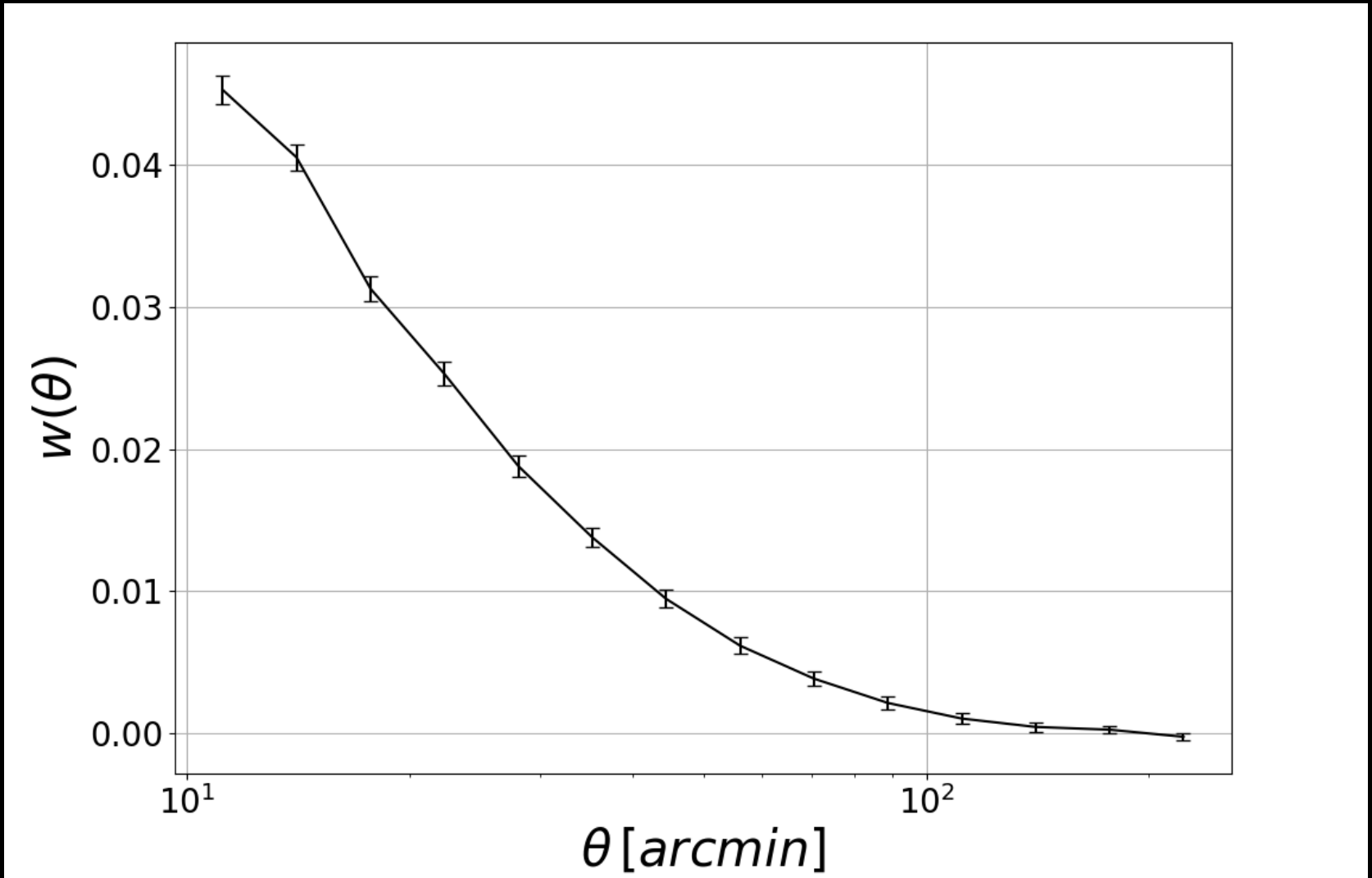
$$\tilde{\delta}_G\left(\vec{\theta}\right) = \int dz\, \phi^i(z)\, \delta_G\left(\vec{\theta}, z\right), \quad \delta(\vec{r}) \equiv \frac{\rho(\vec{r})}{\bar{\rho}} - 1$$

- Radial selection function:

$$\phi^i(z) = \frac{dN_g}{dz} \int dz_{survey}\, P\left(z \middle| z_{survey}\right) W^i(z_{survey}) = n_g^i(z)$$

- Correlation on the 2D sphere → $w(\theta) = \left\langle \tilde{\delta}_G(\vec{\theta}_1)\, \tilde{\delta}_G(\vec{\theta}_2) \right\rangle$
- Access to δ_G → biased tracer of matter ⇒ $\delta_m = b\, \delta_G$ → linear galaxy bias
- We measure the galaxy correlation function → $w_m(\theta) = b^2\, w_G(\theta)$

Two-point angular correlation function

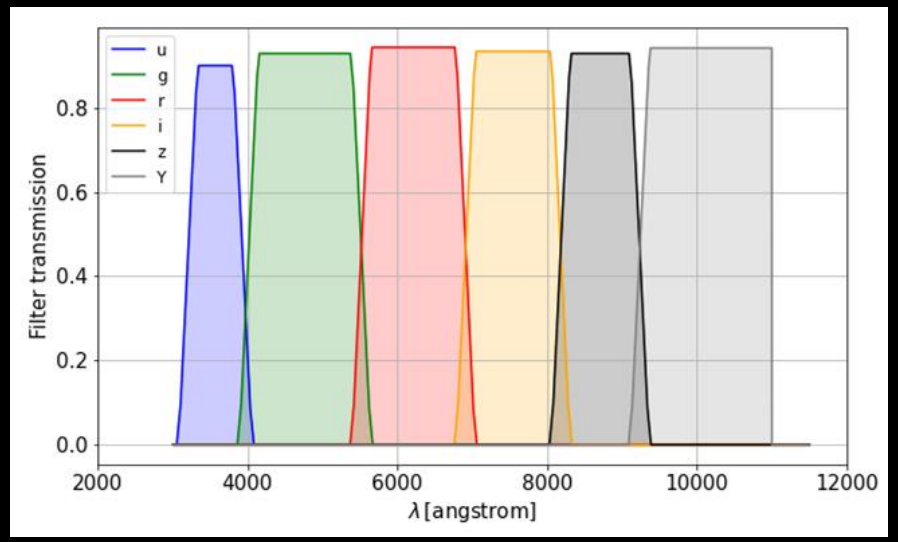


Photometric surveys and the Dark Energy Survey....

Photometric surveys

- Two main ways of surveys:
  - Spectroscopic (eBOSS, DESI....): great redshifts but "lower" statistics (time consuming)
  - Photometric (DES, LSST, Euclid, PAUS, J-PAS....): only photometric redshift estimation but huge statistics
- In both cases, we observe the flux emitted by a source: $F_\nu = \left[\frac{erg}{s \cdot cm^2 \cdot Hz}\right] = [Jy]$
- How bright is an object at a given λ? → Magnitudes: $m_{AB} = -2.5\log(F_\nu) - 48.6$
- In the case of photometry, we measure fluxes through filters:
  - Narrow band filters (PAUS, J-PAS)
  - Broad band filters (DES, LSST....)

Flux in # photons: $F^{obs}_{filter} \propto \int F_\nu(\lambda) \cdot S_{filter}(\lambda) \frac{d\lambda}{h\,\lambda} \rightarrow m^{obs}_{filter}$



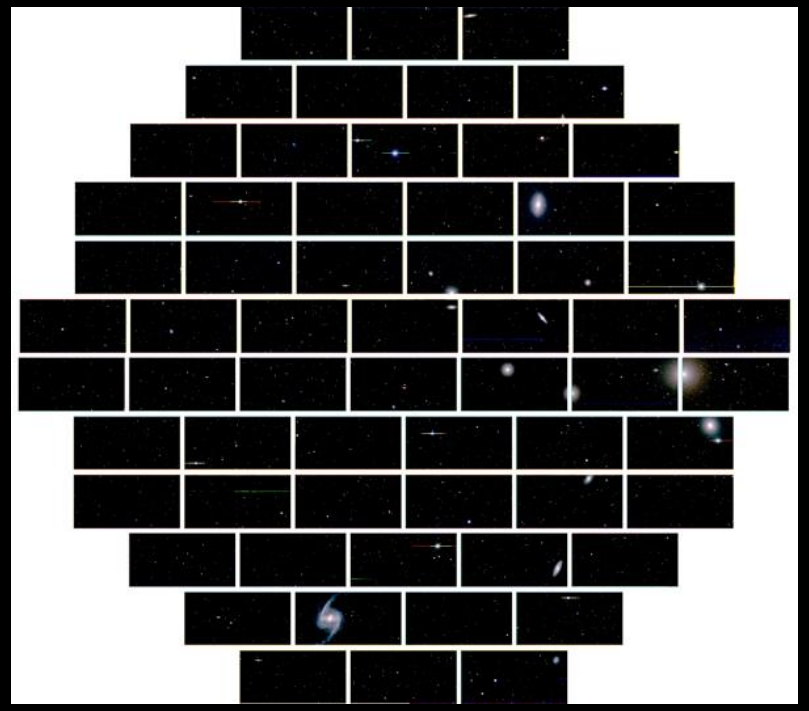
The Dark Energy Survey (DES)

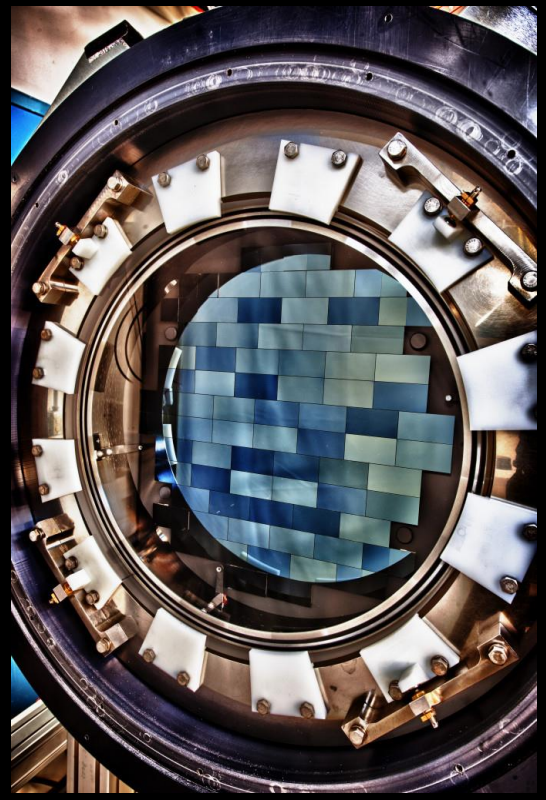
- Photometric galaxy survey
- Goal = tight constraints on the nature of dark energy
- Combination of techniques on the same experiment:
  - Expansion and geometry: type Ia SNe and BAO
  - Matter content and structure: galaxy clustering, weak lensing and counts of clusters
  - Combination with external data (CMB and low-redshift)
- Main cosmology results → $3 \times 2\text{pt} = \xi_{\pm}(\theta) + \underbrace{w(\theta) + \gamma_t(\theta)}_{2 \times 2\text{pt}}$

DECam and the Victor Blanco 4 m Telescope

Dark Energy Camera (DECam):

- 570 megapixel camera
- 74 CCDs in hexagonal pattern
- CCDs optimised for red and NIR
- FoV $\sim 3\ \mathrm{deg}^2$ (~ 14 full Moons)
- Five broad band photometric filters: *grizY*



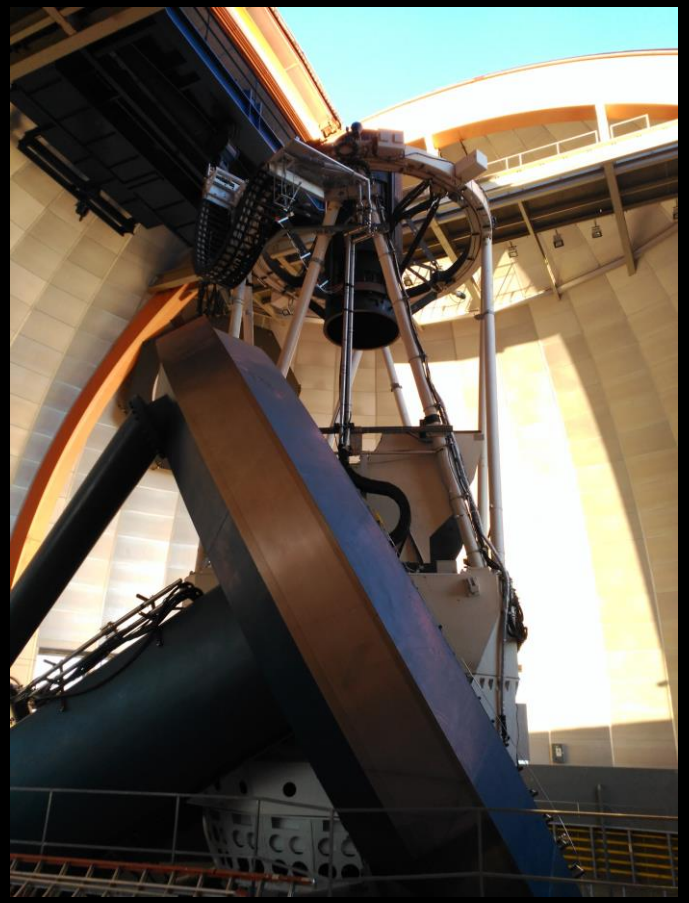


https://www.darkenergysurvey.org/the-des-project/instrument/

DECam and the Victor Blanco 4 m Telescope

Victor Blanco 4 m Telescope:

- Placed at Cerro Tololo Inter-American Observatory (CTIO), Chile
- 4 m primary mirror and equatorial mount
- DECam mounted at prime focus

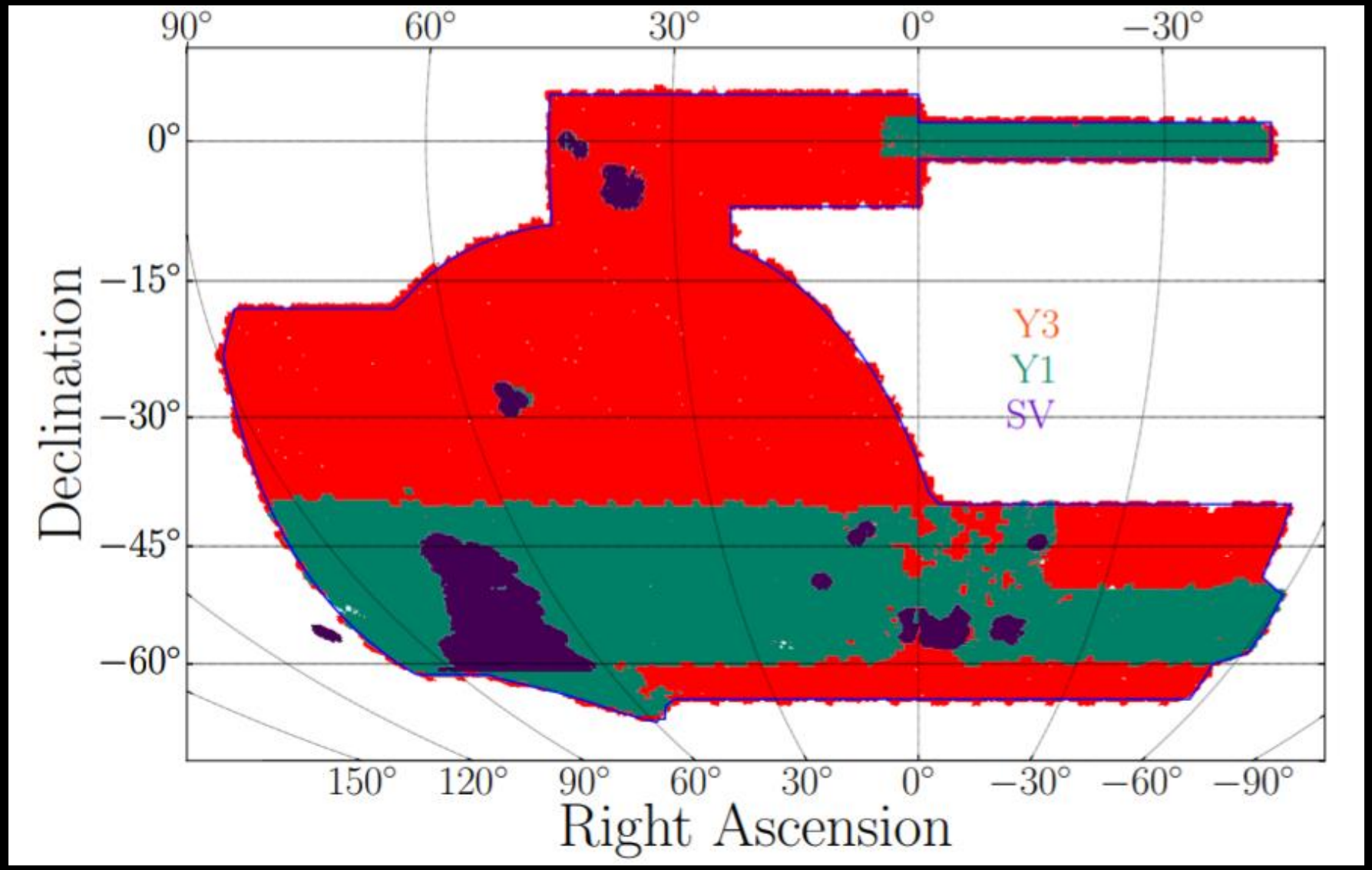


Survey strategy

Period	From	to	Area [deg^2]	Depth (i-band)	# objects
SV	Nov. 2012	Feb. 2013	~ 250	23.68	25M
Y1	Aug. 2013	Feb. 2014	1800	23.29	$\sim$ 137M
Y3	Aug. 2013	Feb. 2016	5000	23.44	$\sim$ 399M
Y6	Aug. 2013	Jan. 2019	5000	23.80	691M

