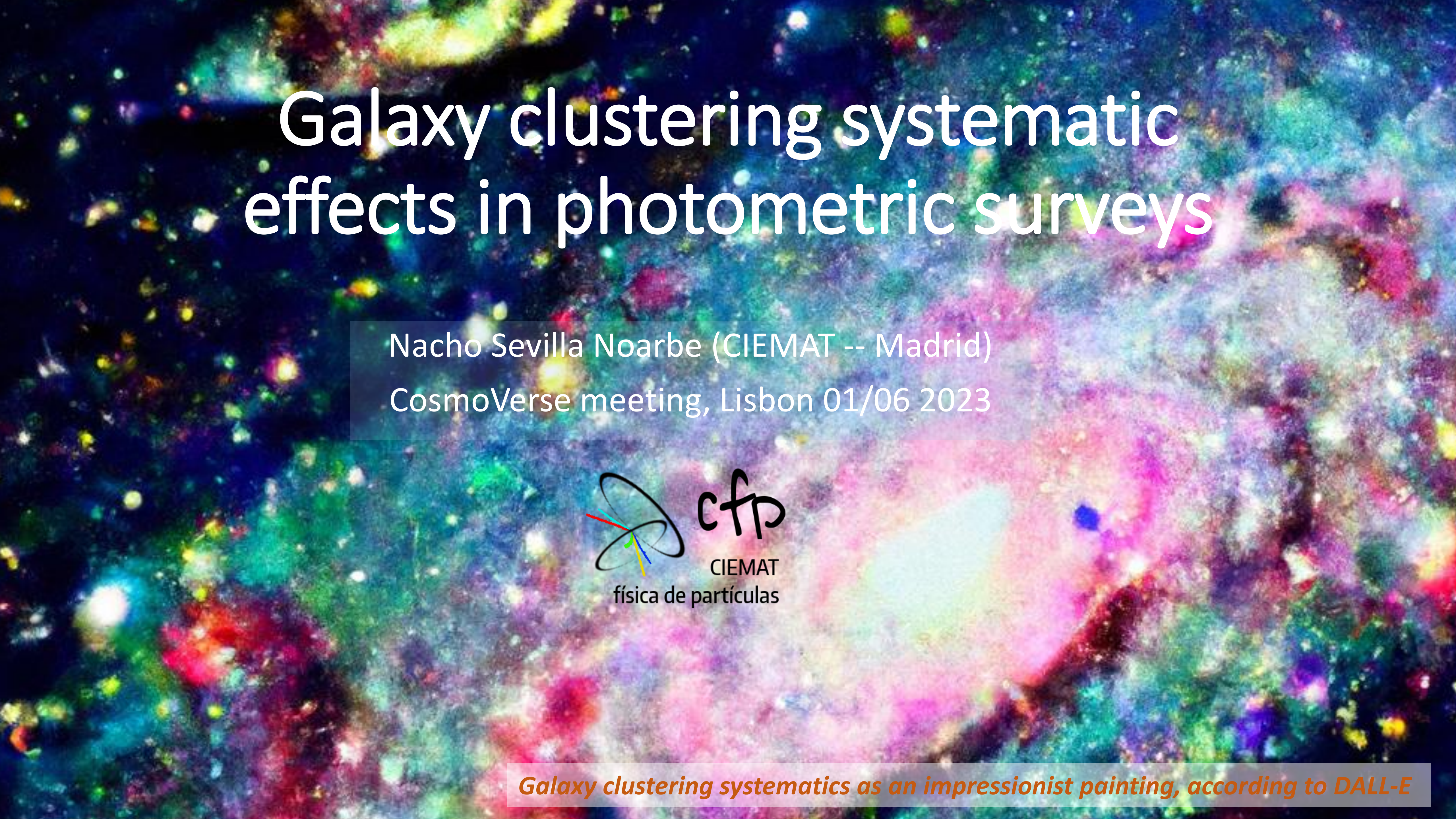
Galaxy clustering systematic effects in photometric surveys

Nacho Sevilla Noarbe (CIEMAT -- Madrid)

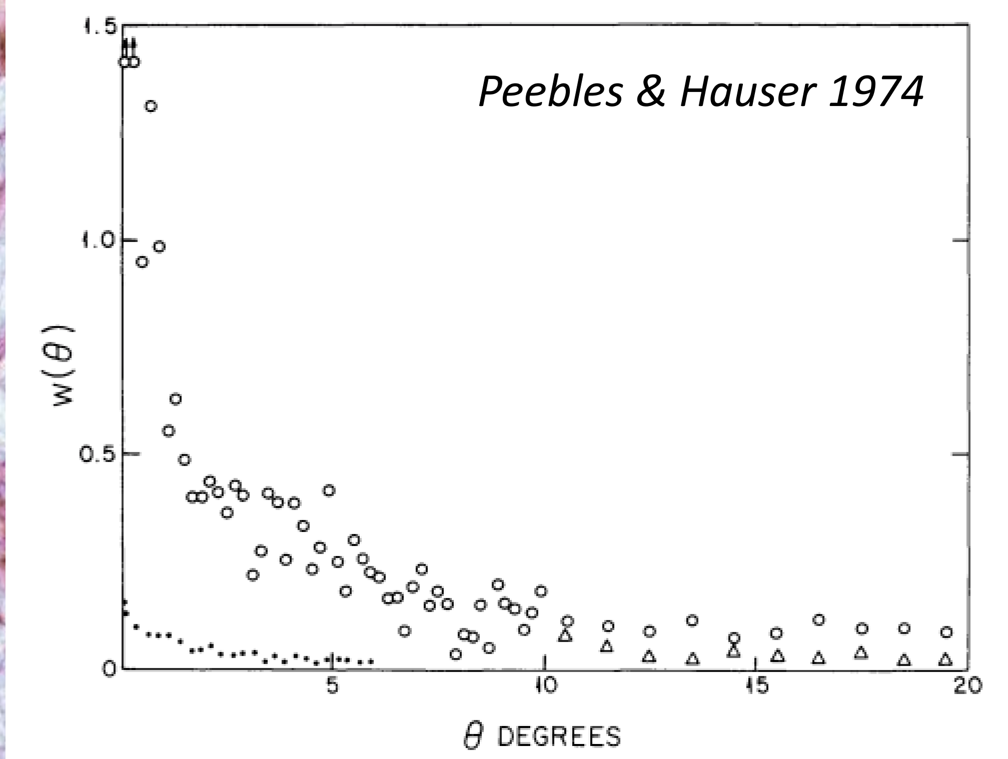
CosmoVerse meeting, Lisbon 01/06 2023

cfp
CIEMAT
física de partículas

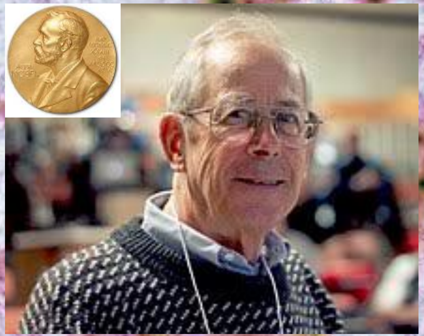


Galaxy clustering systematics as an impressionist painting, according to DALL-E

Large Scale Structure of galaxies is a major component of current cosmology analyses