- Wide-area survey covers 5000 deg^2 of the Southern Hemisphere
  - Footprint → avoid Milky Way plane + overlap with SPT and Stripe 82 (SDSS)
  - Each part of the footprint observed 10 times in each band: 90 s *griz* and 45 s *Y*
  - We use the first three years of data (DES-Y3) → $\sim$ 5000 deg^2 covered
- Deep survey (time domain):
  - 10 regions of the sky → 27 deg^2
  - Six-nights intervals in *griz* –bands
  - Greater depths and thousands of supernovae

Survey strategy



Sevilla-Noarbe et al. 2021

DES-Y3 data

The lens galaxy samples used in DES-Y3 are based on the Y3-Gold catalogue (Sevilla-Noarbe et al. 2021) → ~ 400 M objects over ~5000 deg^2 and depth $i = 23.0$

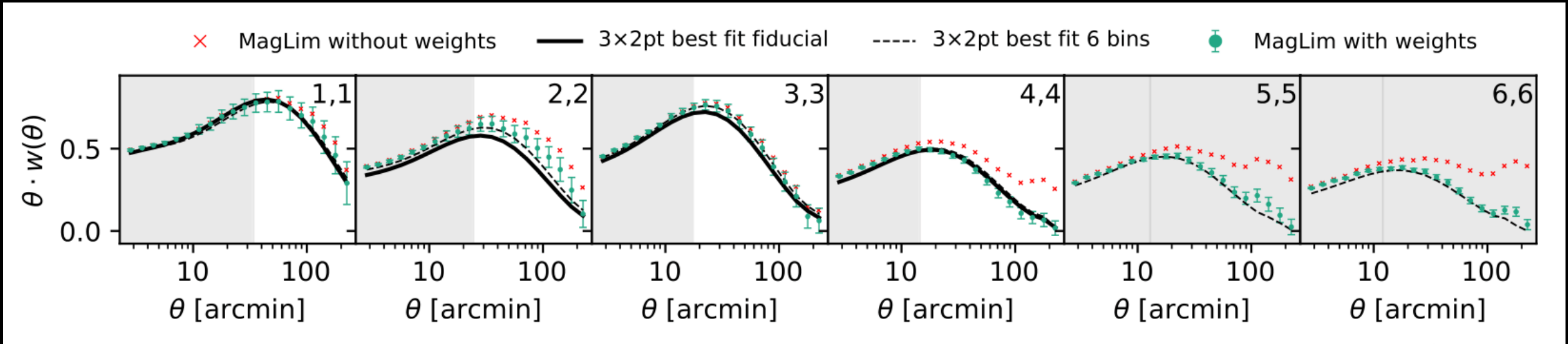
- MagLim sample → fiducial sample (Porredon et al. 2021):
  - Magnitude limited sample. Optimised for cosmological analyses
  - Photometric redshifts → Directional Neighbourhood Fitting (DNF) photo-z code
  - Selection:
    - $i < 4 \cdot z_{photo} + 18$
    - $i > 17.5$
  - $z_{photo} = [0.20, 0.40, 0.55, 0.70, 0.85, 0.95, 1.05]$
  - ~ 11 M galaxies
- redMaGiC sample → secondary sample (Rozo, Rykoff et al. 2016):
  - Selection of LRGs
  - High-quality photo-z → redMaGiC algorithm
  - $z_{photo} = [0.15, 0.35, 0.50, 0.65, 0.80, 0.90]$
  - ~ 3 M galaxies

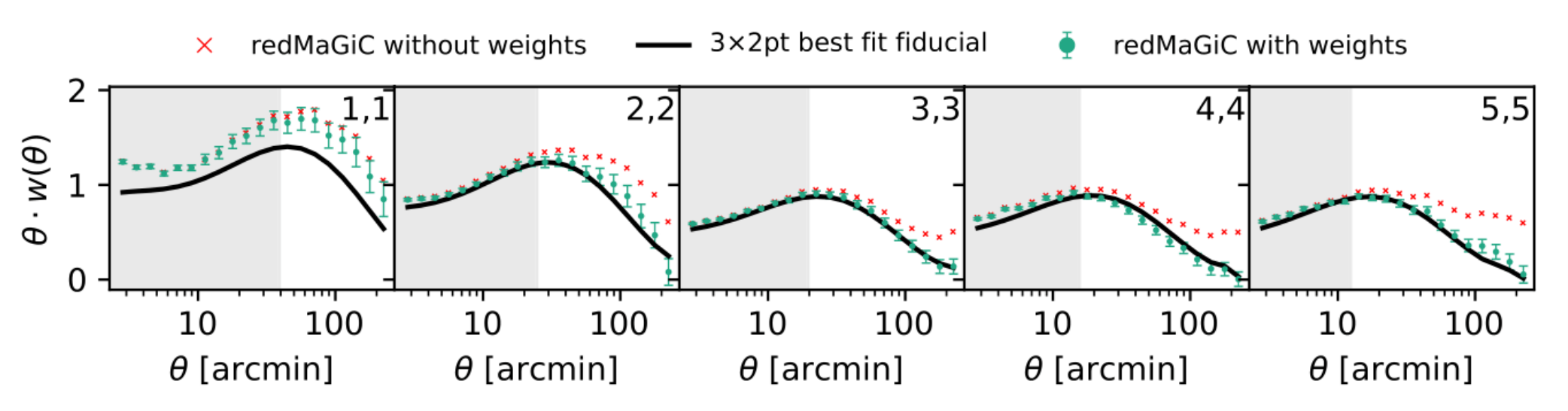
DES-Y3 results on galaxy clustering and cosmological parameters....

DES-Y3 galaxy clustering and 3×2pt

- $w(\theta) \rightarrow 3 \times 2\text{pt} \Rightarrow \{w(\theta), \gamma_t(\theta), \xi_{\pm}(\theta)\}$ and $2 \times 2\text{pt} \Rightarrow \{w(\theta), \gamma_t(\theta)\}$
- Covariance for 3×2pt and 2×2pt includes the systematic terms derived from the clustering analysis
- Two cosmological models are fitted:
  - Flat ΛCDM $\rightarrow \Omega_m, \Omega_b, \Omega_\nu h^2, h, A_s, n_s$
  - wCDM $\rightarrow \Omega_m, \Omega_b, \Omega_\nu h^2, h, A_s, n_s, w$
  - 25 nuisance parameters
- DES-Y3 fiducial results $\Rightarrow$ MagLim first 4 bins
- Additional results with redMaGiC as a robustness check
- Small angular scales excluded to avoid non-linear effects

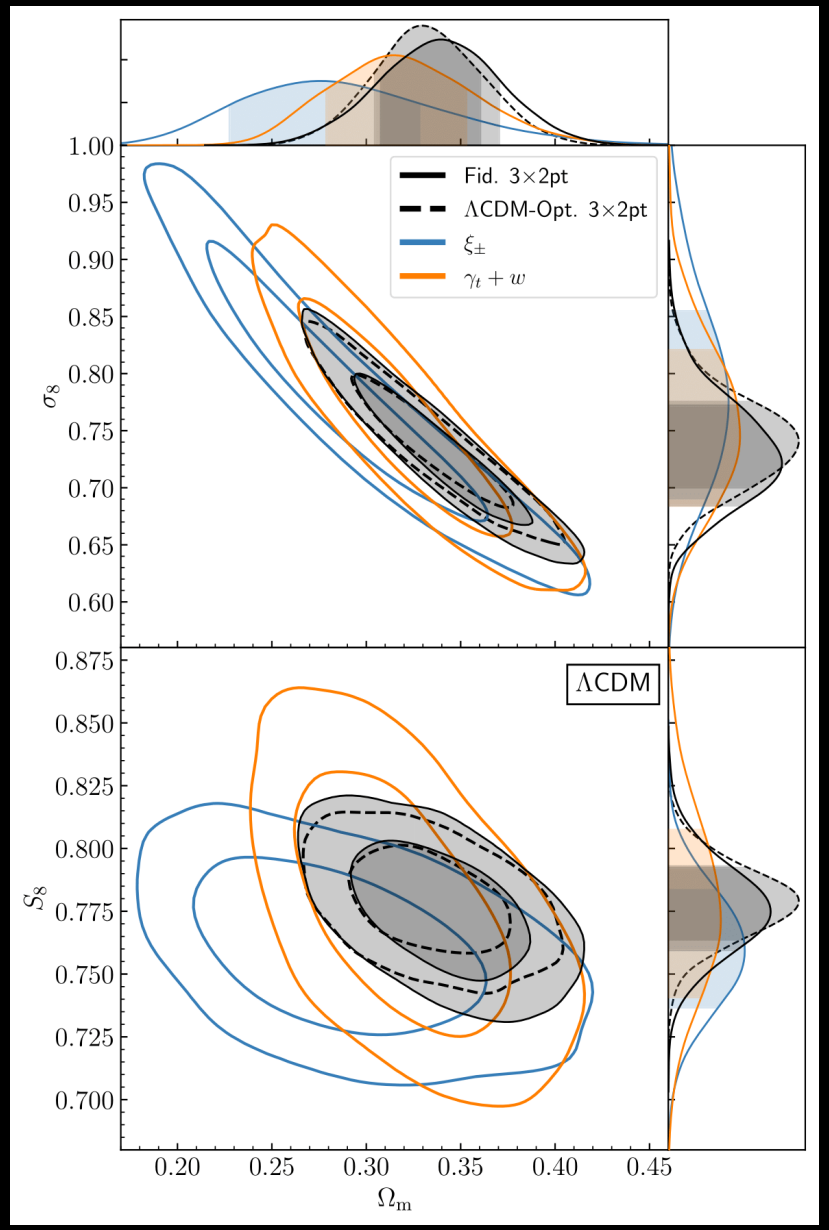
DES-Y3 results: galaxy clustering





Rodríguez-Monroy et al. 2021

DES-Y3 results: 3 × 2pt



- DES-Y3 fiducial ΛCDM cosmology results (mean posterior ±68% C.L.):

$$\Omega_m = 0.339^{+0.032}_{-0.031}$$

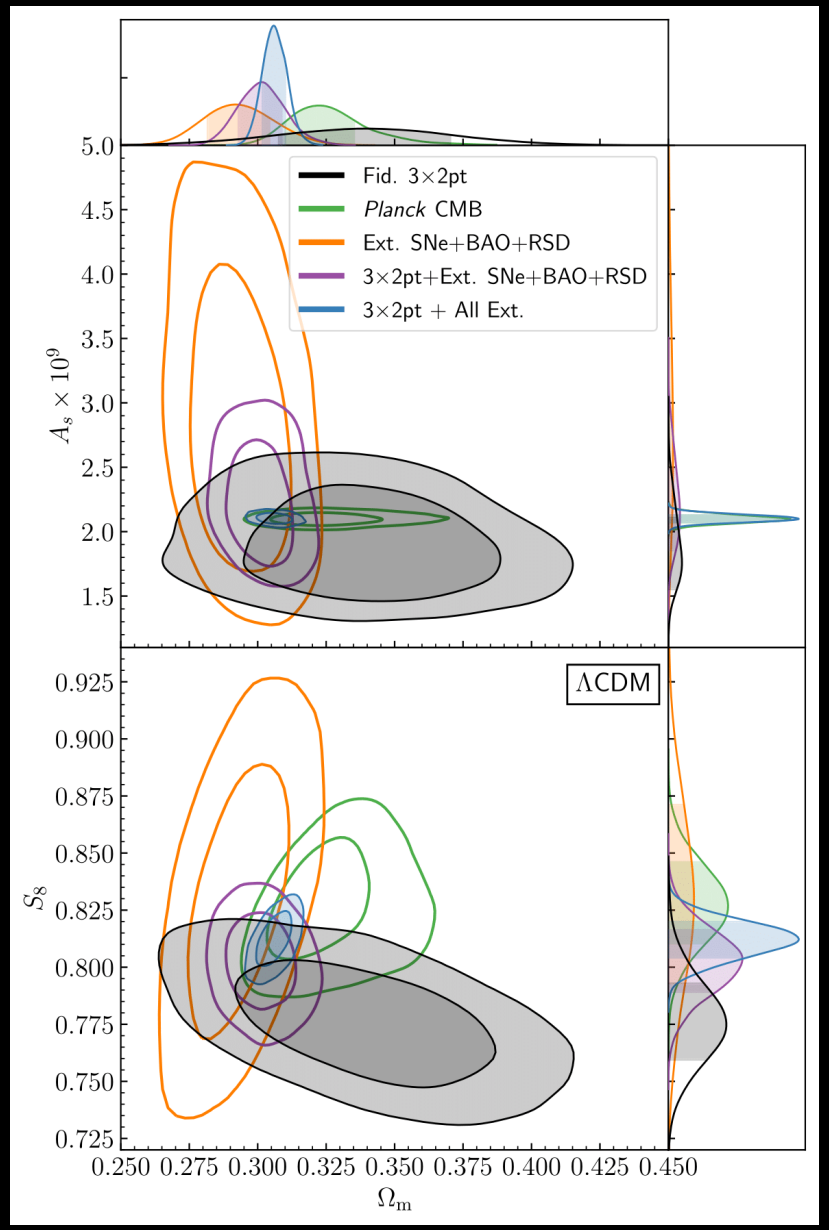
$$\sigma_8 = 0.733^{+0.039}_{-0.049}$$

$$S_8 = 0.776^{+0.017}_{-0.017}$$

- Compatibility of 2×2pt and $\xi_\pm(\theta)$ probes ensured before combining
- $S/N = 83 \Rightarrow$ factor 2.1 improvement w.r.t. Y1
- Goodness-of-fit PPD p-value = 0.04

DES Collaboration et al. 2021

DES-Y3 results: 3 × 2pt



- DES-Y3 combined analysis with external data (Planck+SNe+BAO+RSD) in ΛCDM:

$$\Omega_m = 0.306^{+0.004}_{-0.005}$$

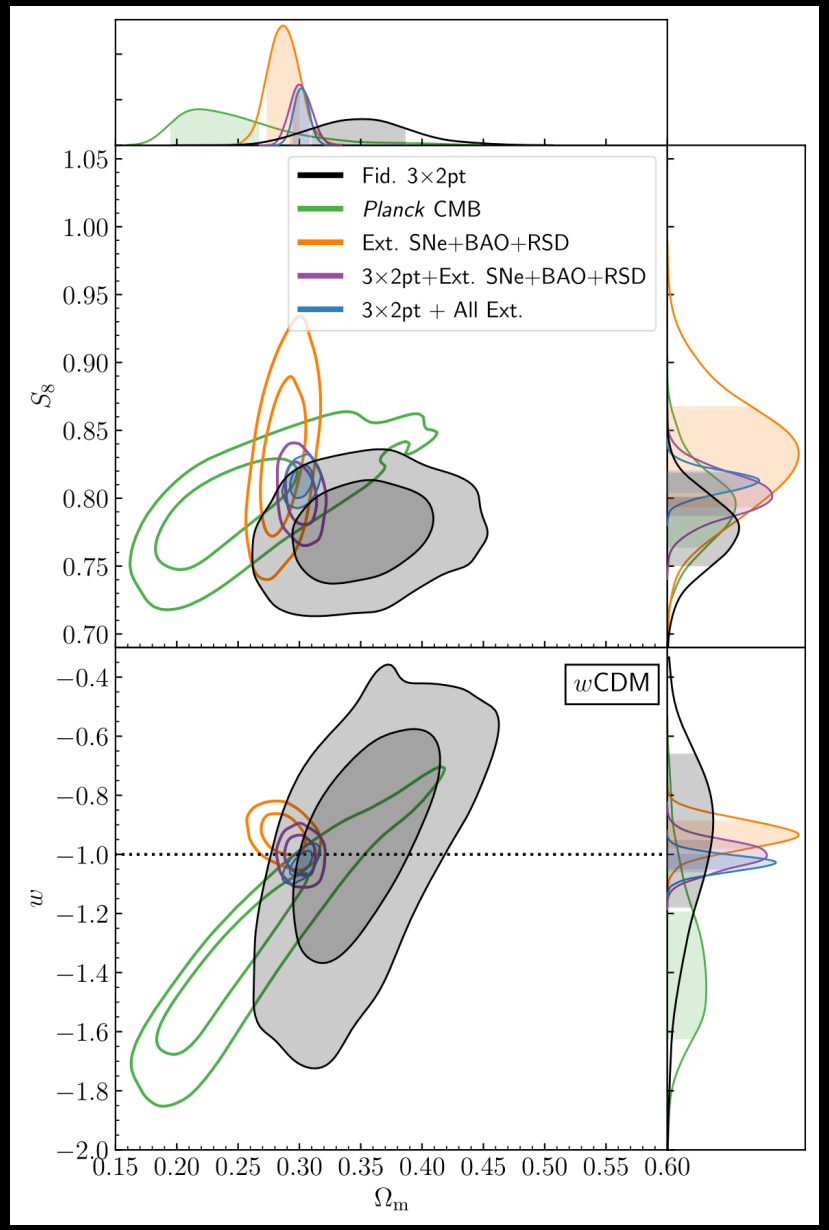
$$\sigma_8 = 0.804^{+0.008}_{-0.008}$$

$$S_8 = 0.812^{+0.008}_{-0.008}$$

- Compatibility of the different data sets ensured → $1.5\,\sigma$ difference DES-Y3 – Planck
- Most precise constraints to date on these parameters → $1\% - 3\%$ in all cosmological parameters