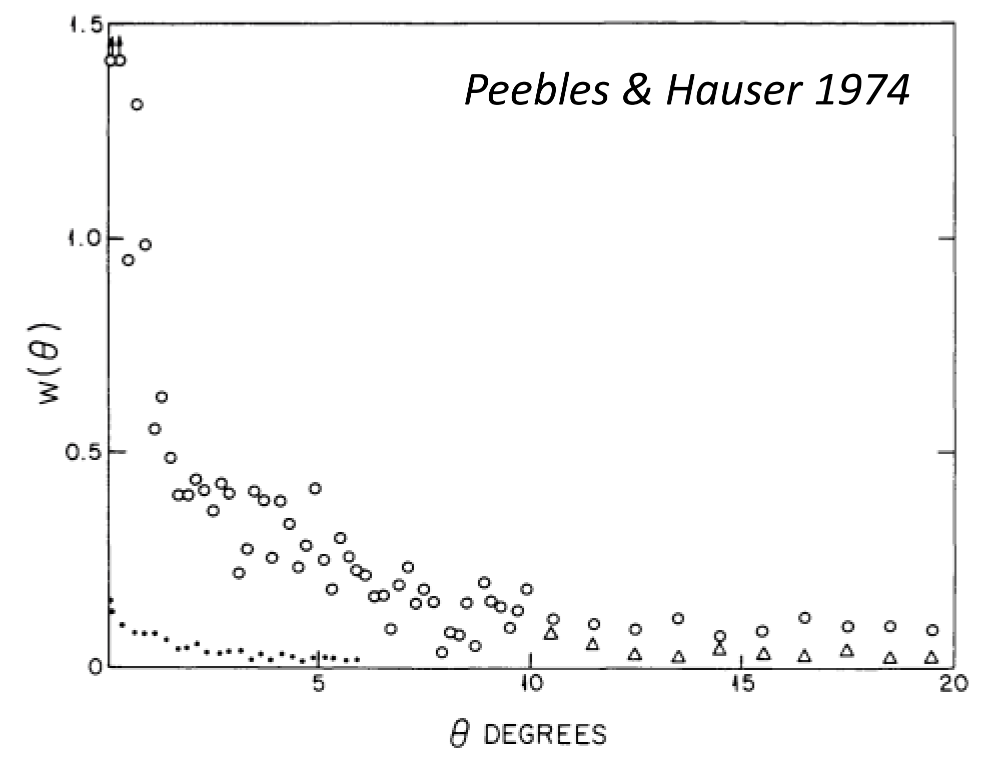


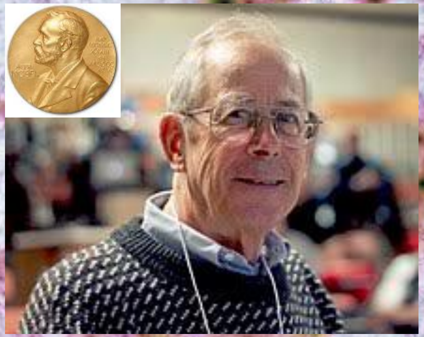
Peebles & Hauser 1974



Illustris simulation

Large Scale Structure of galaxies is a major component of current cosmology analyses





What are the observational effects that can modulate this signal and result in shifts in the S8-Om parameter space and how do we correct them?

Illustris simulation

The Dark Energy Survey is …

… a 1” resolution map of 1/8 of the southern sky

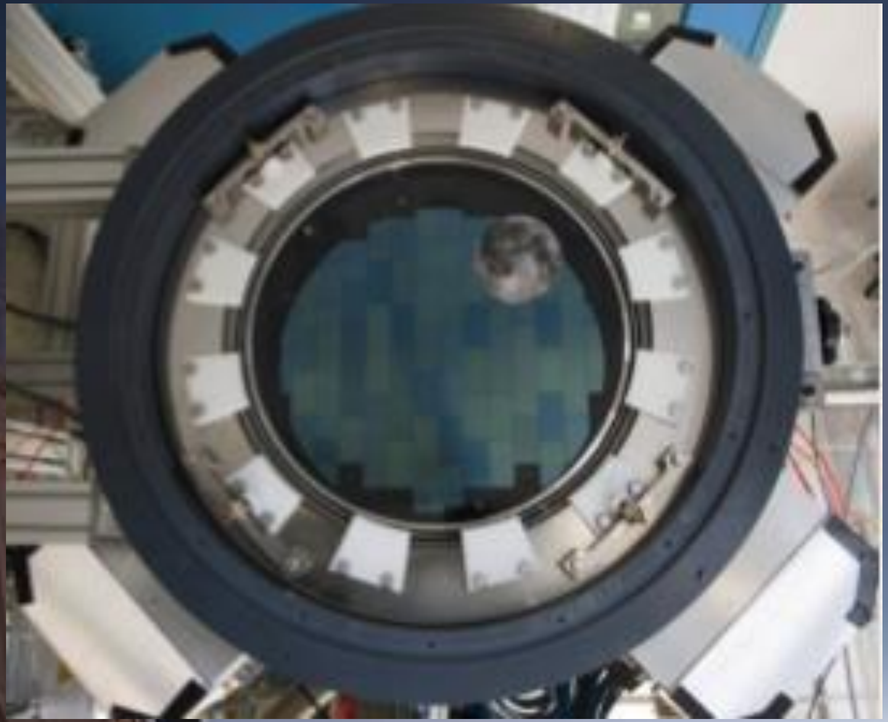
… up to depths of i_{AB} ~ 24

… in 5 photometric bands (*g, r ,i , z , Y*) → photo-zs

… to explore dark energy using several probes:

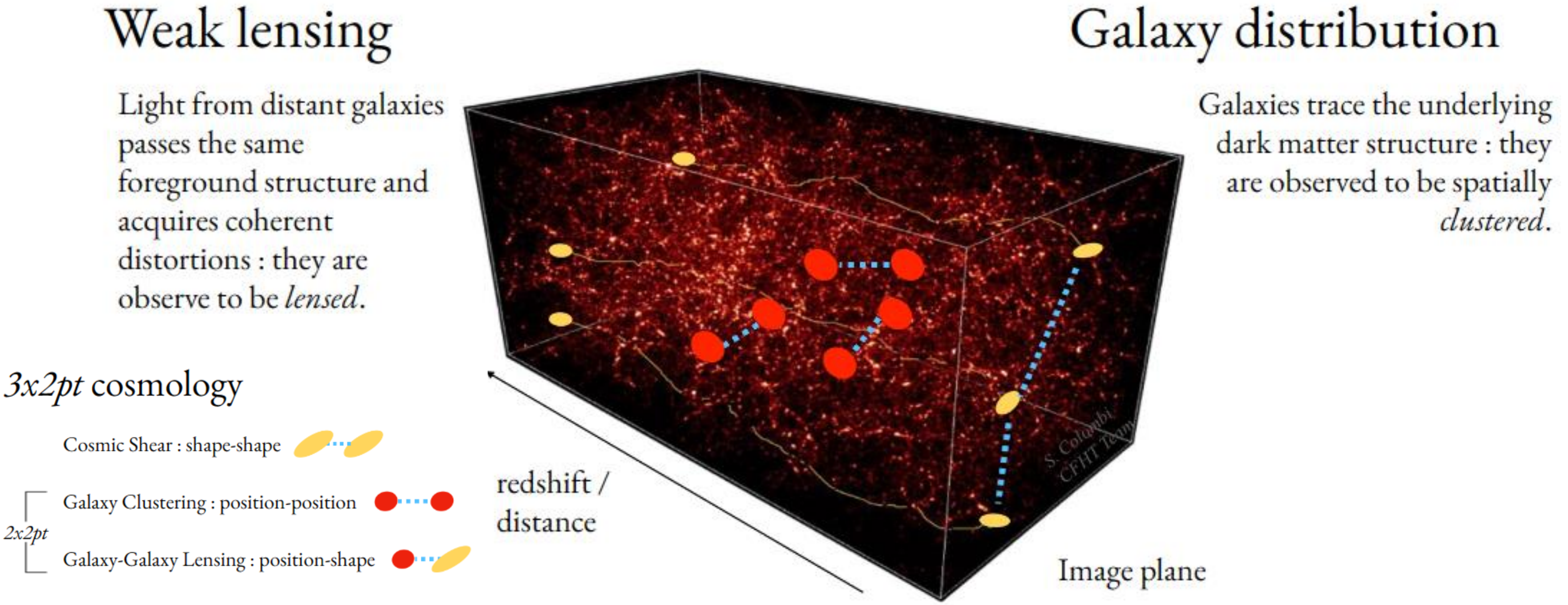
- Supernovae Ia
- BAOs
- Cluster counts
- **Weak lensing + clustering**

… during 525+(~50) nights in 6 years (2013-2019)



Credit: Tim Abbott

In photometric surveys, we combine with shear measurements to produce powerful constraints



Weak lensing

Light from distant galaxies passes the same foreground structure and acquires coherent distortions : they are observe to be *lensed*.

Galaxy distribution

Galaxies trace the underlying dark matter structure : they are observed to be spatially *clustered*.