DES Collaboration et al. 2021

DES-Y3 results: 3 × 2pt



- DES-Y3 combined analysis with external data (Planck+SNe+BAO+RSD) in wCDM:

$$\Omega_m = 0.302^{+0.006}_{-0.006}$$

$$w = -1.031^{+0.030}_{-0.027}$$

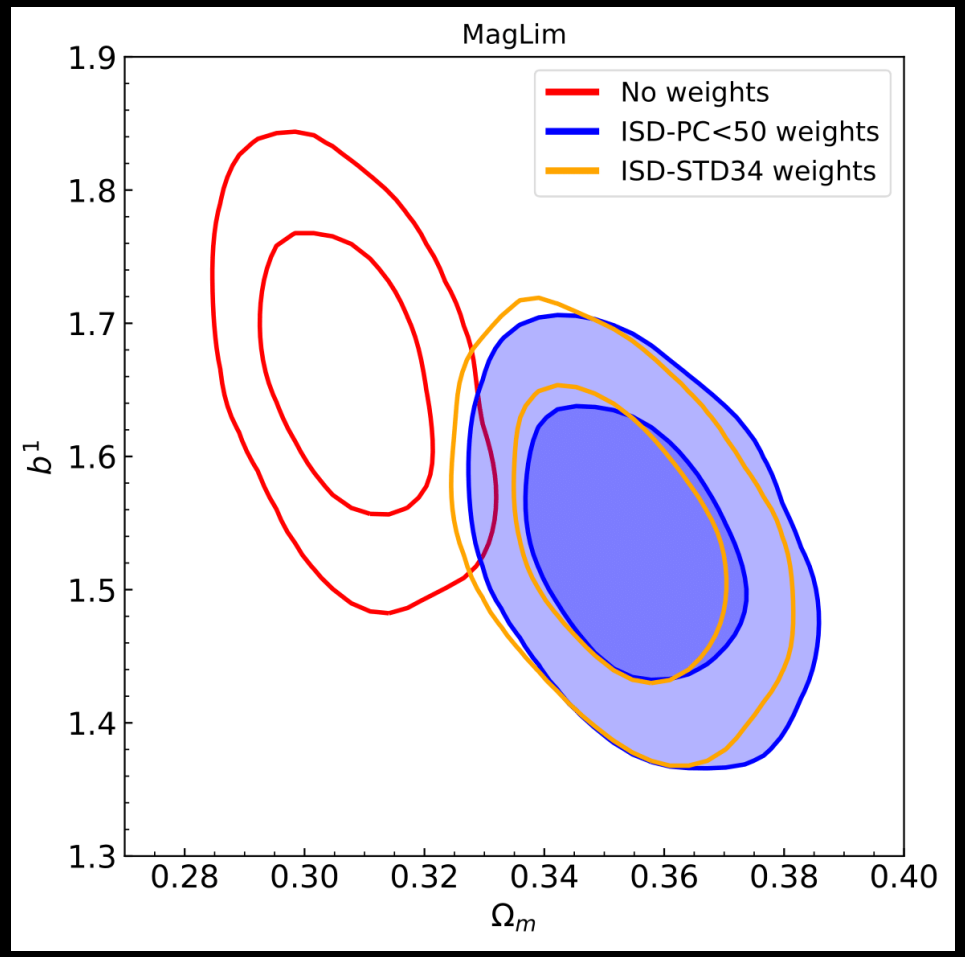
$$S_8 = 0.812^{+0.008}_{-0.008}$$

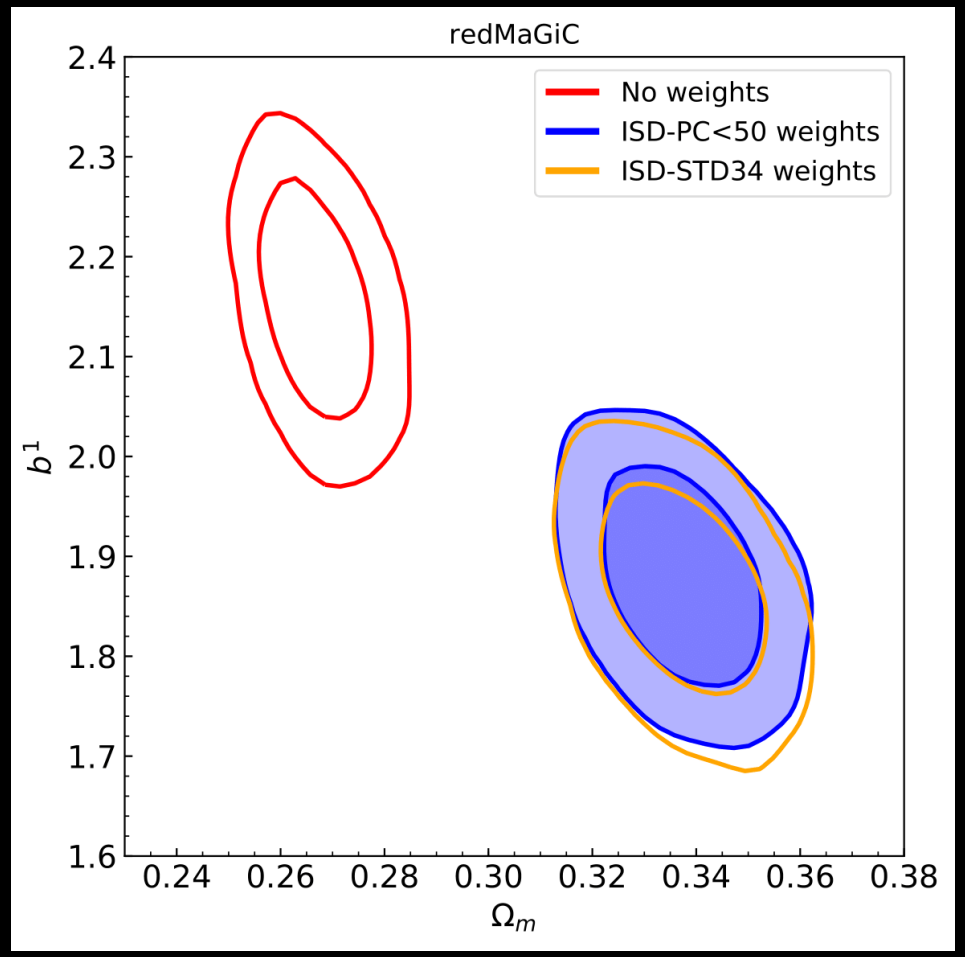
- Bayes factor, $R = 7.8 \Rightarrow$ full joint data analysis has no preference for wCDM over ΛCDM
- This is the most powerful and precise test for the standard cosmological model to date

DES Collaboration et al. 2021

Impact of observational systematics

- Observing conditions can bias cosmological results remarkably





Systematic effects on galaxy clustering....

Systematic effects on galaxy clustering

- Systematic effects in cosmology surveys = non-cosmological fluctuations in n_{gal} and the galaxy properties
- Main sources of systematic uncertainty in galaxy clustering with DES data:
  - Photometric redshift estimation
    - The width, shape and mean of $n_g^i(z)$ can alter the clustering signal
    - We marginalise over bias, Δz^i, and width, σ_z^i, uncertainties on $n_g^i(z)$:

$$n_g^i(z) \rightarrow n_g^i\left(z - \Delta z^i\right)$$
$$n_g^i(z) \rightarrow n_g^i\left(\sigma_z^i\left[z - \langle z \rangle\right] + \langle z \rangle\right)$$

  - Observational systematics

Correcting for observing conditions

- Several ways of dealing with contamination:
  - Act at the $w(\theta)$ (or other statistics / estimators) level
  - Act at the map level (density field, shear field, etc)
  - Modify the randoms used by the estimators
- Many methods employ template maps of potential contaminants
  - Associated risks:
    - Using all maps can lead to overcorrection
    - Manually selecting them can exclude contaminants, leading to undercorrection
  - Intermediate solution:
    - Study correlation between template maps
    - Data-driven selection → take into account the preferences of the data
  - DES-Y3 methods: *ISD, ENet* (and *NN-weights*)

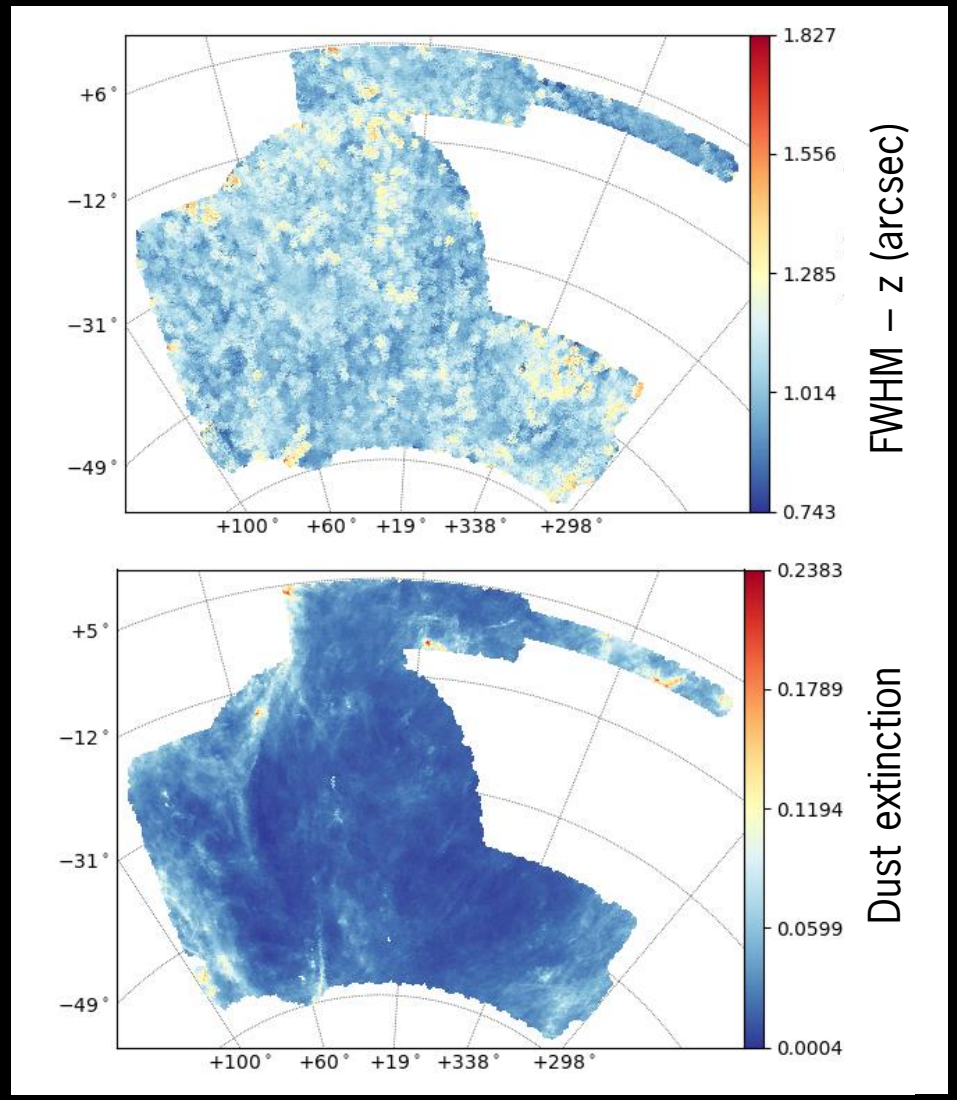
Observational systematics

- Three important sources of systematic:
  - Observing conditions
    - Examples: seeing, airmass, sky brightness, error on the sky brightness
  - Survey properties
    - Examples: exposure time, survey depth, photometric calibrations
  - Astrophysical foregrounds
    - Stellar density:
      - Stars identified as galaxies
      - Bright stars causing obscuration
    - Dust extinction

SP map	Units	Statistics
airmass	$\emptyset$	WMEAN, MIN, MAX
fwhm	arcsec	WMEAN, MIN, MAX
fwhm_fluxrad	arcsec	WMEAN, MIN, MAX
exptime	seconds	SUM
t_eff	$\emptyset$	WMEAN, MIN, MAX
t_eff_exptime	seconds	SUM
skybrite	electrons/CCD pixel	WMEAN
skyvar	(electros/CCD pixel)^2	WMEAN, MIN, MAX
skyvar_sqrt	electrons/CCD pixel	WMEAN
skyvar_uncert	electrons/ s $\cdot$ coadd pixel	
sigma_mag_zero	mag	QSUM
fgcm_gry	mag	WMEAN, MIN
maglim	mag	
sof_depth	mag	
magauto_depth	mag	
stars_1620	# stars	
stellar_dens	# stars/deg^2	
sfd98	mag	

Observational systematics

- Three important sources of systematic:
  - Observing conditions
  - Survey properties
  - Astrophysical foregrounds
- Characterisation of systematic effects:
  - Solution → Survey property (SP) maps:
    - HEALPix maps
    - Track spatial variations of a statistic of the imaging conditions across the sky
    - We work at $N_{side} = 4096$ and 512, after applying the angular mask



Survey property maps and correlations

- Stacking of images ⇒ we need to use a summary statistic. Several statistics available for some quantities → 26 quantities
- We have 4 photometric bands: *griz* (not using Y - band)
- 3 astrophysical foreground maps considered
- In total: $26 \times 4 + 3 = 107$ SP maps available for DES-Y3
- However, many of these maps are correlated
- We wish to reduce the number of SP maps
  - Optimise the decontamination process
  - Using too many maps can lead to overcorrection

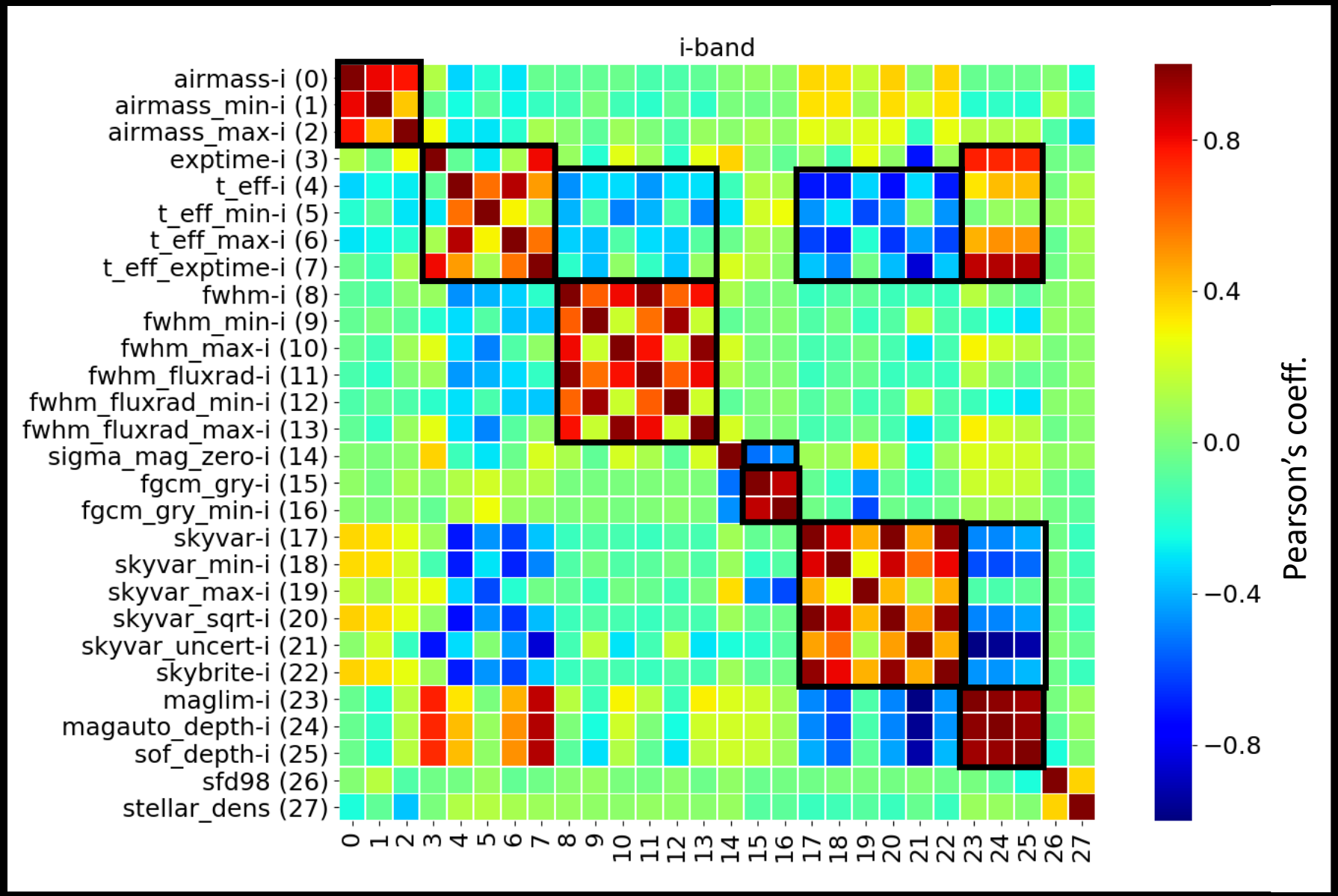
Survey property maps and correlations

- Stacking of images ⇒ we need to use a summary statistic. Several statistics available for some quantities → 26 quantities
- We have 4 photometric bands: *griz* (not using Y - band)
- 3 astrophysical foreground maps considered
- In total: $26 \times 4 + 3 = 107$ SP maps available for DES-Y3
- However, many of these maps are correlated
- We wish to reduce the number of SP maps
- We use two different SP map bases
  - Standard (STD) basis → STD maps (original maps)
  - Principal component (PC) basis → PC maps
- We perform different dimensionality reductions in these two SP map bases