3x2pt cosmology

Cosmic Shear : shape-shape

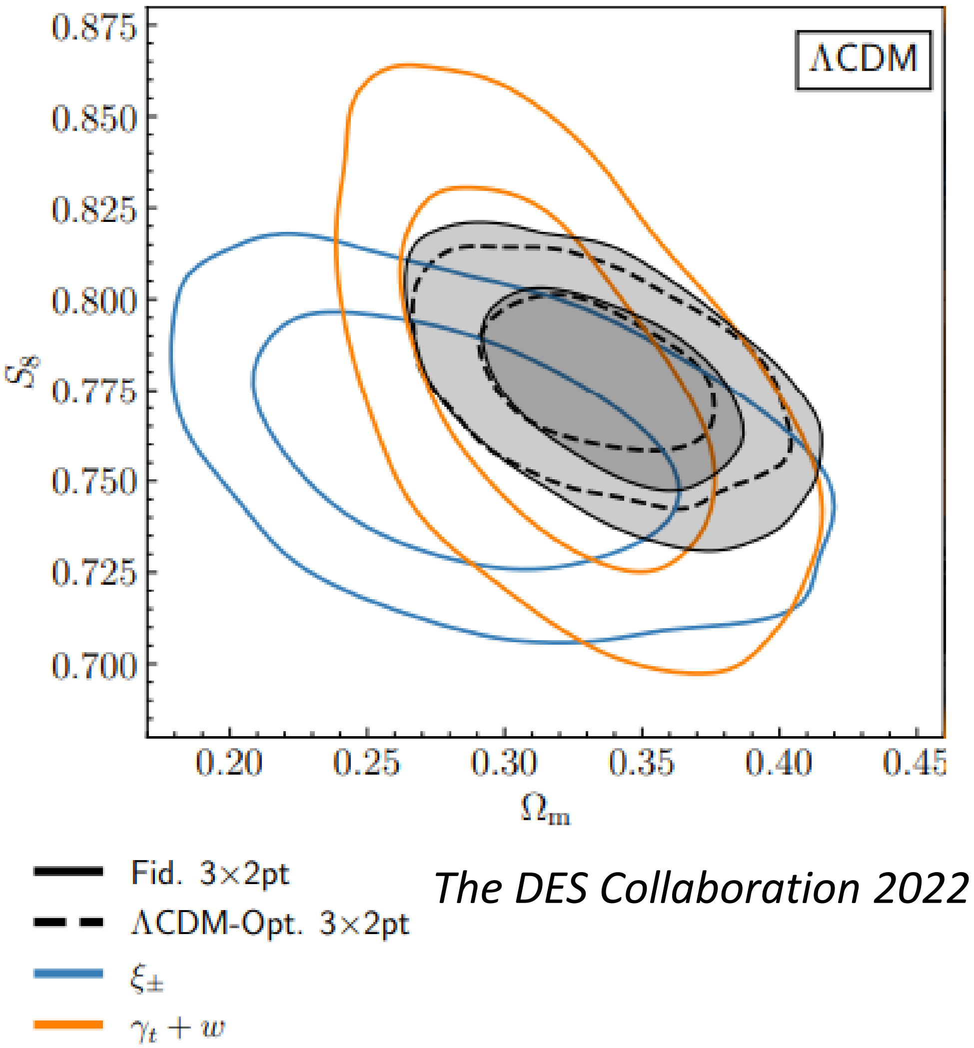
2x2pt

Galaxy Clustering : position-position

Galaxy-Galaxy Lensing : position-shape

Credit: The DES Collaboration

In photometric surveys, we combine with shear measurements to produce powerful constraints



The DES Collaboration 2022

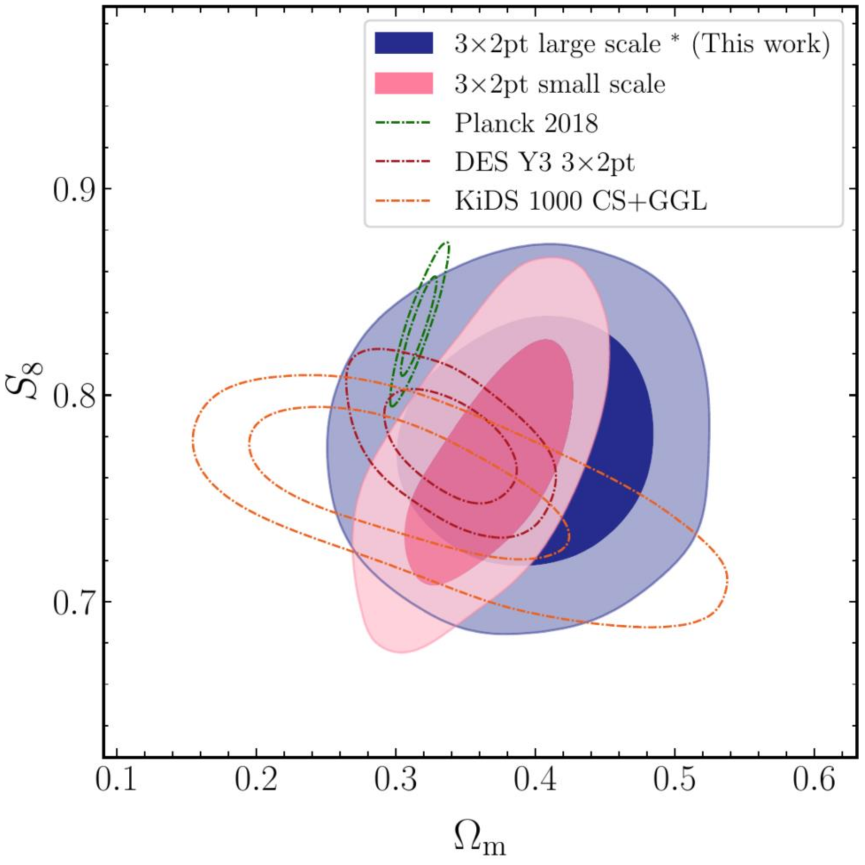
DES-Y3 results on its own are compatible with high-z inferred cosmological parameters.

Y3 catalogs available tens of millions of galaxies with photo-zs:

https://des.ncsa.illinois.edu/releases/y3a2

DES-Y6 will double the depth and introduce refinements in modelling, systematics treatment, publication next year.

In photometric surveys, we combine with shear measurements to produce powerful constraints



S.Sugiyama et al. 2023

The current observational picture is one in which high redshift observations from Planck are consistent with this low redshift probe.

However, three different surveys with different techniques and areas consistently report a 2-sigma discrepancy.

(a lot of literature on statistical effects)

Many astrophysical effects can shift things around by 0.5-1 sigma.

Can observational systematics on clustering affect here?

There are two major sources of systematic effects for LSS in DES

$$\boxed{w^i(\theta)} = \left(b^i\right)^2 \int \frac{dl\, l}{2\pi} J_0(l\theta) \int d\chi \frac{\left[\boxed{n^i_{\rm g}(z(\chi))} dz/d\chi\right]^2}{\chi^2} P_{\rm NL}\left(\frac{l+1/2}{\chi}, z(\chi)\right)$$

Photometric redshifts: incorrect bin assignment; inaccurate redshift distribution for signal prediction.

$$n^i_g(z) = n^i_{PZ}(z - \Delta z^i)$$

Effects that remove or add clustering power (removes/adds galaxies, in general with some spatial pattern).

$$\delta_g = W \cdot s + n$$

Observed

Also NB: different galaxy samples will respond differently to the above.