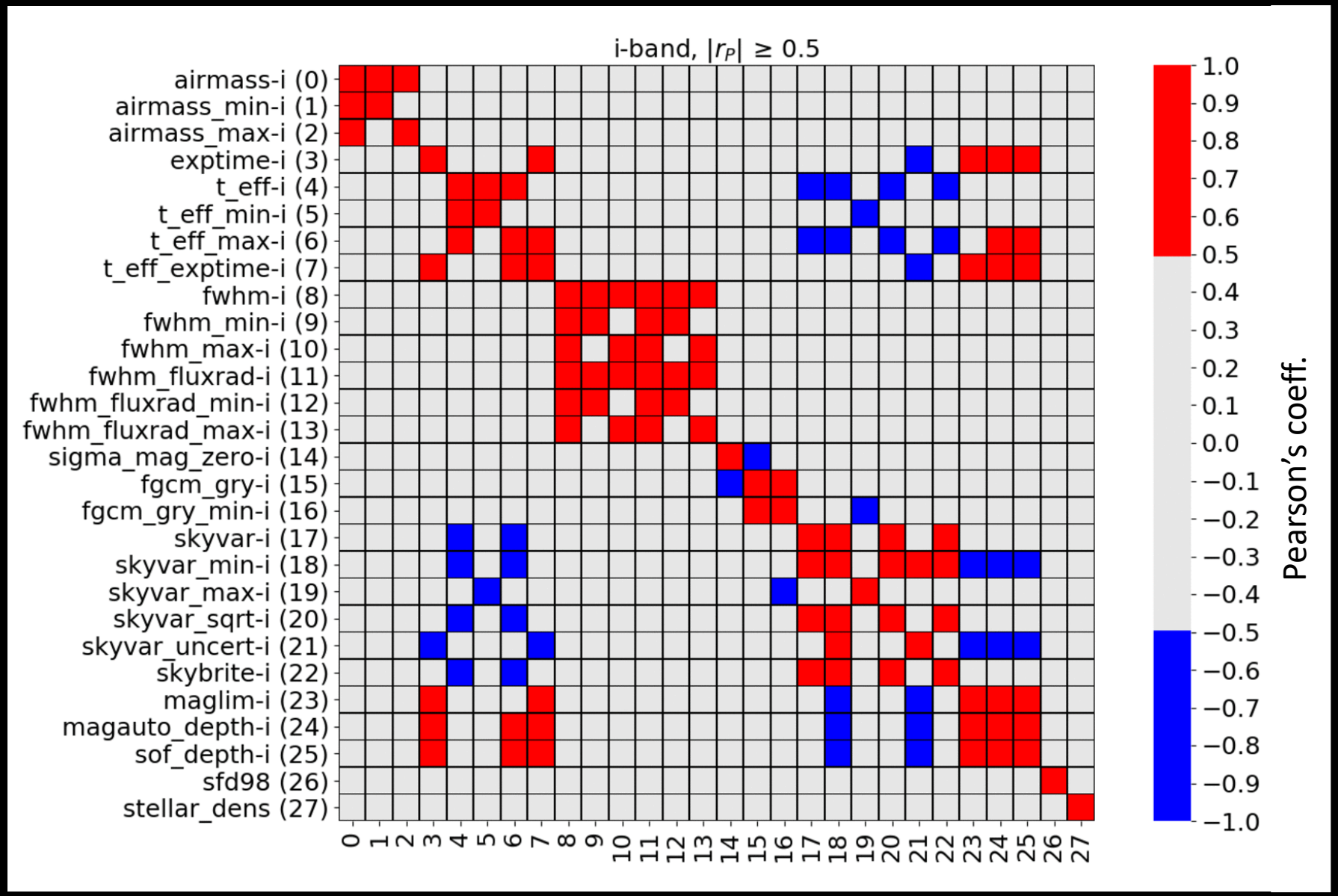
STD maps

- STD maps = original SP maps
- How can we reduce their number?
  - We can automatically exclude some STD maps (*maglim*, *magauto_depth* and *stars_1620*)
  - Idea → identify SP map "families"
  - STD maps corresponding to the same physical magnitude form families
  - To define these families and to identify further correlations, we use the Pearson's correlation coefficient

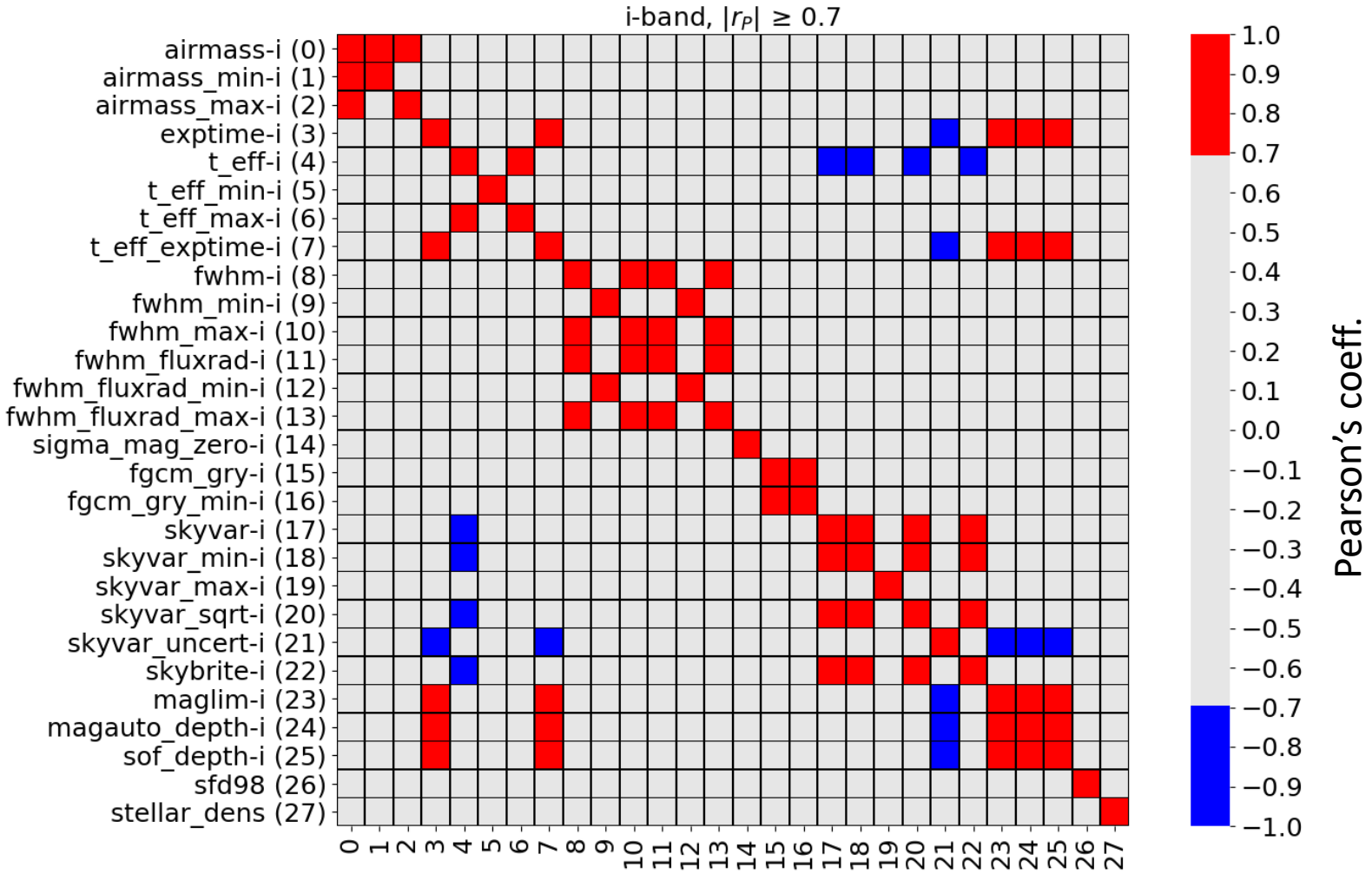
STD maps


i-band
airmass-i (0)
airmass_min-i (1)
airmass_max-i (2)
exptime-i (3)
t_eff-i (4)
t_eff_min-i (5)
t_eff_max-i (6)
t_eff_exptime-i (7)
fwhm-i (8)
fwhm_min-i (9)
fwhm_max-i (10)
fwhm_fluxrad-i (11)
fwhm_fluxrad_min-i (12)
fwhm_fluxrad_max-i (13)
sigma_mag_zero-i (14)
fgcm_gry-i (15)
fgcm_gry_min-i (16)
skyvar-i (17)
skyvar_min-i (18)
skyvar_max-i (19)
skyvar_sqrt-i (20)
skyvar_uncert-i (21)
skybrite-i (22)
maglim-i (23)
magauto_depth-i (24)
sof_depth-i (25)
sfd98 (26)
stellar_dens (27)
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27
0.8
0.4
0.0
−0.4
−0.8
Pearson's coeff.

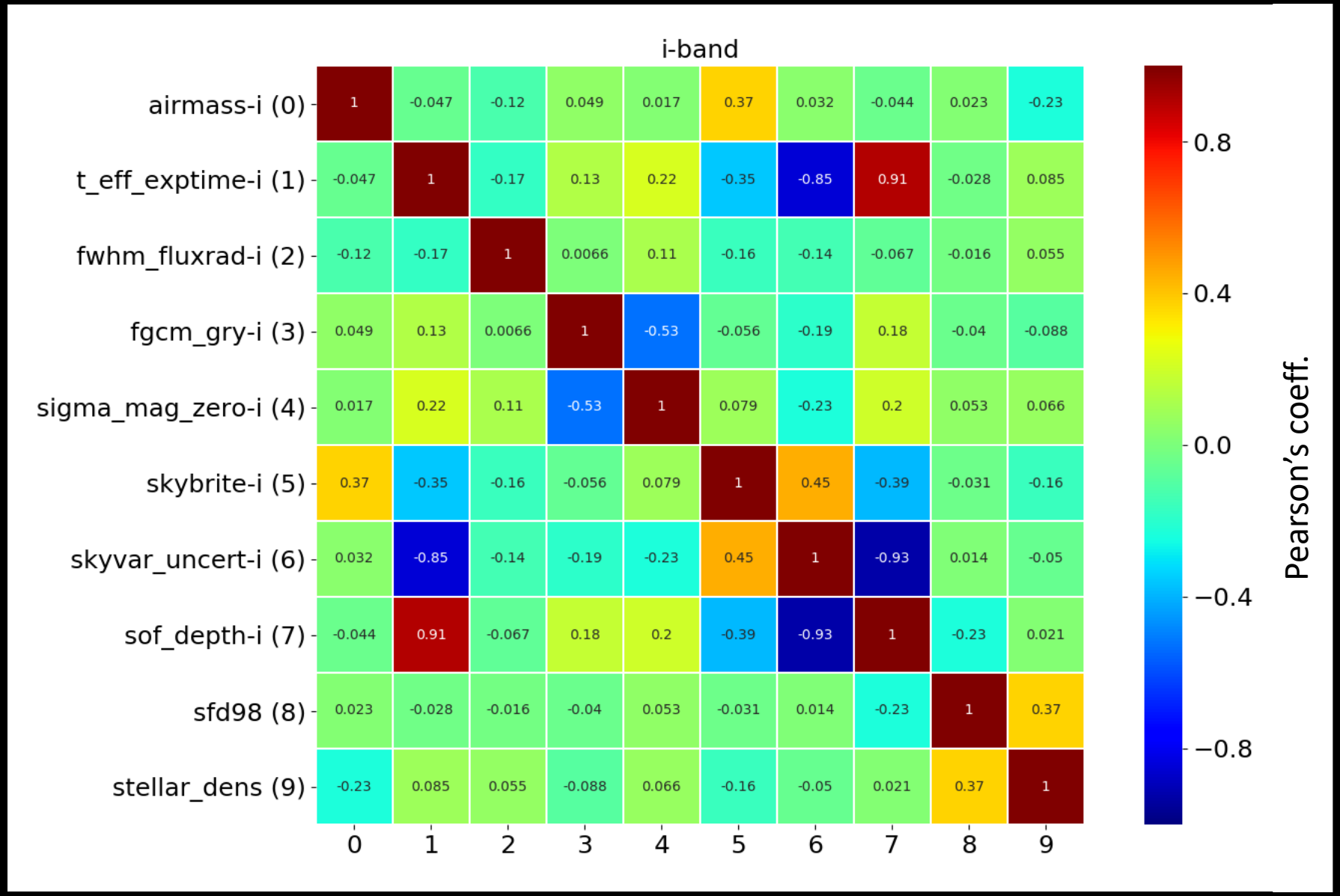
STD maps


i-band, $|r_P| \geq 0.5$
airmass-i (0)
airmass_min-i (1)
airmass_max-i (2)
exptime-i (3)
t_eff-i (4)
t_eff_min-i (5)
t_eff_max-i (6)
t_eff_exptime-i (7)
fwhm-i (8)
fwhm_min-i (9)
fwhm_max-i (10)
fwhm_fluxrad-i (11)
fwhm_fluxrad_min-i (12)
fwhm_fluxrad_max-i (13)
sigma_mag_zero-i (14)
fgcm_gry-i (15)
fgcm_gry_min-i (16)
skyvar-i (17)
skyvar_min-i (18)
skyvar_max-i (19)
skyvar_sqrt-i (20)
skyvar_uncert-i (21)
skybrite-i (22)
maglim-i (23)
magauto_depth-i (24)
sof_depth-i (25)
sfd98 (26)
stellar_dens (27)
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27
1.0 0.9 0.8 0.7 0.6 0.5 0.4 0.3 0.2 0.1 0.0 −0.1 −0.2 −0.3 −0.4 −0.5 −0.6 −0.7 −0.8 −0.9 −1.0
Pearson's coeff.

STD maps

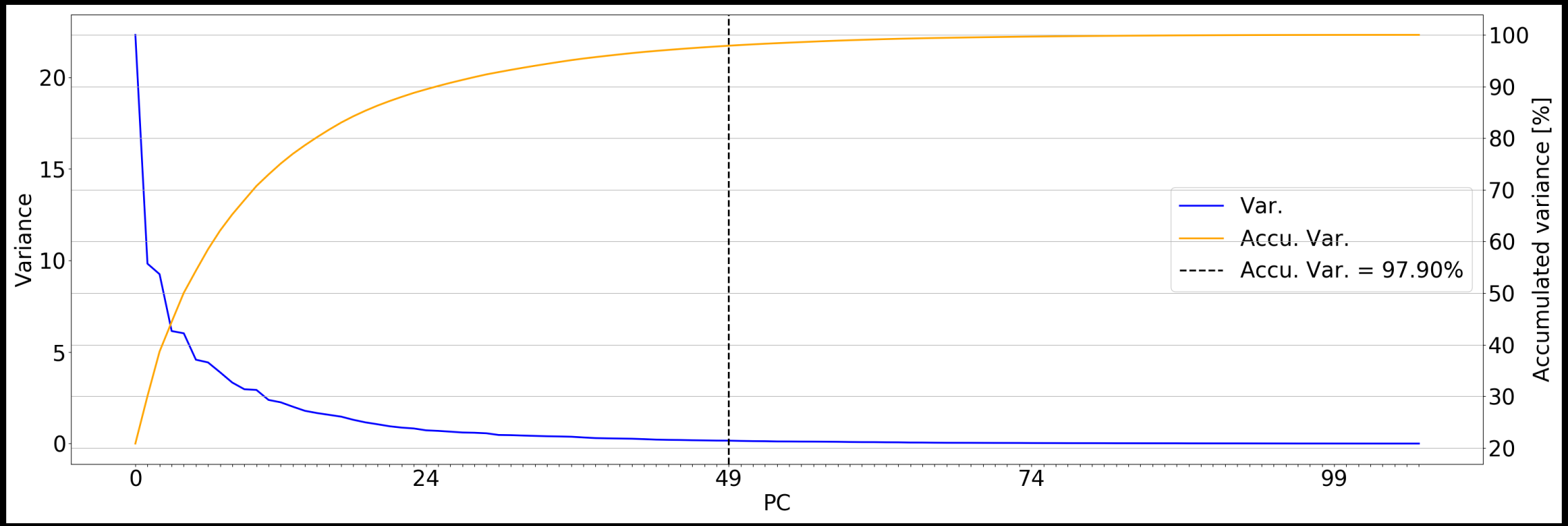

i-band, $|r_P| \geq 0.7$
airmass-i (0)
airmass_min-i (1)
airmass_max-i (2)
exptime-i (3)
t_eff-i (4)
t_eff_min-i (5)
t_eff_max-i (6)
t_eff_exptime-i (7)
fwhm-i (8)
fwhm_min-i (9)
fwhm_max-i (10)
fwhm_fluxrad-i (11)
fwhm_fluxrad_min-i (12)
fwhm_fluxrad_max-i (13)
sigma_mag_zero-i (14)
fgcm_gry-i (15)
fgcm_gry_min-i (16)
skyvar-i (17)
skyvar_min-i (18)
skyvar_max-i (19)
skyvar_sqrt-i (20)
skyvar_uncert-i (21)
skybrite-i (22)
maglim-i (23)
magauto_depth-i (24)
sof_depth-i (25)
sfd98 (26)
stellar_dens (27)
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27
1.0 0.9 0.8 0.7 0.6 0.5 0.4 0.3 0.2 0.1 0.0 −0.1 −0.2 −0.3 −0.4 −0.5 −0.6 −0.7 −0.8 −0.9 −1.0
Pearson's coeff.

Representative STD maps: *STD34*



PC maps selection: *PC<50*

- We use the 107 STD maps ⇒ 107 PC maps
- We verify that there are no significant diversions from linearity
- Now it is possible to reduce the number of PC maps



PC maps selection: *PC<50*

- We use the 107 STD maps ⇒ 107 PC maps
- We verify that there are no significant diversions from linearity
- Now it is possible to reduce the number of PC maps
- We retain the 50 first PC maps → $\sim 98\%$ accumulated variance
- For DES-Y3 galaxy clustering and 3×2pt analysis:
  - PC<50 is the fiducial set of SP maps for systematics mitigation
  - STD34 maps are also used as a cross-check and as robustness test
  - Robustness of both SP map bases and selections is later tested on simulations and data

Methods for observational systematics mitigation....