There are multiple approaches to redshift distribution estimation

Empirically calibrated analytical expression:
$N(z) \propto z^a \cdot \exp(-z/z0)^b$

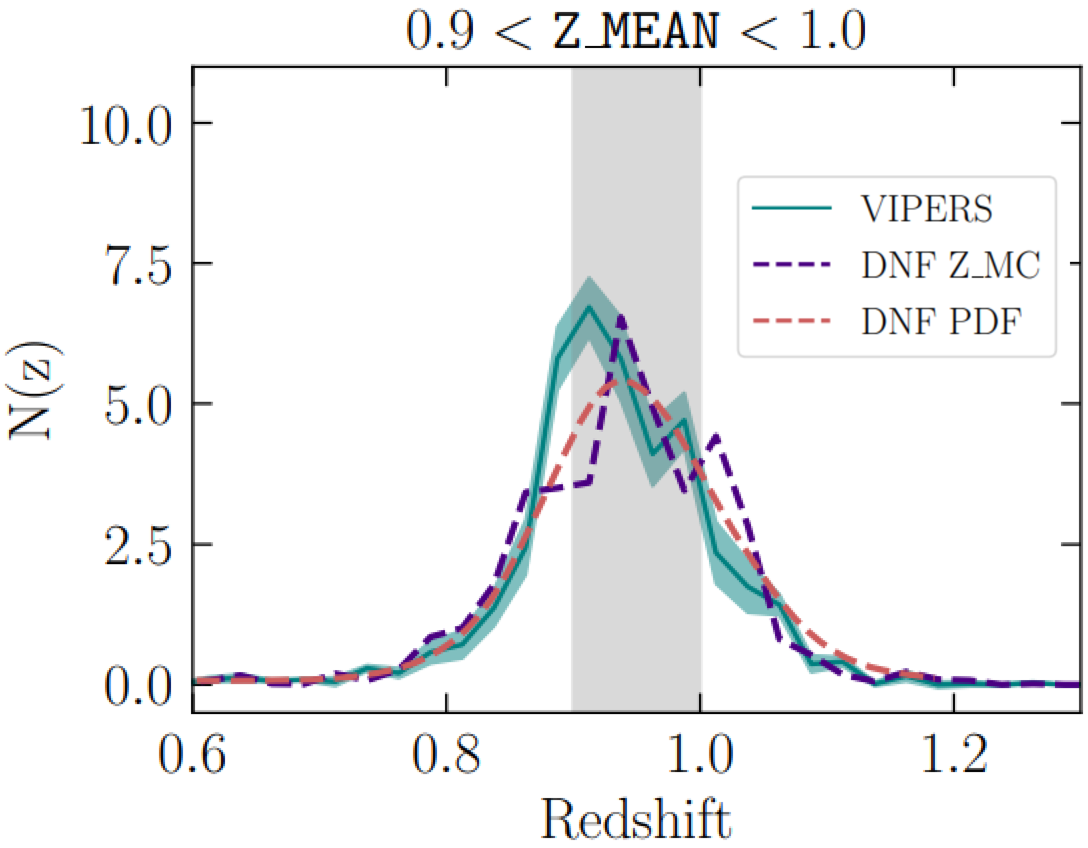
Simulations of galaxy fluxes —> estimations of photo-z as if observed

Spectroscopic sample

Estimate from photo-z code

Cross-correlations with a spec-z LSS sample

Calibrations through richer datasets (SOMpz)



A.Carnero-Rosell et al. 2022
J. de Vicente, E. Sánchez, I.S-N 2015

We obtain an estimate on N(z), and marginalize over uncertainties on the bias and width.

There are multiple approaches to redshift distribution estimation

Empirically calibrated analytical expression:
$N(z) \propto z^a \cdot \exp(-z/z0)^b$

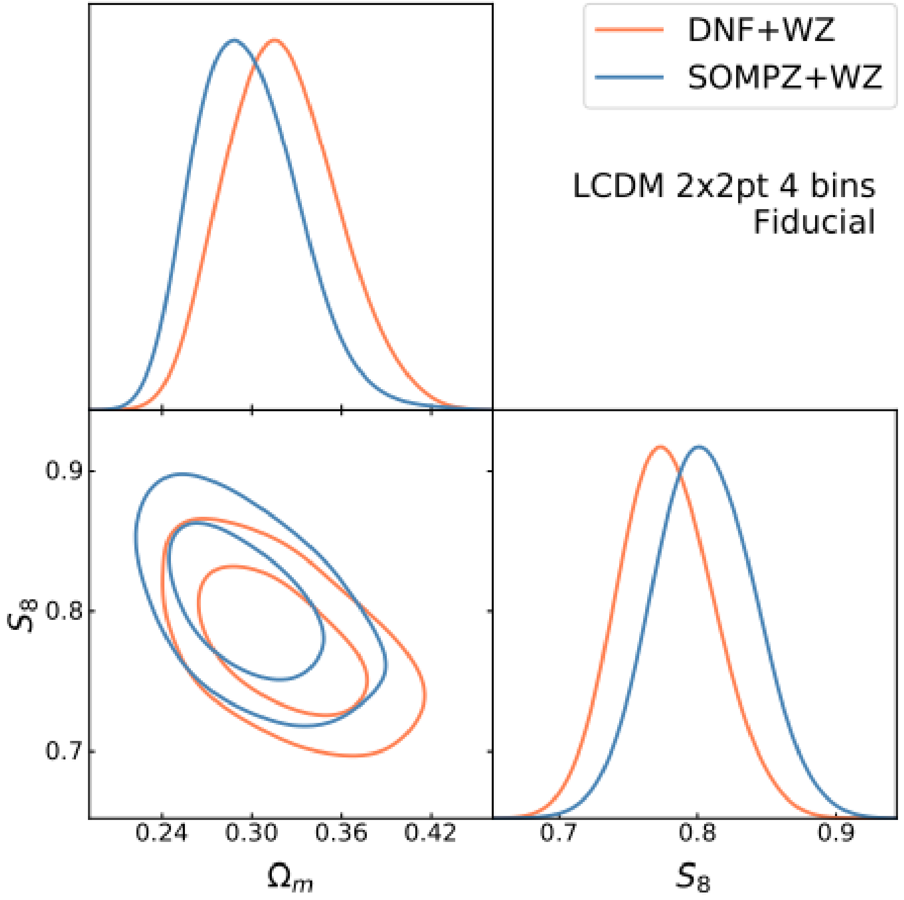
Simulations of galaxy fluxes —> estimations of photo-z as if observed

Spectroscopic sample

Estimate from photo-z code

Cross-correlations with a spec-z LSS sample

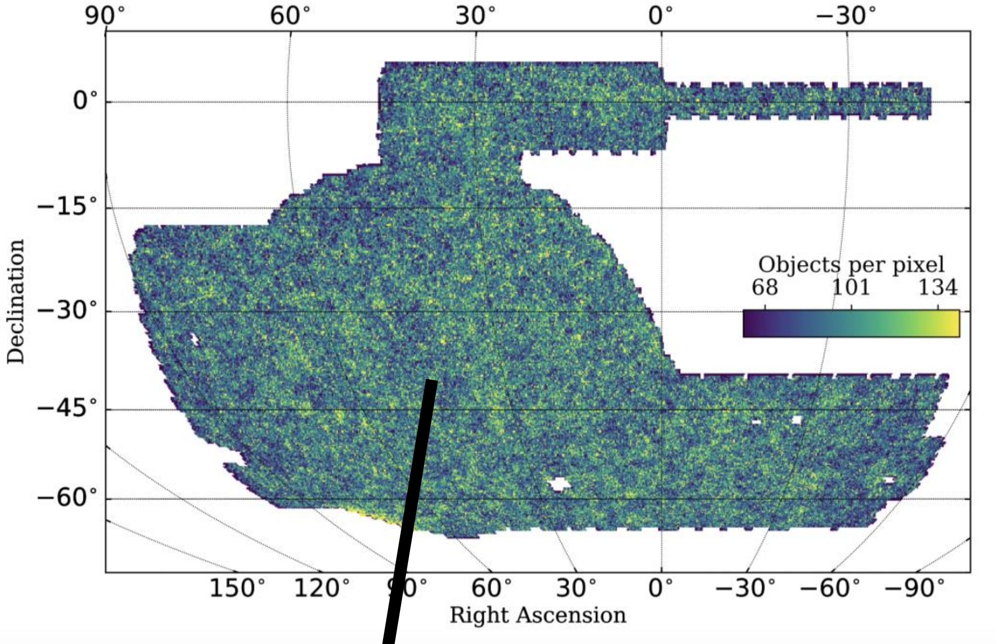
Calibrations through richer datasets (SOMpz)



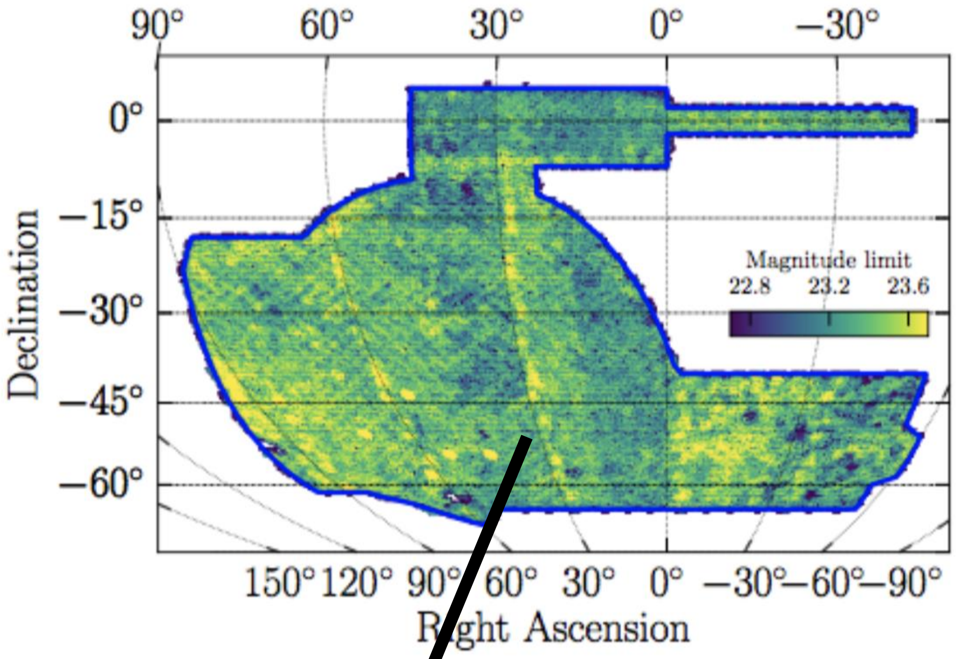
G.Giannini et al. 2022

We can encounter small shifts, that are relevant!

We can remove systematic effects on clustering by mapping survey properties



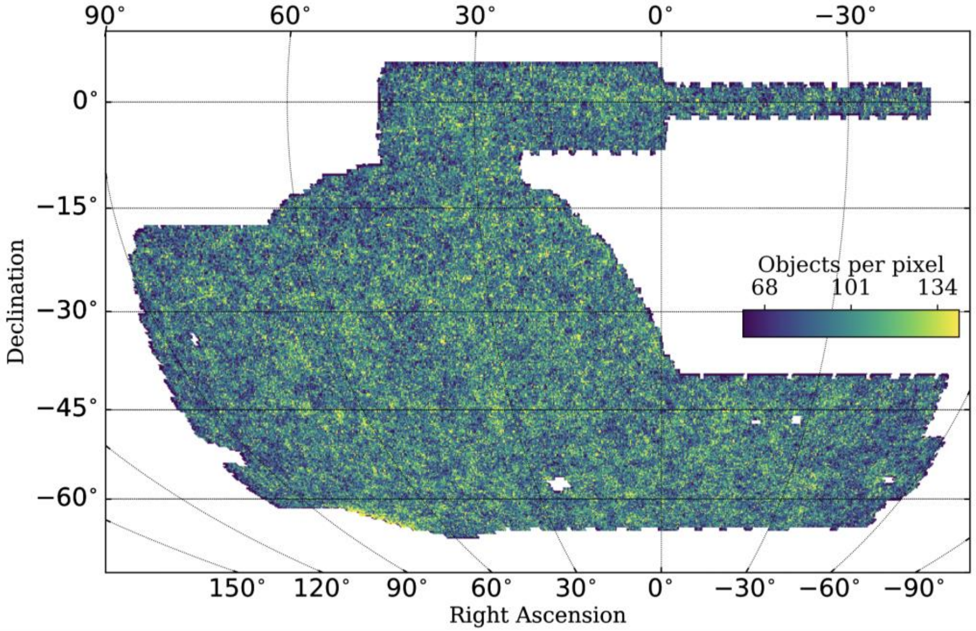
The DES Collaboration 2021



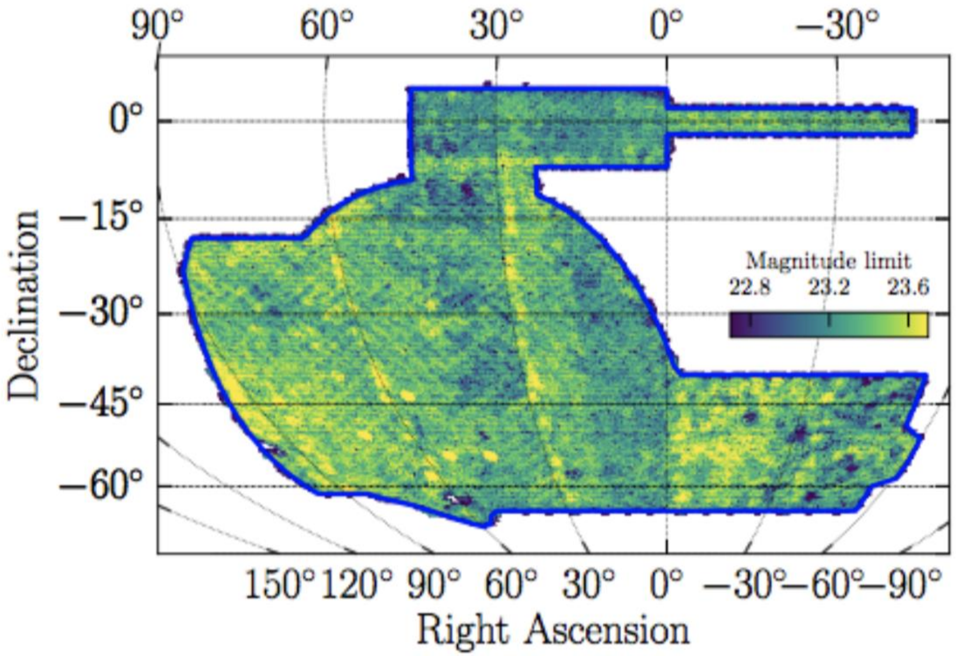
I.S-N et al. 2021

$$\delta^{\mathrm{obs}}(\bar{\mathbf{u}}) = \delta^{\mathrm{true}}(\bar{\mathbf{u}}) + \sum_{i=1}^{N} \varepsilon_i \delta_{\mathrm{i}}^{\mathrm{sys}}(\bar{\mathbf{u}})$$