Systematics mitigation methods

For the DES-Y3 galaxy clustering analysis we have considered three different methods for observational systematics mitigation:

- Iterative Systematics Decontamination, *ISD* (Elvin-Poole et al. 2017, Rodríguez-Monroy et al. 2021):
  - Fiducial method of DES-Y3 analysis
  - Thoroughly validated
- Elastic Net regularisation, *ENet* (Weaverdyck et al. 2020):
  - Alternative method for systematics decontamination
  - Employed for validation and systematic uncertainty estimation
  - Thoroughly validated (with Y1 data)
- Neural net weights, *NN-weights* (Rezaie et al. 2020):
  - Method employed as an additional cross-check
  - Less validated yet

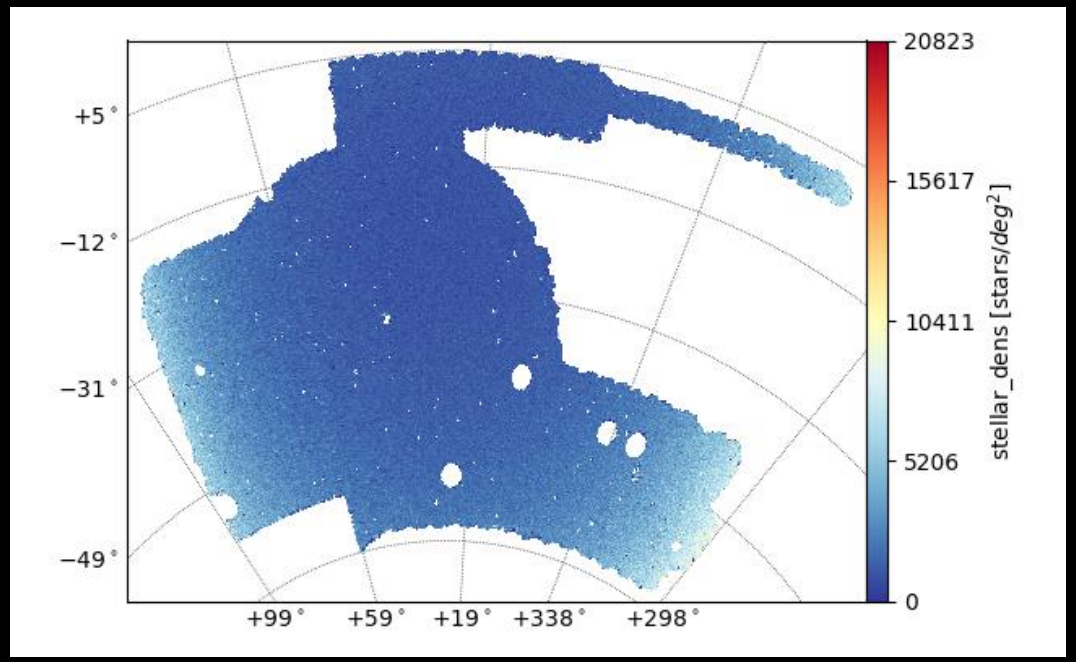
Iterative Systematics Decontamination, *ISD*

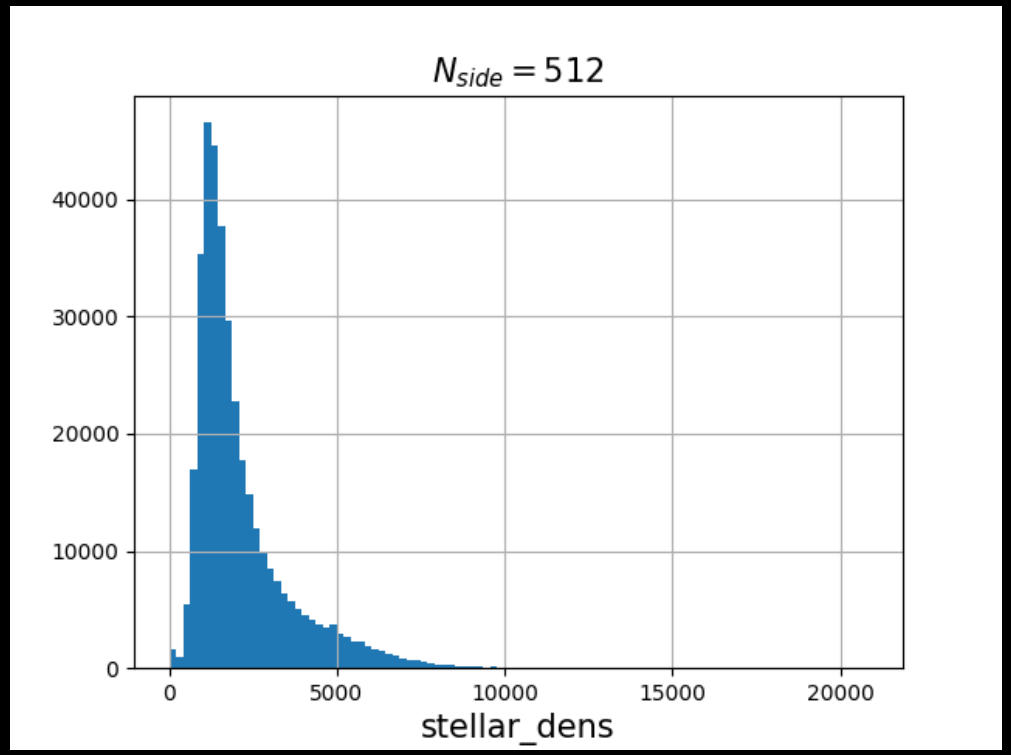
- DES-Y3 fiducial method for systematics decontamination → ISD
- Organised as iterative pipeline
- Evaluates correlation between observed galaxy number density and SP map values
- Three main inputs:
  - Galaxy sample to be decontaminated
  - List of SP maps
  - Set of mock galaxy catalogues: log-normal mocks
- Log-normal mocks:
  - Chance correlations can result in overcorrection
  - Solution → compute the correlation between SP maps and simulated realisations of the same power spectrum
  - Also used for validation tests

ISD: iterative pipeline

- ISD identifies the most contaminant SP map and corrects for it step-wise
- Steps:
  1. Identify the SP map that causes the most significant impact
     - Evaluate galaxy number density, n_{gal}, as function of SP map values
     - We call this relation *1D relation*. For a given SP map, s_1:
       I. Bin its values. We employ an equal area binning scheme
       II. Identify the pixels, p, on the sky such that $s_1^p \in bin_{SP}$
       III. Calculate average number density, $\bar{n}_{gal}^i$, on those regions

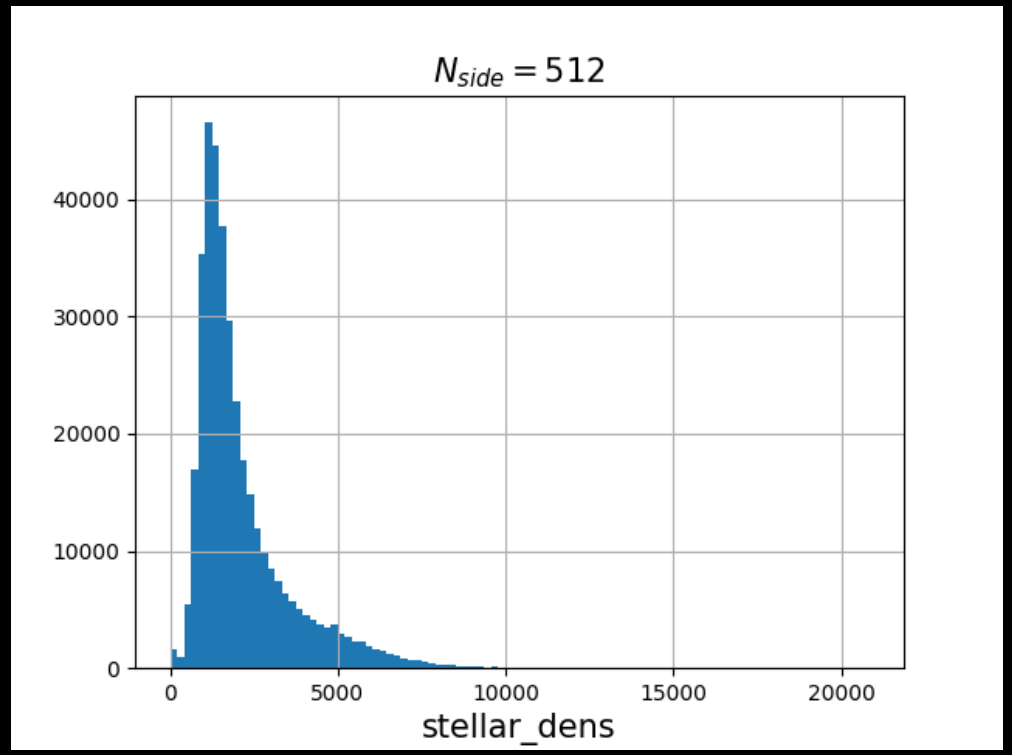
1D relation visually

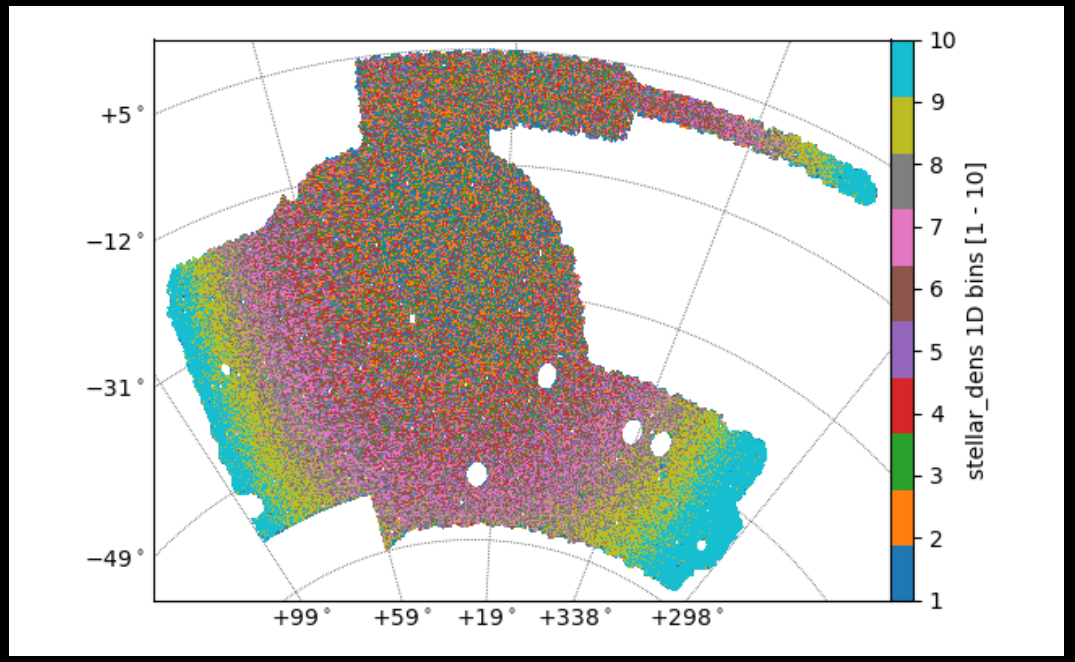




SP map → Value distribution

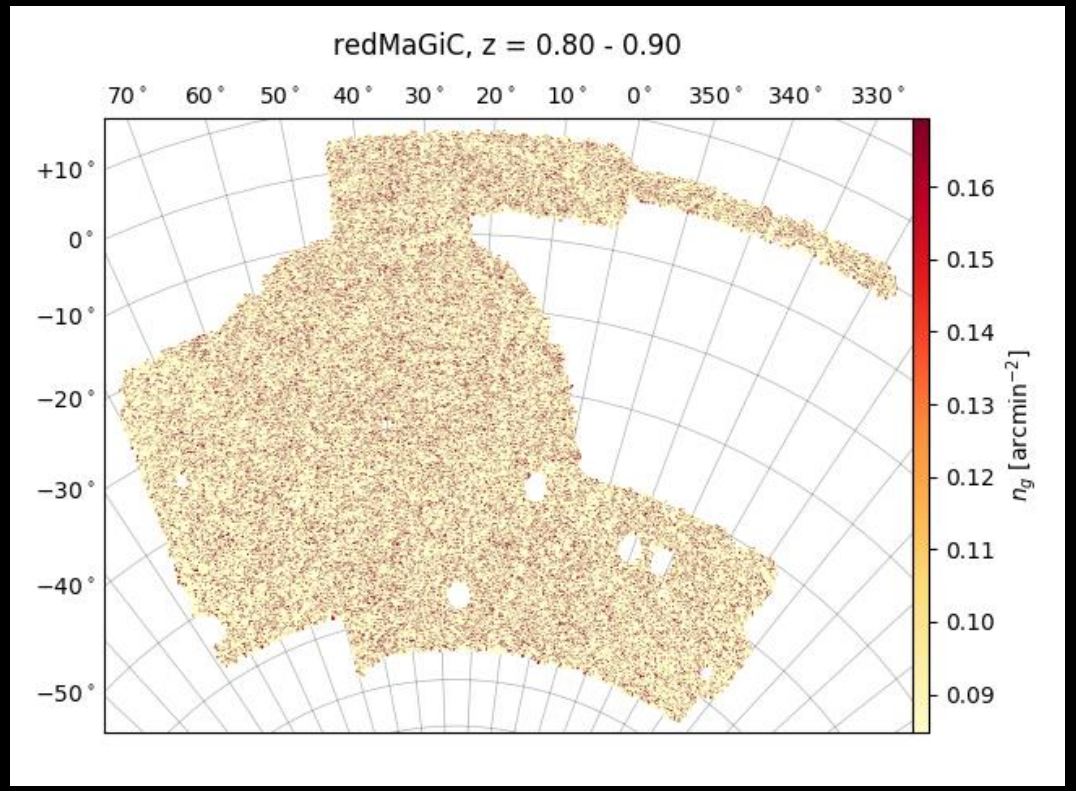
1D relation visually

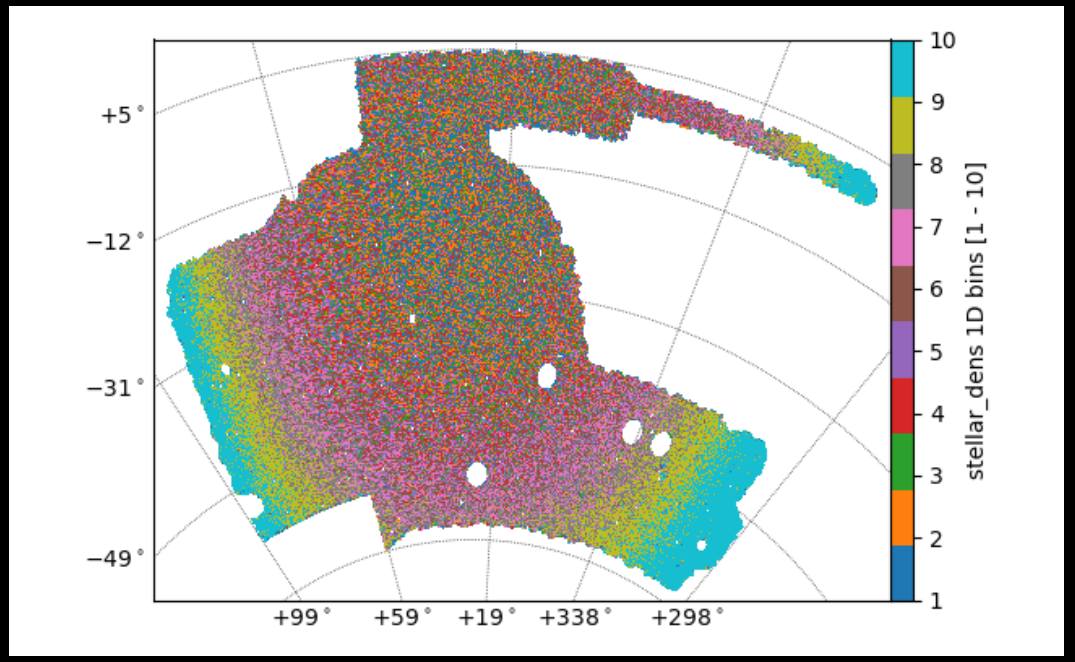




Value distribution → Equal area SP bins on the sky

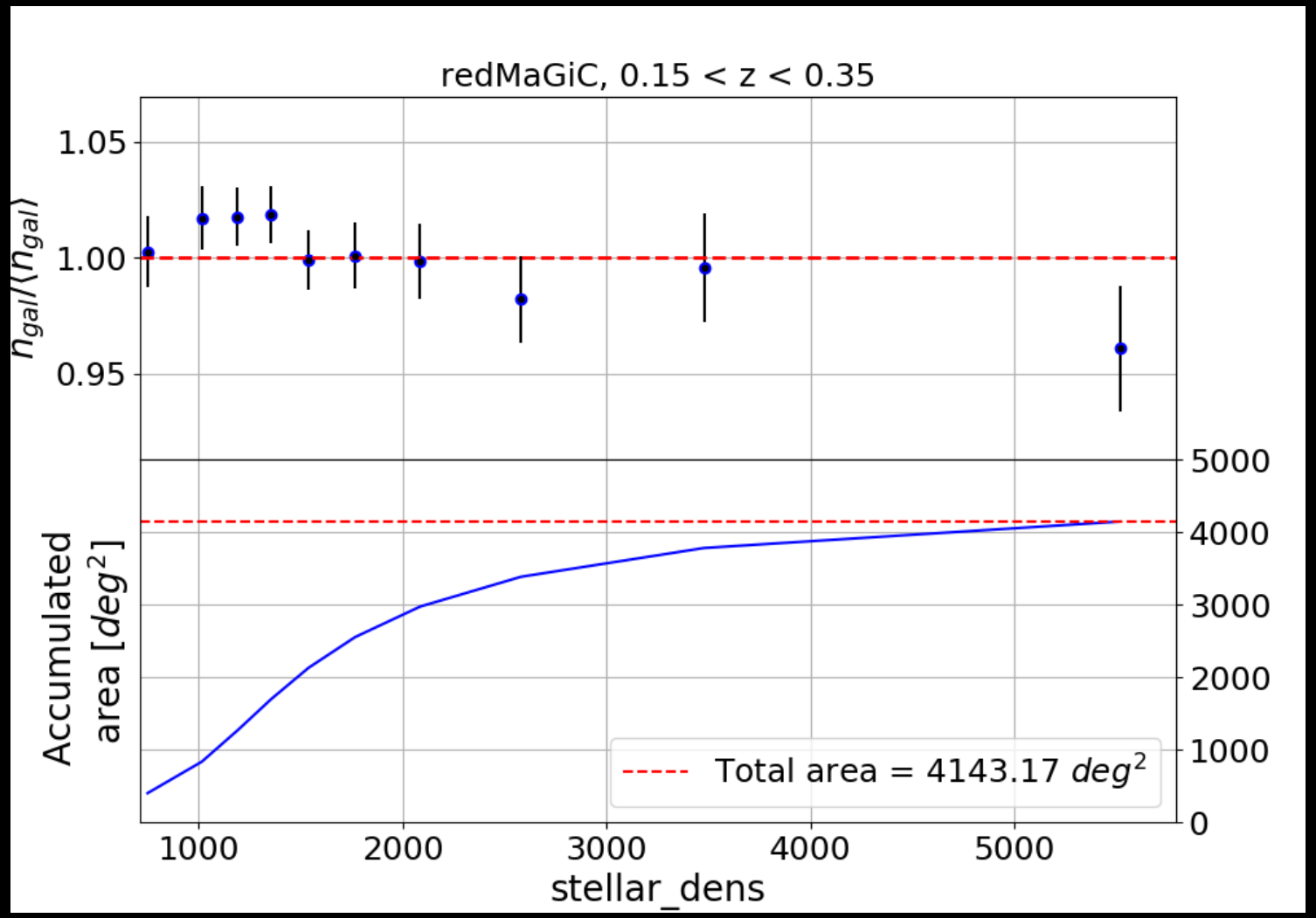
1D relation visually





Galaxy number density ⟷ Equal area SP bins on the sky

1D relation visually



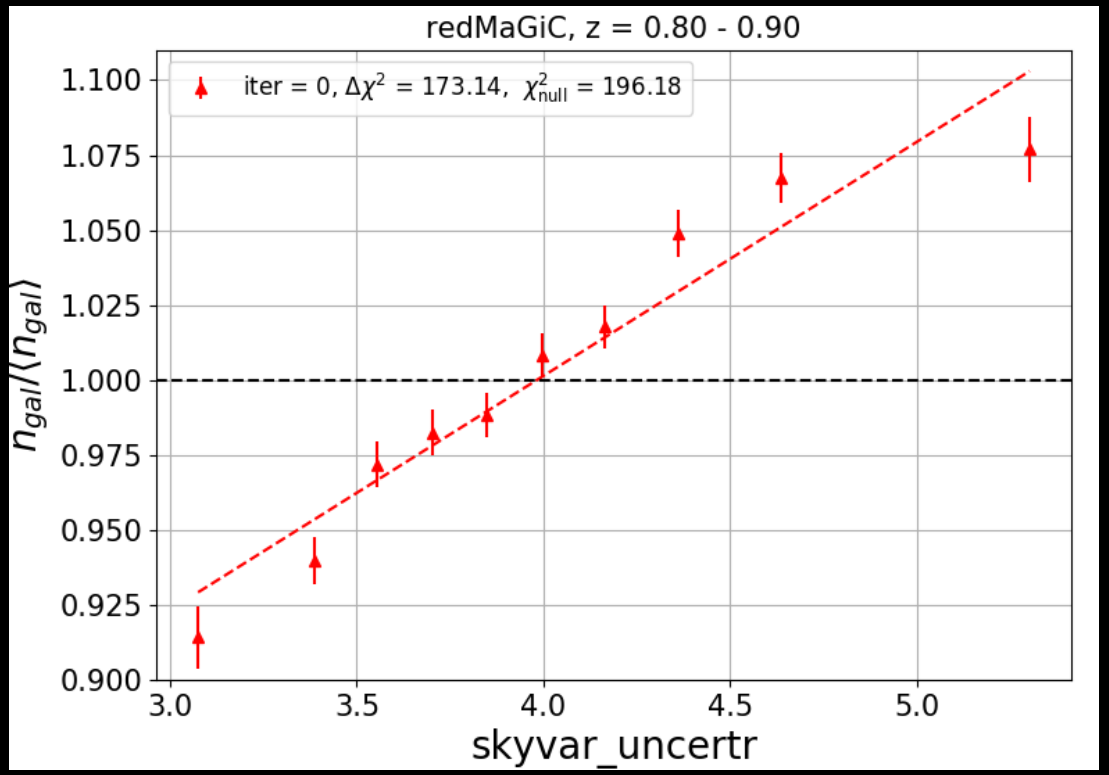
ISD: iterative pipeline

- ISD identifies the most contaminant SP map and corrects for it step-wise
- Steps:
  1. Identify the SP map that causes the most significant impact
     - Evaluate galaxy number density, n_{gal}, as function of SP map values
     - We call this relation *1D relation*. For a given SP map, s_1:
       I. Bin its values. We employ an equal area binning scheme
       II. Identify the pixels, p, on the sky such that $s_1^p \in bin_{SP}$
       III. Calculate average number density, $\bar{n}_{gal}^i$, on those regions
     - Evaluate significance of the contamination

Significance of the impact of an SP map, S_{1D}

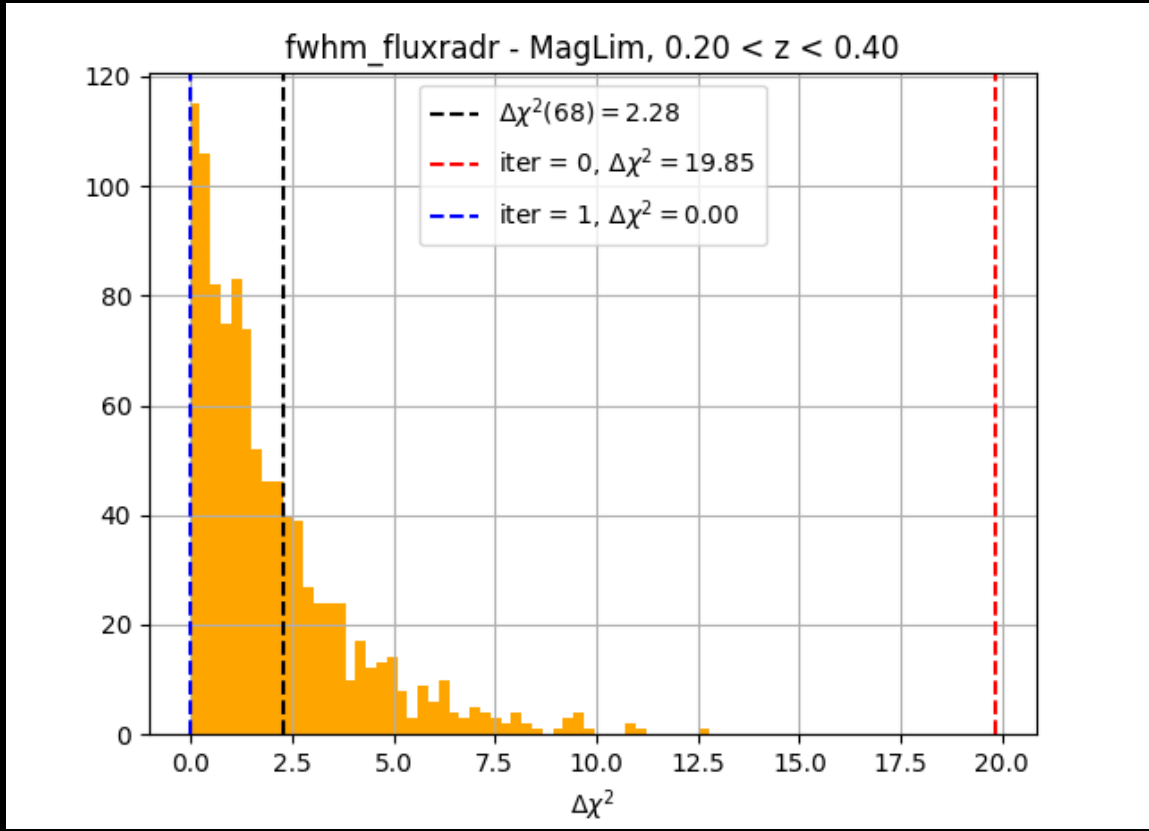
- We need to quantify the level of impact of an SP map, s_1, on the data
- To do it, we compute the 1D relation $\frac{n_o^i(s_1^i)}{\langle n_o \rangle}$, $i = 1, \ldots, N_{bins}^{1D}$
- We fit this relation to:
  - Null test: $\frac{n_o^i(s_1^i)}{\langle n_o \rangle} = 1 \quad \forall\, s_1^i \rightarrow \chi^2_{null}$
  - Model: $\frac{n_o^i(s_1^i)}{\langle n_o \rangle} = m\, s_1^i + c \rightarrow \chi^2_{model}$
- Level of impact on the data:

$$\Delta\chi^2 = \chi^2_{null} - \chi^2_{model}$$



Significance of the impact of an SP map, S_{1D}

- $\Delta\chi^2$ tells nothing about the significance of the contamination
- A non-zero signal can simply be due to chance correlations
- Compute $\Delta\chi^2_{mock}$ for s_1 with respect to 1000 mocks
- Obtain $\Delta\chi^2(68)$



Significance of the impact of an SP map, S_{1D}

- $\Delta\chi^2$ tells nothing about the significance of the contamination
- A non-zero signal can simply be due to chance correlations
- Compute $\Delta\chi^2_{mock}$ for s_1 with respect to 1000 mocks
- Obtain $\Delta\chi^2(68)$
- We define the significance of the contamination as

$$S_{1D} = \frac{\Delta\chi^2}{\Delta\chi^2(68)}$$

- Then, an SP map is considered to be significantly contaminant if $S_{1D} > T_{1D}$
- The significance threshold, T_{1D}, is fixed beforehand
- Our fiducial value is $T_{1D} = 2$

ISD: iterative pipeline

- ISD identifies the most contaminant SP map and corrects for it step-wise
- Steps:
  1. Identify the SP map that causes the most significant impact
  2. Obtain a weight map to correct for the impact of that SP map
     - s_1 identified as the most contaminant SP map
     - Calculate weight map, w_1, defined as

$$w_1 \equiv \frac{1}{F(s_1)}$$

     - $F(s_1)$ is fitted to the 1D relation of s_1
     - $F(s) = \begin{cases} m\sqrt{s} + c & \text{if } s = \textit{t\_eff\_exptime, skybrite} \text{ or } \textit{skyvar\_uncert} \\ m\,s + c & \text{if } s = \text{any other STD map and all PC maps} \end{cases}$
     - w_1 is normalised such that $\bar{w}_1 = 1$

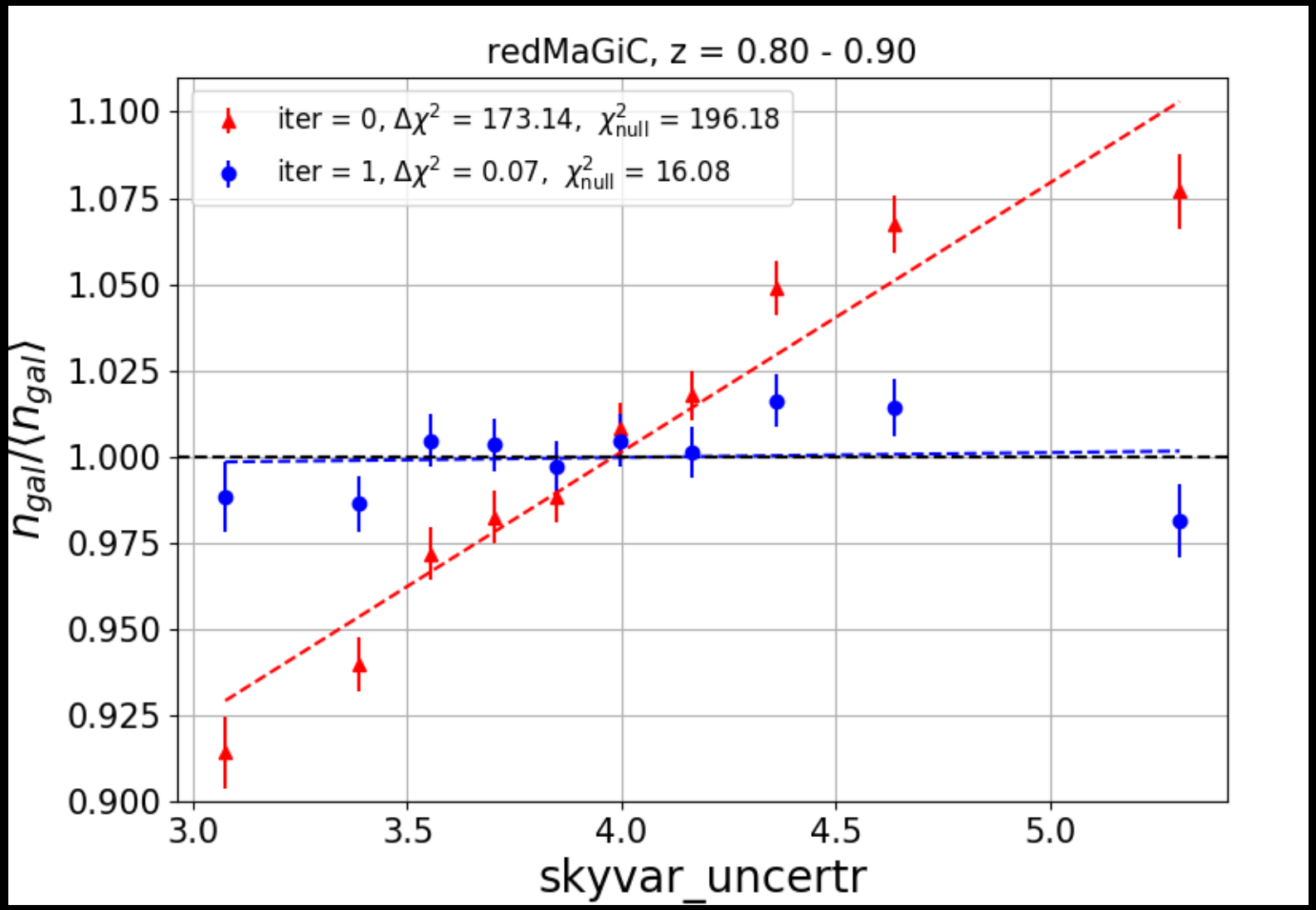
ISD: iterative pipeline

- ISD identifies the most contaminant SP map and corrects for it step-wise
- Steps:
  1. Identify the SP map that causes the most significant impact
  2. Obtain a weight map to correct for the impact of that SP map
  3. Apply the weight map to the data
     - The observed number of galaxies at pixel p, N_o^p, is re-scaled as

$$N_o^p \rightarrow N_d^p = w_1^p \cdot N_o^p$$

     - We do this at $N_{side} = 4096$

Correction of the data

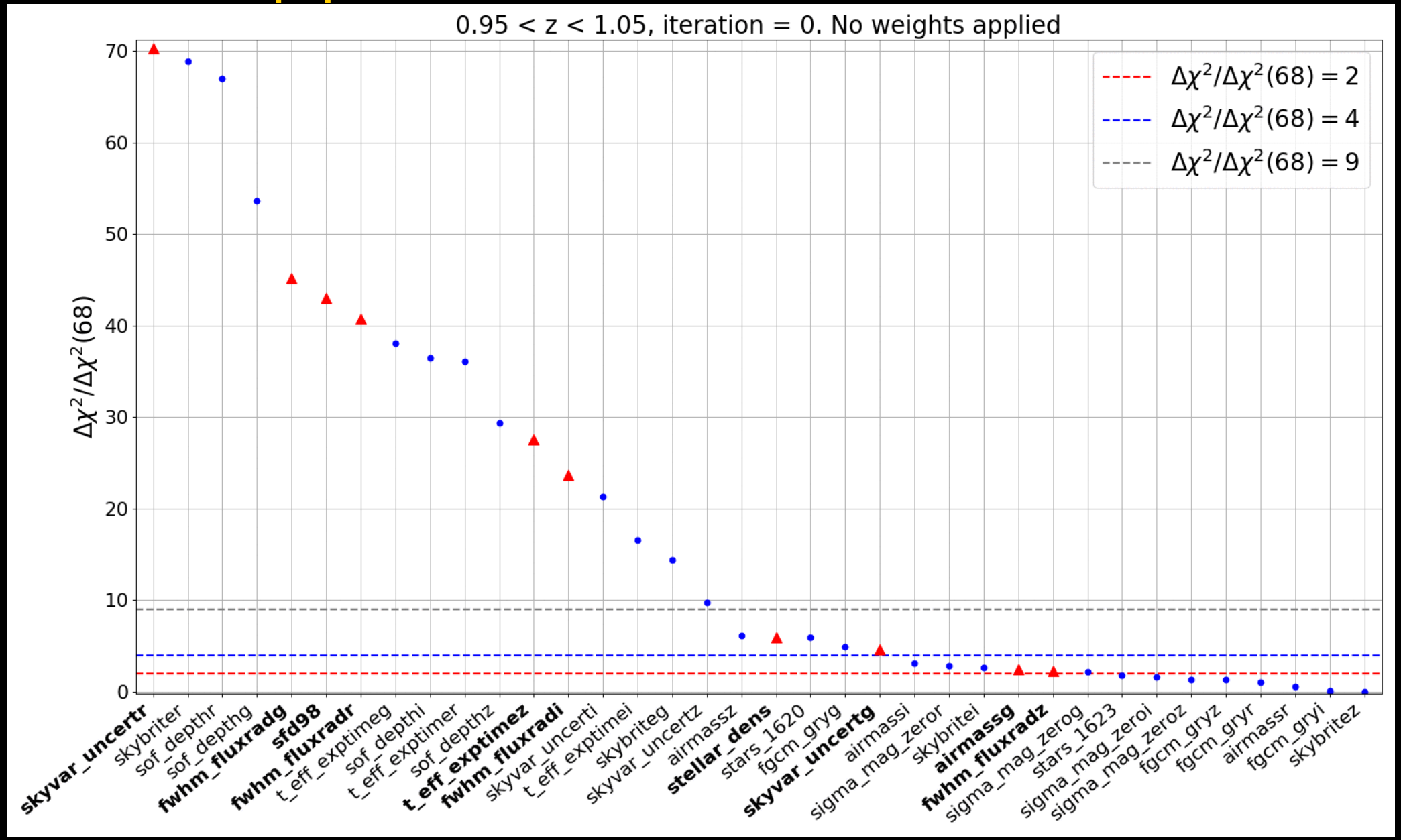


ISD: iterative pipeline

- ISD identifies the most contaminant SP map and corrects for it step-wise
- Steps:
  1. Identify the SP map that causes the most significant impact
  2. Obtain a weight map to correct for the impact of that SP map
  3. Apply the weight map to the data
  4. Re-evaluate until the contamination is lower than a pre-fixed threshold
     - Iteration 0: evaluate S_{1D} of all SP maps + correct for s_1
     - Iteration 1: go back to step 1 $\Rightarrow$ s_2 labelled as most contaminating SP map $\rightarrow$ obtain weight map w_2 $\rightarrow$ apply to the data
     - Repeat N times (N iterations) until all SP maps have $S_{1D} \leq T_{1D}$
     - Final weight map, w_T, is

$$w_T = \prod_{i=1}^{N} w_i \text{ , with } \overline{w}_T = 1$$

ISD: iterative pipeline



Elastic Net regularisation, *ENet*

- 1D marginals → contamination could be missed if weakly distributed across maps
- Multilinear fit → fit all SP maps and obtained contamination amplitudes, α_i
- High number of SP maps ⇒ higher risk of overcorrection (overfitting)
- Solution → Elastic net (ENet) regularisation (LASSO+ridge):
  - Minimise loss function: $Loss = \frac{1}{2\,N_{pix}} \|\delta_{obs} - S\alpha\|_2^2 + \lambda_1 \|\alpha\|_1 + \frac{\lambda_2}{2} \|\alpha\|_2^2$
  - LASSO (λ_1) → penalises non-zero α_i ⇒ favours reduction of SP maps
  - Ridge (λ_2) → penalises correlation between maps
- Multilinear fit → less sensitive to SP basis
- Different configurations to both galaxy samples ($N_{side} = 512$). We use ENet for:
  - Weights to contaminate mocks
  - Evaluate difference between methods

  → Systematic terms to the covariance

Validation of the weights and robustness tests....