We can remove systematic effects on clustering by mapping survey properties



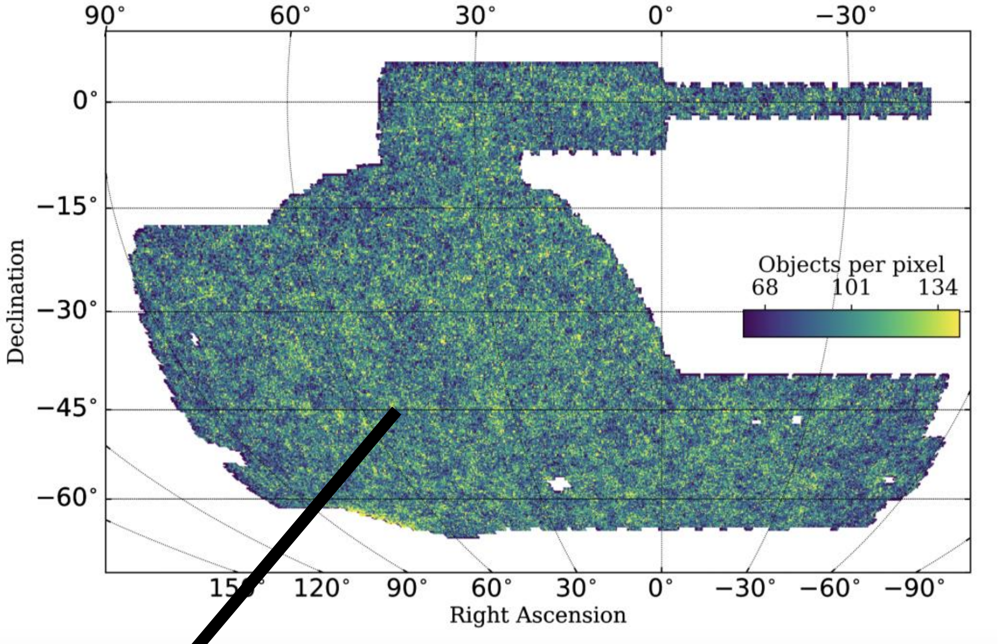
The DES Collaboration 2021



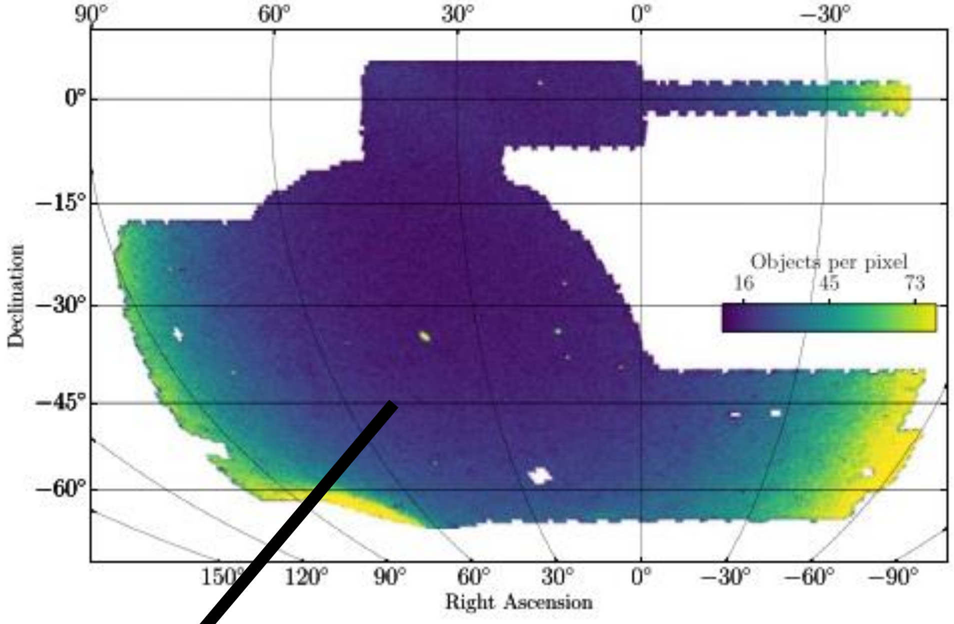
I.S-N et al. 2021

$$\omega^{\mathrm{obs}}(\theta) = \omega^{\mathrm{true}}(\theta) + \sum_{i} \sum_{j} \varepsilon_i \varepsilon_j \langle \delta_{\mathrm{i}}^{\mathrm{sys}} \delta_{\mathrm{j}}^{\mathrm{sys}} \rangle$$

We can remove systematic effects on clustering by mapping survey properties



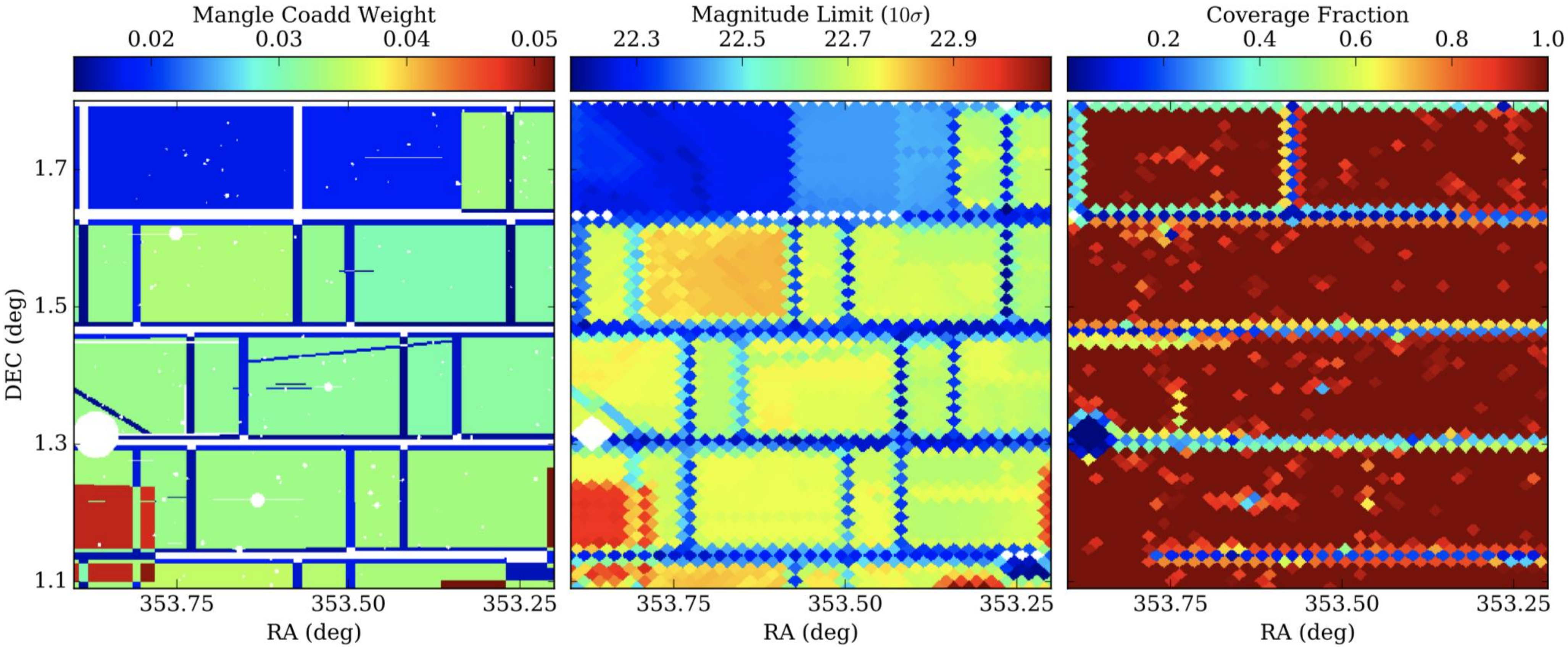
The DES Collaboration 2021



The DES Collaboration 2021

$$\delta^{obs}(\bar{\mathbf{u}}) = \delta^{true}(\bar{\mathbf{u}})(1\text{–fstar}) + \delta^{star}(\bar{\mathbf{u}})\text{fstar} + \sum_{i=1}^{N} \varepsilon_i \delta_i^{sys}(\bar{\mathbf{u}})$$

These maps can be built by approximating detailed measurements of survey conditions

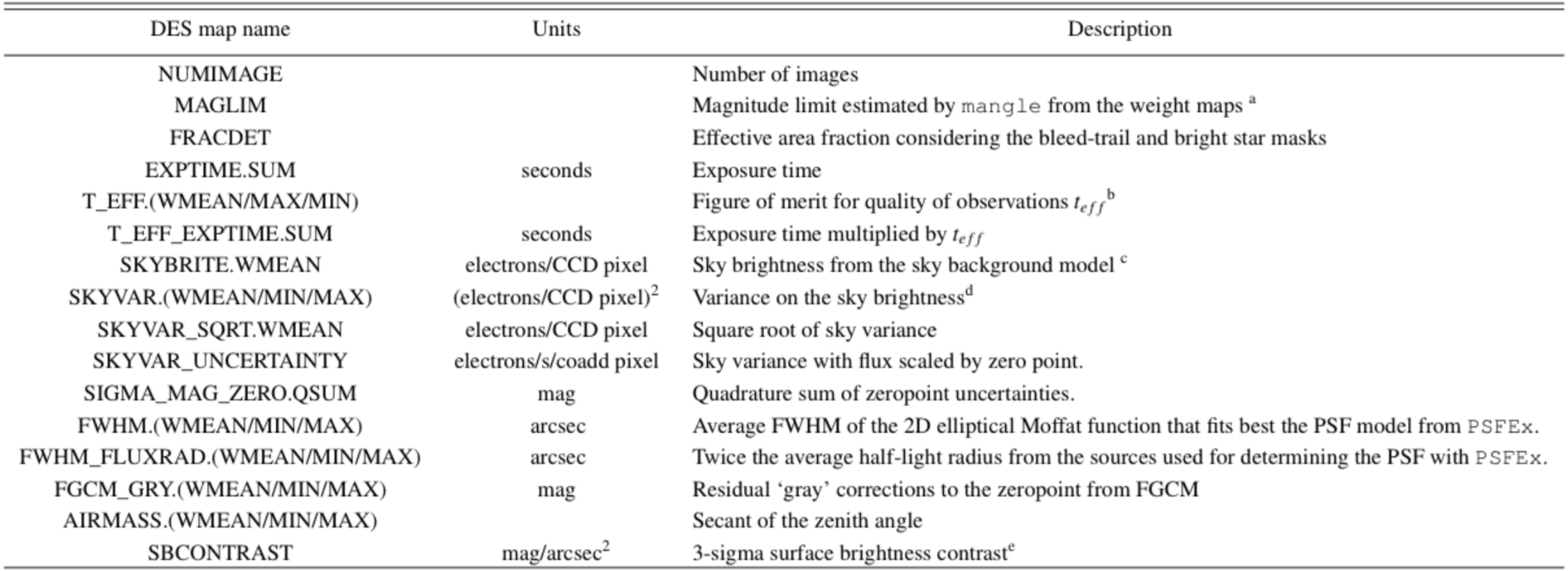


A.Drlica-Wagner, I.S-N et al. 2018

B.Leistedt, ..., I.S-N et al. 2016

Pixelization is realized through Mangle (Tegmark et al. 2005) + Healpix (Gorski et al. 2005)

Survey property maps should include any physical circumstance affecting the observation



DES map name	Units	Description
NUMIMAGE		Number of images
MAGLIM		Magnitude limit estimated by `mangle` from the weight maps [a]
FRACDET		Effective area fraction considering the bleed-trail and bright star masks
EXPTIME.SUM	seconds	Exposure time
T_EFF.(WMEAN/MAX/MIN)		Figure of merit for quality of observations t_{eff}[b]
T_EFF_EXPTIME.SUM	seconds	Exposure time multiplied by t_{eff}
SKYBRITE.WMEAN	electrons/CCD pixel	Sky brightness from the sky background model [c]
SKYVAR.(WMEAN/MIN/MAX)	(electrons/CCD pixel)^2	Variance on the sky brightness[d]
SKYVAR_SQRT.WMEAN	electrons/CCD pixel	Square root of sky variance
SKYVAR_UNCERTAINTY	electrons/s/coadd pixel	Sky variance with flux scaled by zero point.
SIGMA_MAG_ZERO.QSUM	mag	Quadrature sum of zeropoint uncertainties.
FWHM.(WMEAN/MIN/MAX)	arcsec	Average FWHM of the 2D elliptical Moffat function that fits best the PSF model from `PSFEx`.
FWHM_FLUXRAD.(WMEAN/MIN/MAX)	arcsec	Twice the average half-light radius from the sources used for determining the PSF with `PSFEx`.
FGCM_GRY.(WMEAN/MIN/MAX)	mag	Residual 'gray' corrections to the zeropoint from FGCM
AIRMASS.(WMEAN/MIN/MAX)		Secant of the zenith angle
SBCONTRAST	mag/arcsec^2	3-sigma surface brightness contrast[e]

I.S-N et al. 2021

Add to this stellar maps and extinction.

Mitigation can follow several strategies

A few approaches proposed (compared in A.Ross et al. 2011)

selective elimination of area

$$\omega^{\mathrm{true}}(\theta) \simeq \omega^{\mathrm{obs}}(\theta) - \cancel{\sum_i \sum_j \varepsilon_i \varepsilon_j \langle \delta_i^{\mathrm{sys}} \delta_j^{\mathrm{sys}} \rangle}$$

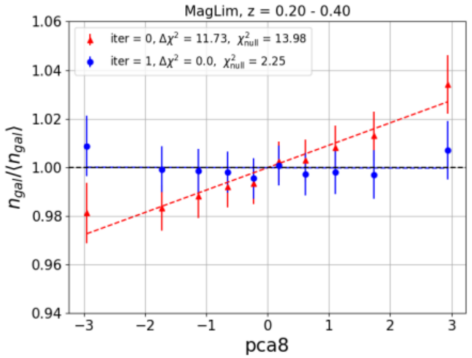
- Masking (Myers et al. 2006)

computable with sys X obs

$$\omega^{\mathrm{true}}(\theta) = \omega^{\mathrm{obs}}(\theta) - \sum_i \sum_j \varepsilon_i \varepsilon_j \langle \delta_i^{\mathrm{sys}} \delta_j^{\mathrm{sys}} \rangle$$

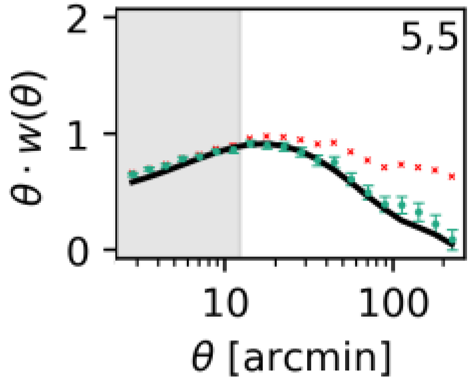
- Correcting (e.g. M.Crocce et al. 2015 using cross-correlations)
- Weighting (e.g. J.Elvin-Poole et al. 2018, M.Rodríguez-Monroy et al. 2022 on galaxies)