Weights validation

- Systematics mitigation ⇒ remove non-cosmological signal
- However, the correction methods can induce biases on the recovered $w(\theta)$
  - Undercorrection
  - Overcorrection
- Therefore, it is necessary to perform an exhaustive validation process
- Elements of the validation process:
  - Simulations:
    - Uncontaminated mocks: same log-normal mocks used for S_{1D}
    - Contaminated mocks: mocks that incorporate contamination detected on the data
    - Decontaminated mocks: mocks that undergo the decontamination process

Weights validation

- Systematics mitigation ⇒ remove non-cosmological signal
- However, the correction methods can induce biases on the recovered $w(\theta)$
  - Undercorrection
  - Overcorrection
- Therefore, it is necessary to perform an exhaustive validation process
- Elements of the validation process:
  - Simulations
  - Criterion to quantify the impact of the bias: $\chi^2 > 3$ ⇒ marginalise over it / account for it
  - Definition of the biases and tests to detect them
  - Procedure to account for the bias: systematic terms to the covariance

Contaminated mocks

- We wish to have realisations of the Universe affected by contamination
- Idea → use the weight maps to contaminate log-normal mocks:

$$N_t^p \rightarrow N_c^p = N_t^p \cdot \frac{1}{w_T^p}$$

- We generate contaminated log-normal mocks:
  - We can run a decontamination method on them ⇒ Decontaminated mocks

Contaminated mocks

- We wish to have realisations of the Universe affected by contamination
- Idea → use the weight maps to contaminate log-normal mocks:

$$N_t^p \rightarrow N_c^p = N_t^p \cdot \frac{1}{w_T^p}$$

- We generate contaminated log-normal mocks
- Same method to contaminate and decontaminate ⇒ potential flaw of the validation:
  - We test sensitivity to forms of contamination we know a priori we are sensitive to

Contaminated mocks

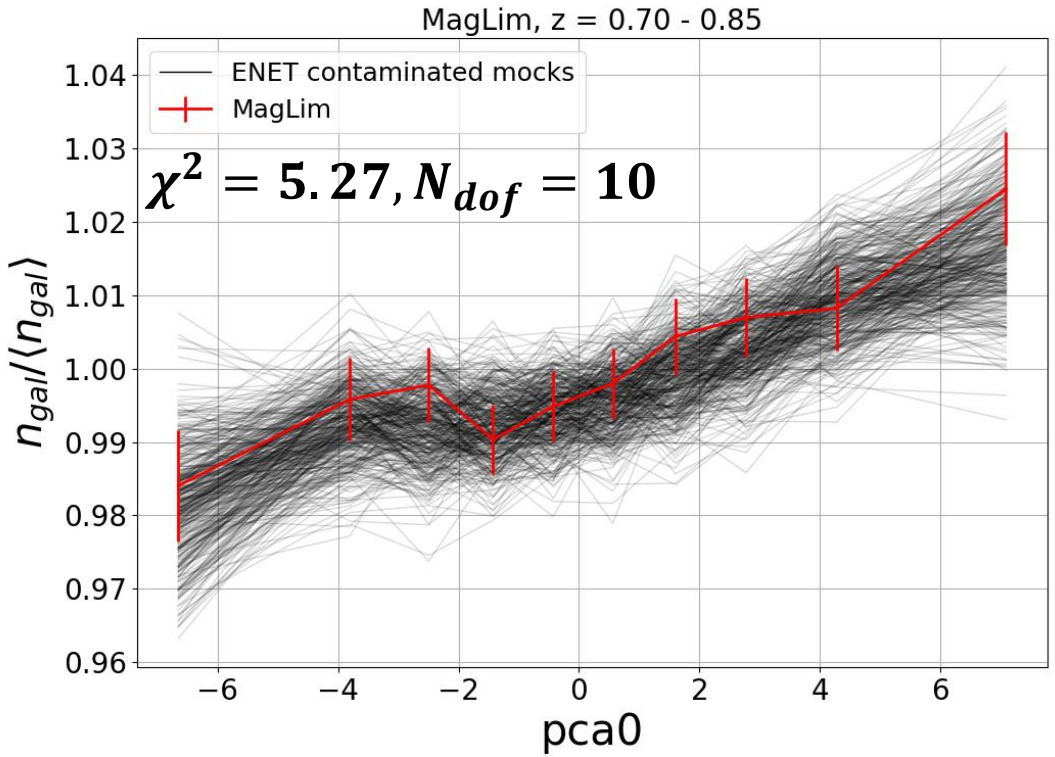
- We wish to have realisations of the Universe affected by contamination
- Idea → use the weight maps to contaminate log-normal mocks:

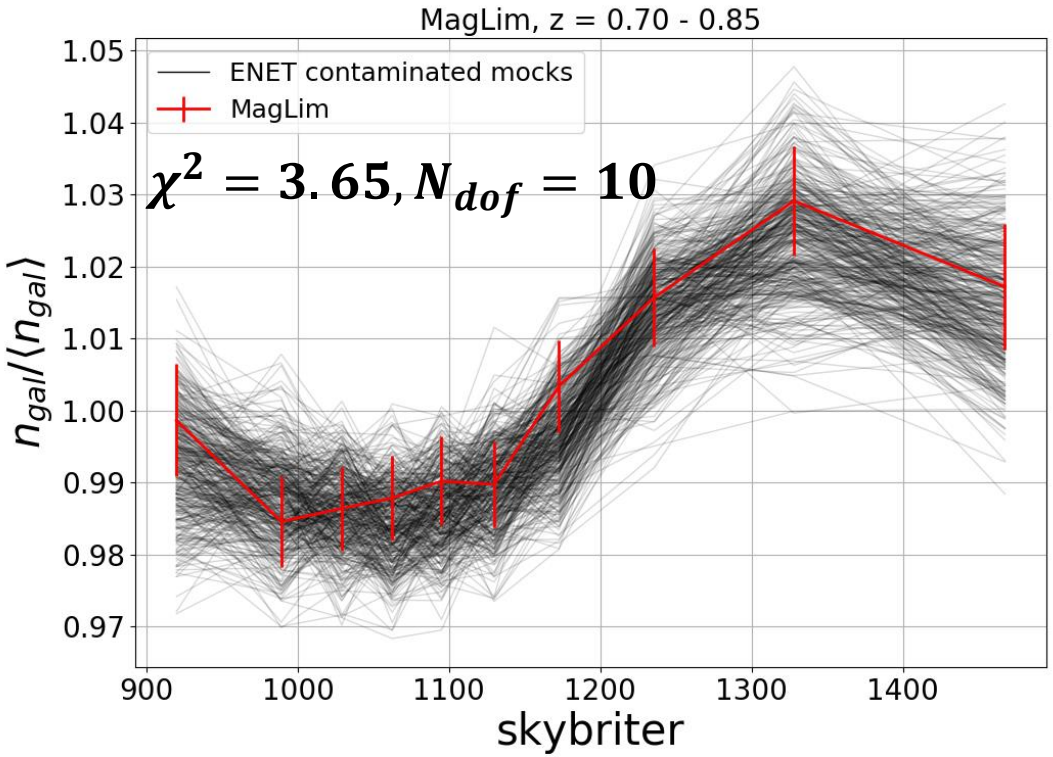
$$N_t^p \rightarrow N_c^p = N_t^p \cdot \frac{1}{w_T^p}$$

- We generate contaminated log-normal mocks
- Same method to contaminate and decontaminate ⇒ potential flaw of the validation
- Solution: contaminate with weights from ENet and decontaminate with ISD
  - Both methods determine the level of contamination in different ways
  - Avoid blind spots in the validation → unveil biases

Contaminated mocks

- We generate 400 ENet-STD107 contaminated mocks for both MagLim and redMaGiC
- Verify that the mocks reproduce the contamination of the data





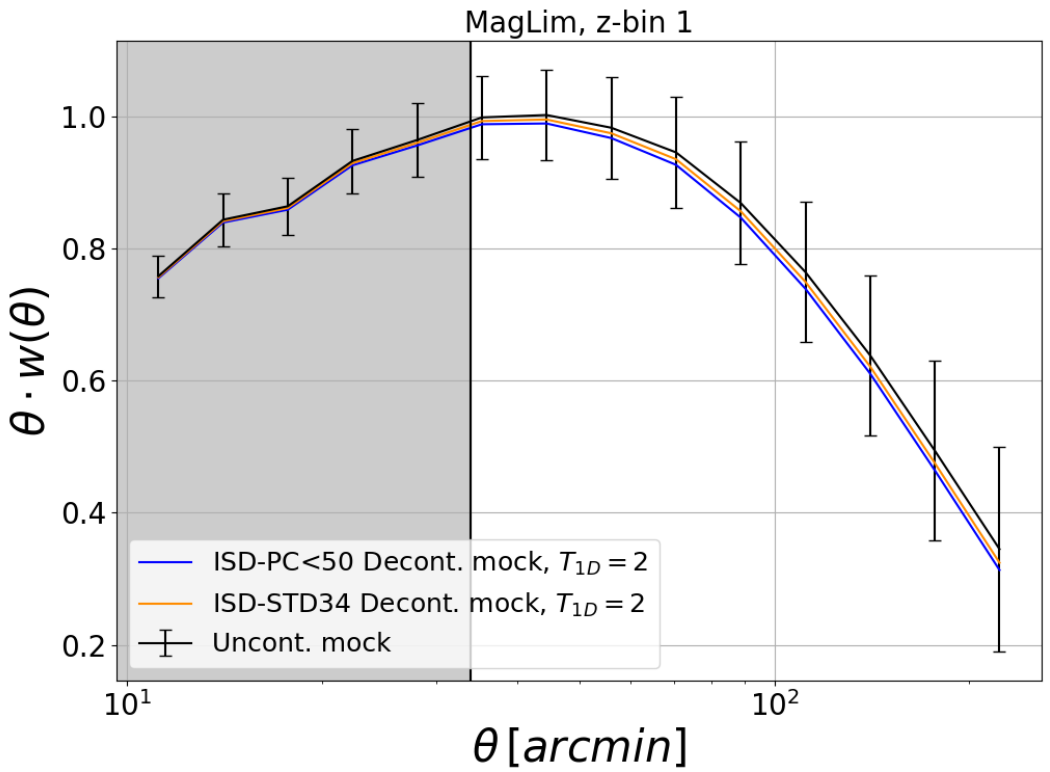
False correction bias

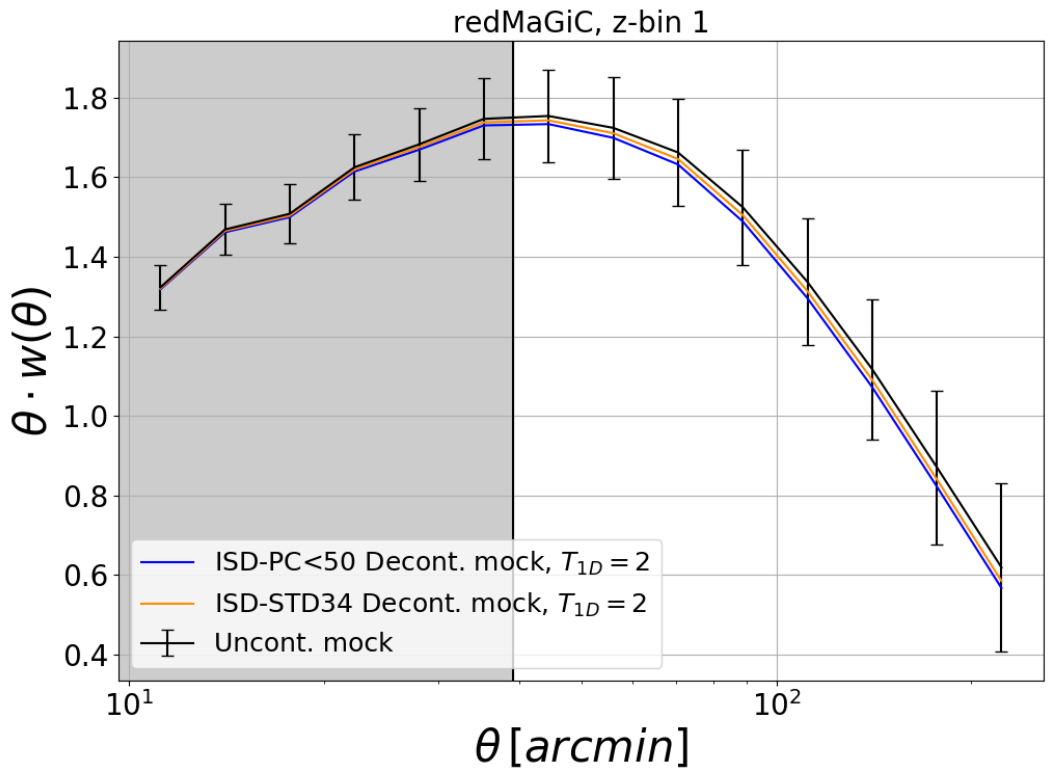
- Chance correlations, i.e. the structure of an SP map resembling that of the data, can lead to overcorrection
- Large number of SP maps and strict T_{1D} increase probabilities of chance correlations
- Overcorrection ⇒ remove actual cosmological structure
- We use $N = 400$ uncontaminated mocks
- We run ISD on these mocks with configurations ISD-PC<50 and ISD-STD34
- False correction bias estimator:

$$w_{f.c.bias}^{T_{1D}}(\theta) = \frac{1}{N}\left(\sum_{i=1}^{N} w_{w,unc,i}^{T_{1D}} - \sum_{j=1}^{N} w_{unc,j}\right)(\theta)$$

False correction bias

Error bars contain systematic contribution (as we will explain)





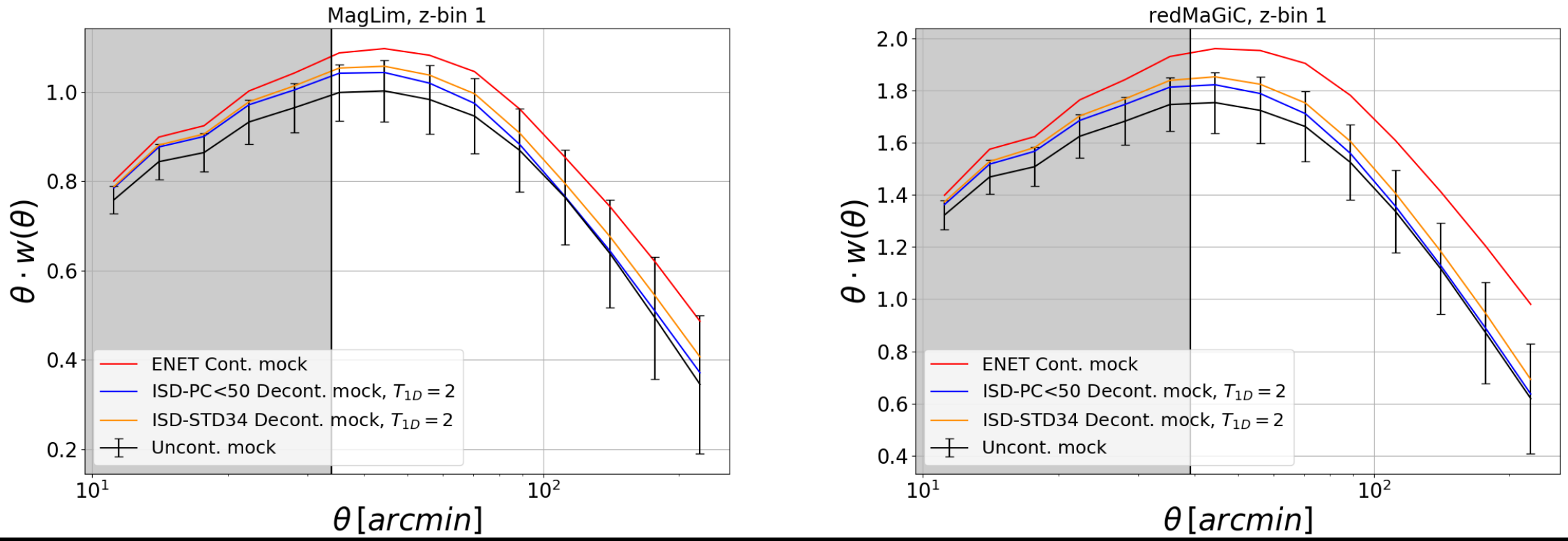
Residual systematic bias

- Unidentified contaminating SP maps could lead to undercorrection. When using ISD:
  - T_{1D} is too high (too relaxed)
  - Low-significance linear combinations of SP maps (due to marginalisation)
- We use $N = 400$ ENet contaminated mocks:
  - Aggressive level of contamination with ENet-STD107 weights
  - Imprint modes of contamination to which ISD may not be sensitive
  - Combining two methods ⇒ additional level of robustness, avoiding blind spots
- We run ISD on these mocks with configurations ISD-PC<50 and ISD-STD34
- Residual systematic bias estimator:

$$w_{r.s.bias}^{T_{1D}}(\theta) = \frac{1}{N}\left(\sum_{i=1}^{N} w_{dec,i}^{T_{1D}} - \sum_{j=1}^{N} w_{unc,j}\right)(\theta)$$

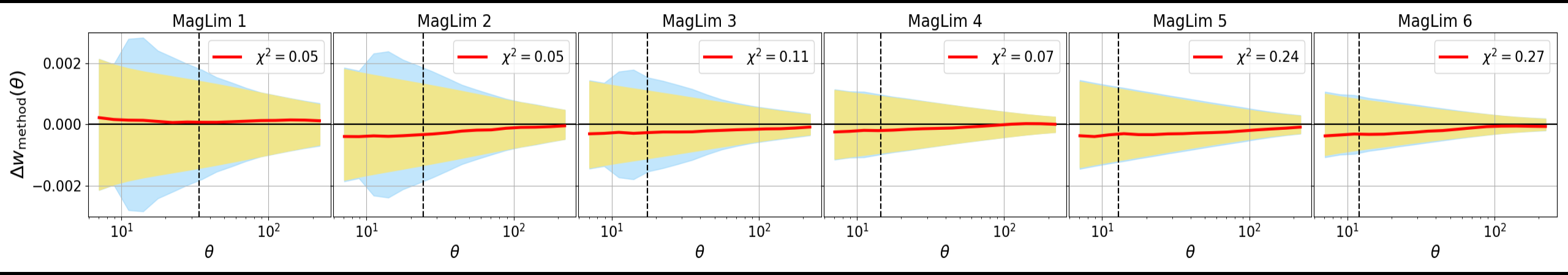
Residual systematic bias

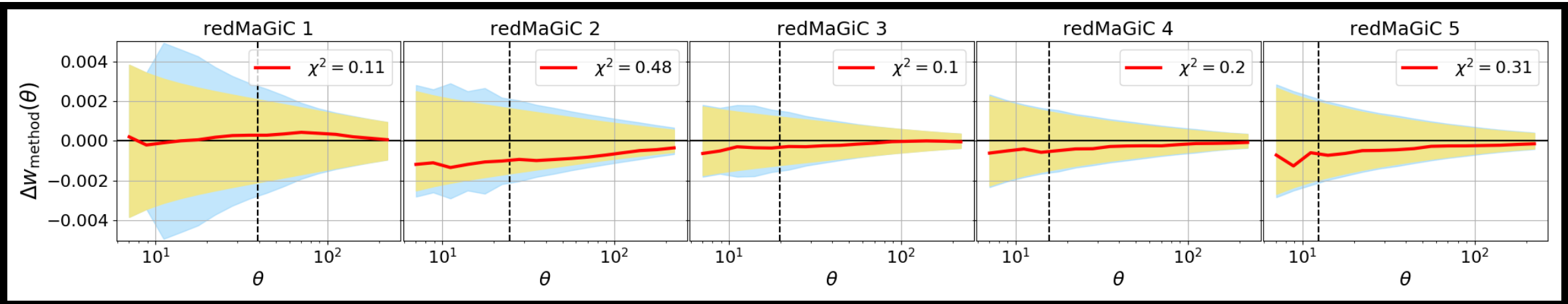
Error bars contain systematic contribution (as we will explain)



Systematic terms to the covariance

Account for systematic uncertainties = modify covariance matrix → systematic contribution to the error budget on $w(\theta)$



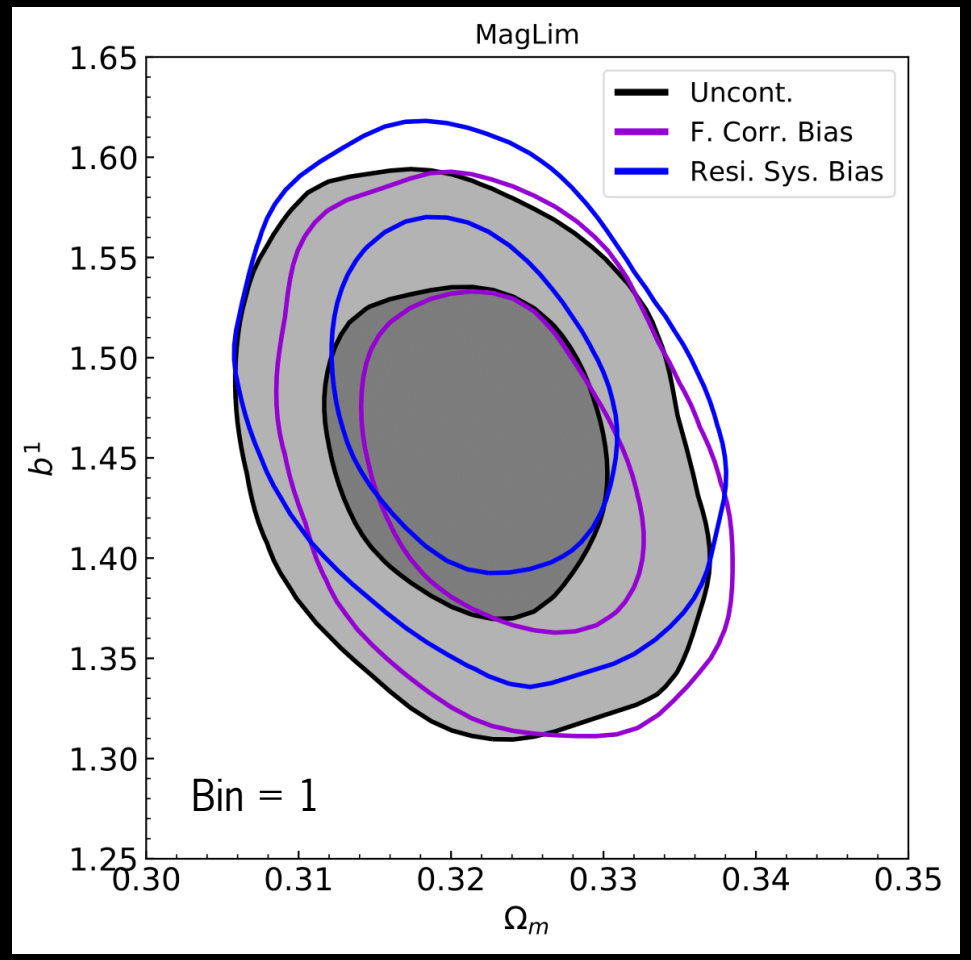


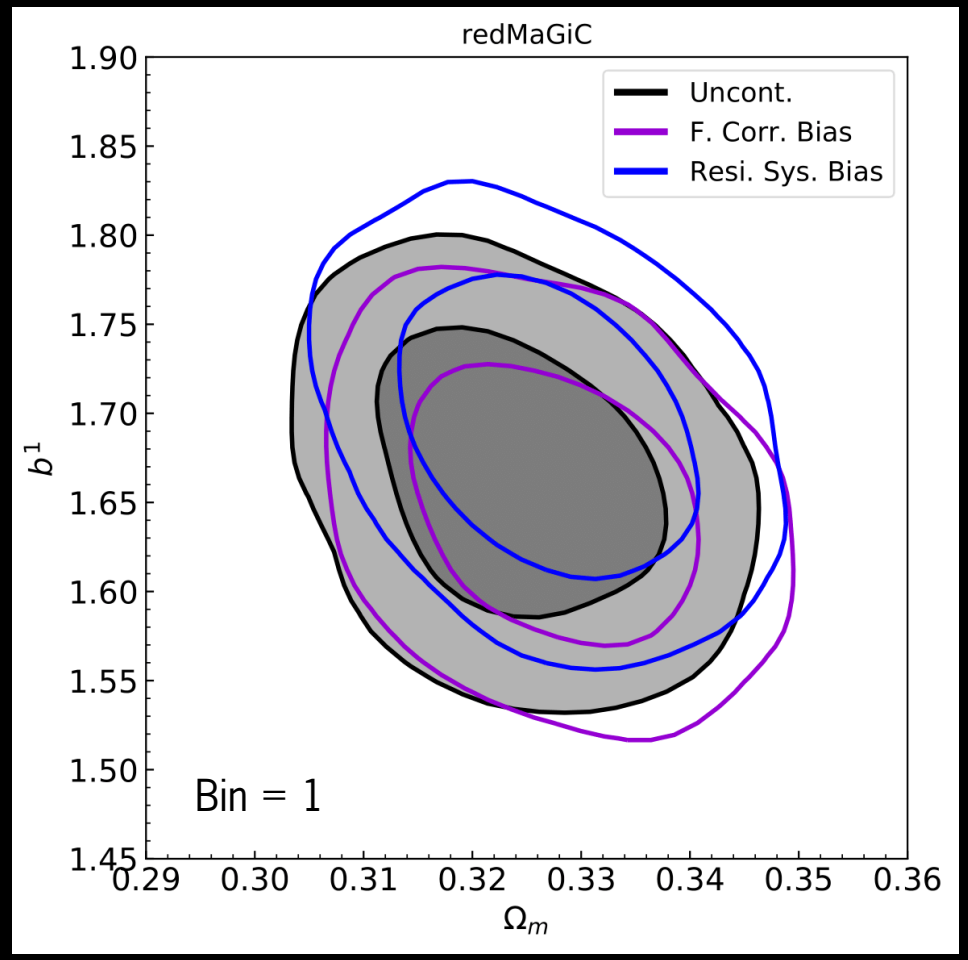
Rodríguez-Monroy et al. 2021

Impact on parameter estimation

- Finally, we wish to check the impact of the biases on the estimation of cosmological parameters
- We employ as data vectors the mean of three different sets of 400 $w(\theta)$:
  - $\bar{w}_{unc}(\theta)$ (uncontaminated mocks)
  - $\bar{w}_{w,unc}(\theta)$ (decontaminated mocks from false correction bias test)
  - $\bar{w}_{dec}(\theta)$ (decontaminated mocks from residual systematic bias test)
- Systematic effects → impact on galaxy clustering amplitude

Impact on parameter estimation



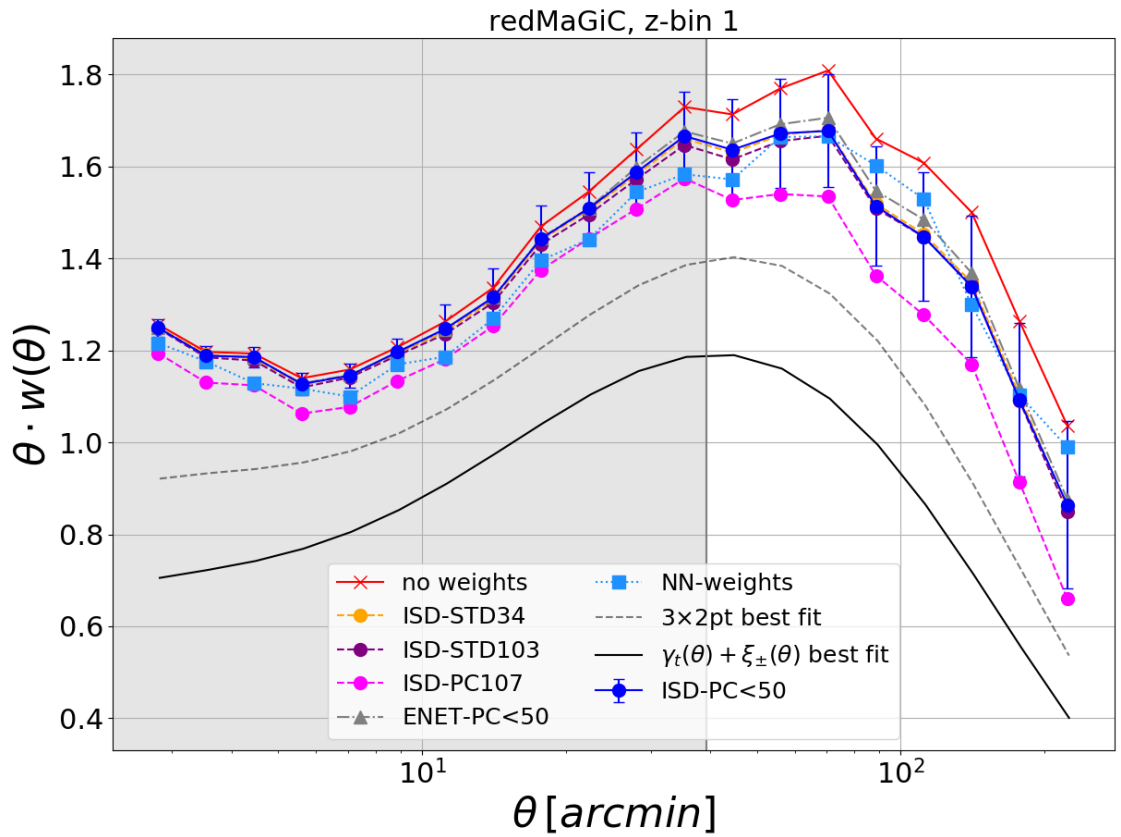


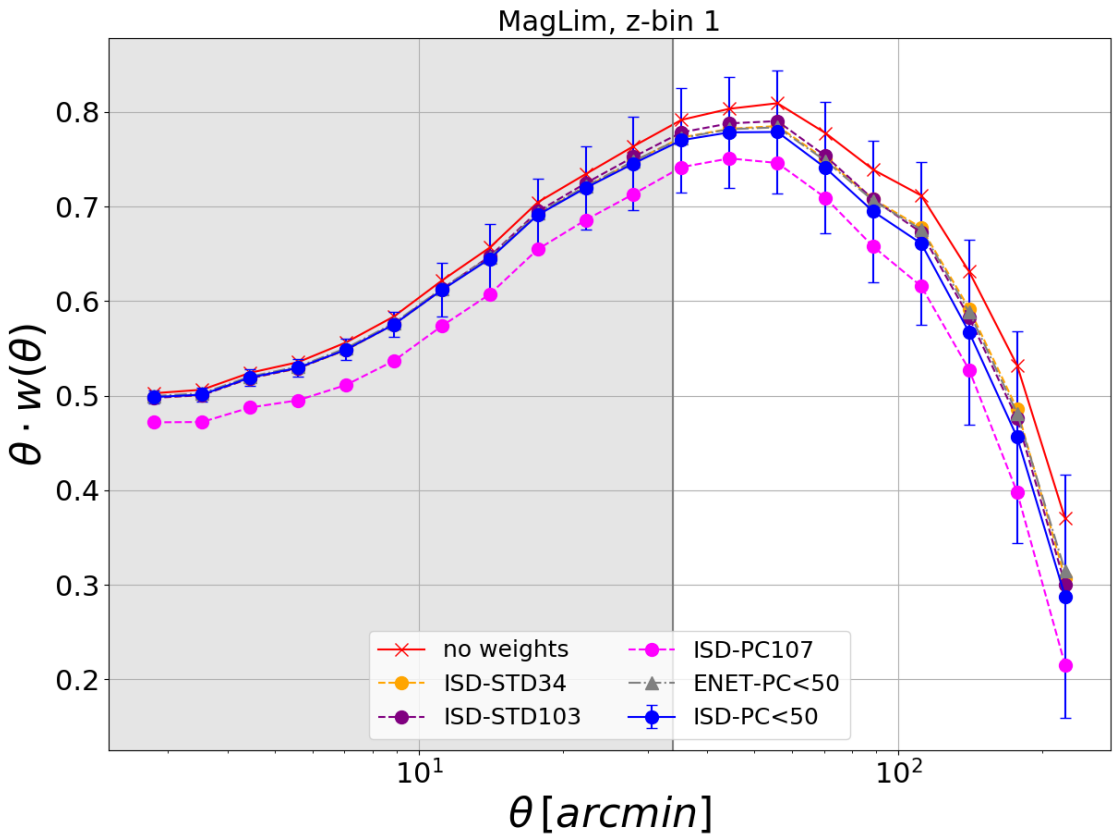
Additional robustness tests

- After unblinding of the redMaGiC sample an inconsistency between the corrected amplitude of $w(\theta)$ and that predicted by weak lensing
- This translates into a decorrelation of the galaxy bias measured by $w(\theta)$ and $\gamma_t(\theta)$
- Observational systematics raise $w(\theta)$ ⇒ further investigate from this point of view
- Additional tests:
  - ISD-STD103
  - ISD-PC107 (results in overcorrection)
  - ENet-PC<50
  - NN-weights (only for redMaGiC)
  - Assumption of linearity → χ^2_{null}
  - ISD with Gaia EDR3 or Planck 2013 maps

→ negligible effect on $w(\theta)$

Additional robustness tests





Lessons learned….

Lessons learned

DES-Y3 has been challenging from the point of view of observational systematics. We should apply the lessons learned not only to DES-Y6, but also to similar surveys:

- Know the potential limitations of the method you are employing: linearity assumption, marginalization vs multilinear fit, calibration with simulations
- It is advantageous to have different methods (relying on different assumptions) applied to the same data and to combine them. This allows to
  - avoid blind spots
  - ensure the robustness
  - obtain systematic uncertainties associated with the choice of method
- Validation on simulations: it is critical to be able to determine the level of different biases that the methods can introduce

Lessons learned

- Know well your SP maps: risk of tracing actual LSS from the data. Especially important when the maps are created from the data, e.g., the FWHM from PSF fitting, or sky-brightness maps
- Use data-driven methods to identify the contaminants
- Be careful and justify well any pre-selection of SP maps (e.g., correlations between them). More risk on excluding too many maps than on using to many of them
- Different SP bases can help to exclude potential problems with some methods
- External information/tracers as cross-check: other surveys, convergence maps, simulations....
- Use more than one lens galaxy sample: exact same methods, with exact same choices and exact same validation process applied to different galaxy samples
- Importance of a correct systematic mitigation, especially for surveys with a shrinking statistical error → DES-Y6, LSST, Euclid....

Merci beaucoup