1. Establish 1D relationship of systematic and density (red line)

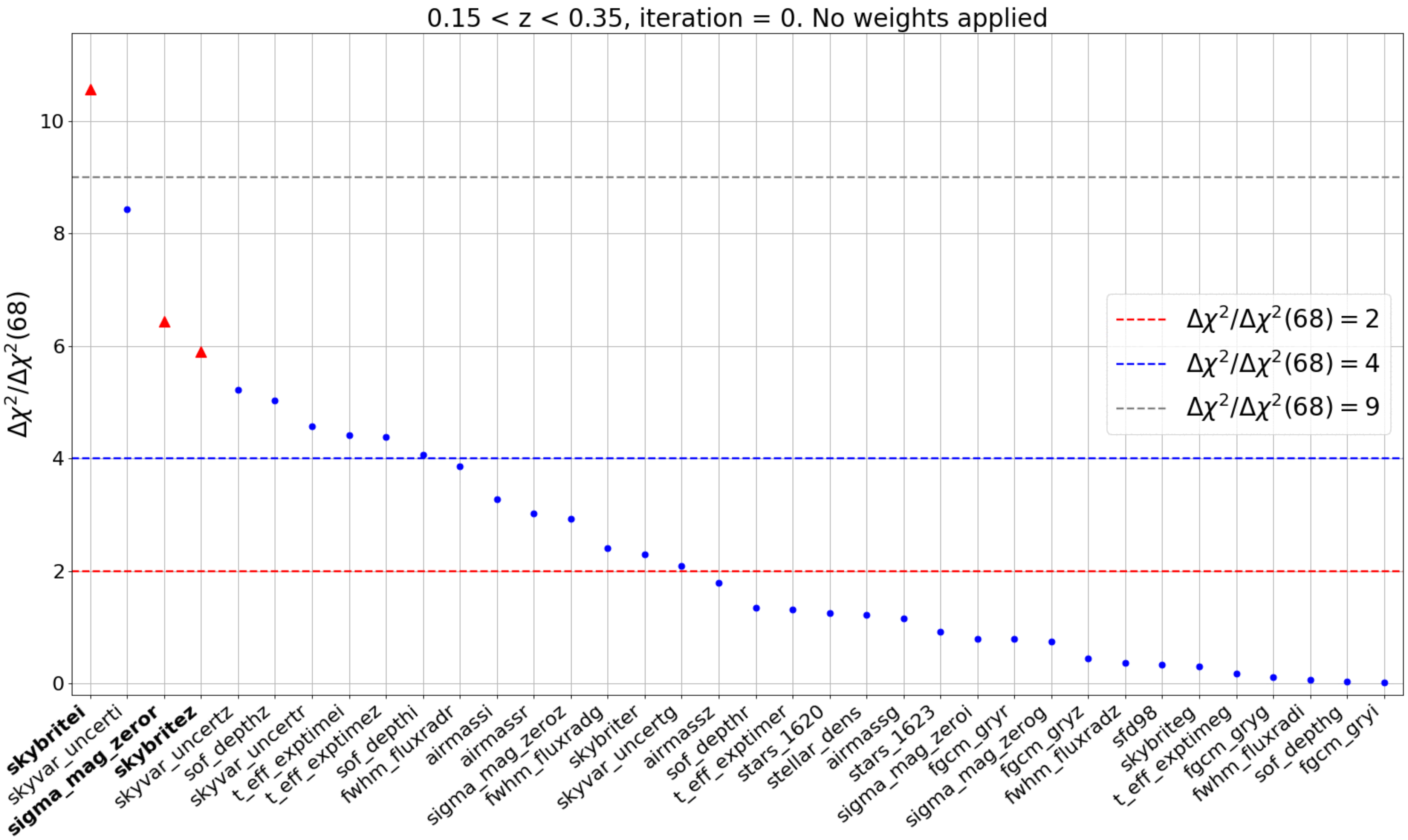


2. Compute weight for galaxies to ‘level’ this relationship (blue line)

3. Estimate correlation function applying these weights to galaxies



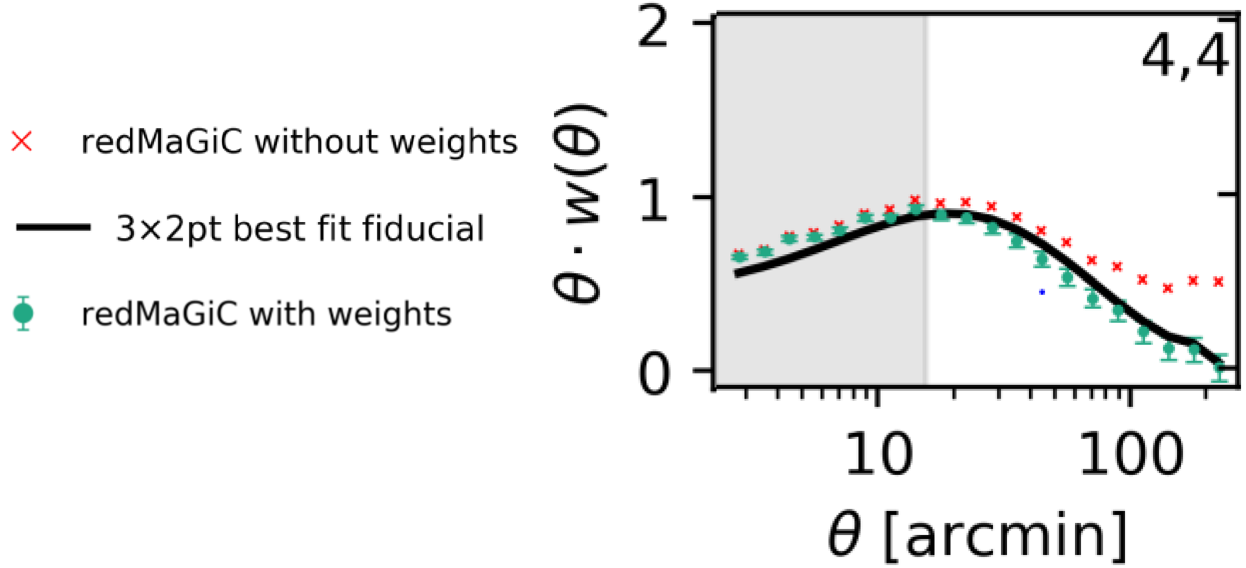
My prediction is that this gif won’t work



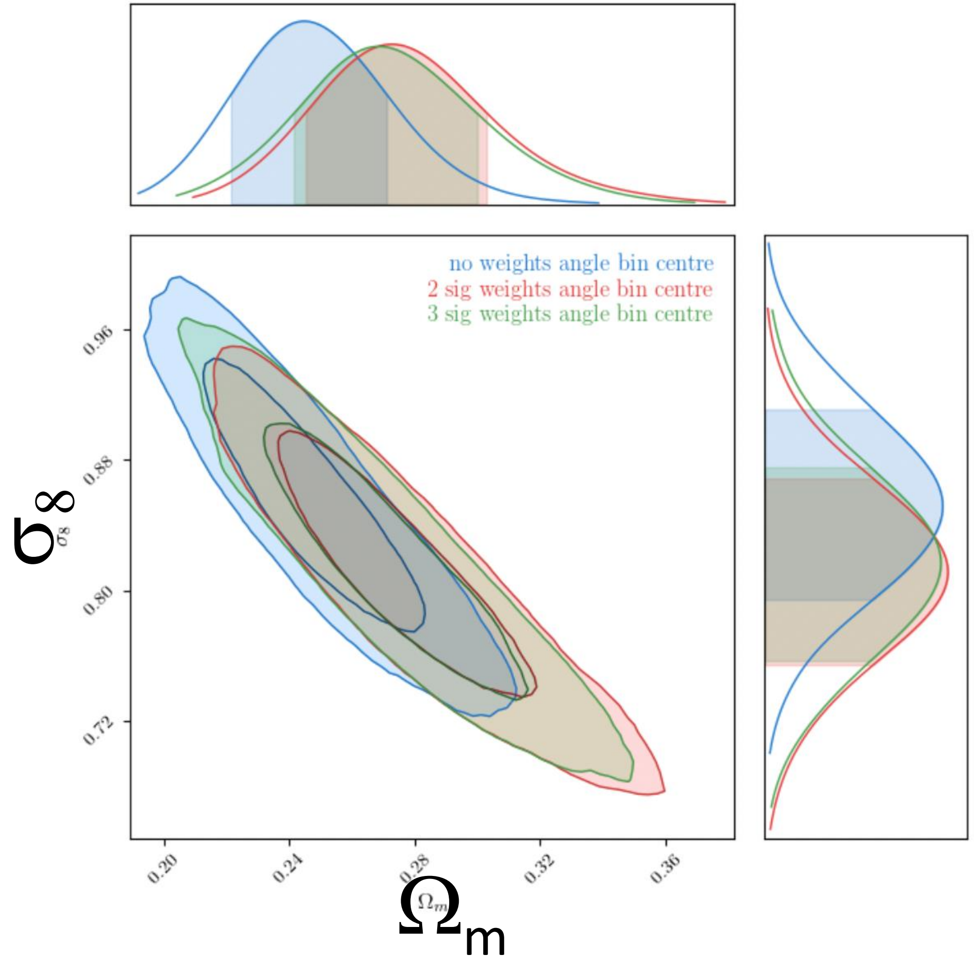
The impact in clustering is significant on derived cosmological parameters

We can have >~ 2σ shifts in parameters.

Primordial non-Gaussianities signal can be completely washed out (M. Rezaie et al. 2021)



M.Rodríguez-Monroy et al. 2022



J.Elvin-Poole PhD thesis, DES Y1

In the land of Cls…

Equivalently, we can perform Template Subtraction (B.Leistedt et al. 2013, B.Leistedt & H.Peiris 2014, F.Elsner et al. 2015):

$$C_l^{true} = C_l^{obs} - \alpha_l^2 C_l^{sys} \times (some\ debiasing\ factor)$$

Or Mode Projection (eg D. Alonso, F.J. Sánchez & A. Slosar 2019):

$$\mathbf{C} \rightarrow \mathbf{C} + \sum_{i=1}^{N} \lim_{\sigma \rightarrow \infty} (\sigma_i t_i t_i^T)$$

All equivalent to among them (N.Weaverdyck & D. Huterer, 2020)

In DES Y3 we checked with additional methods for increased robustness

Other methods were compared in DES (M.Rodríguez-Monroy et al. 2022):

ENet (N. Weaverdyck & D. Huterer 2020)

All previous methods:

Minimize $\|\delta_{obs} - T\alpha\|^2$: find α

ENET (Zou & Hastie 2005):

Incentivize few templates

Minimize correlation

Minimize $\frac{1}{2N}\|\delta_{obs} - T\alpha\|_2^2 + \lambda_1\|\alpha\|_1 + \frac{\lambda_2}{2}\|\alpha\|_2$: find α

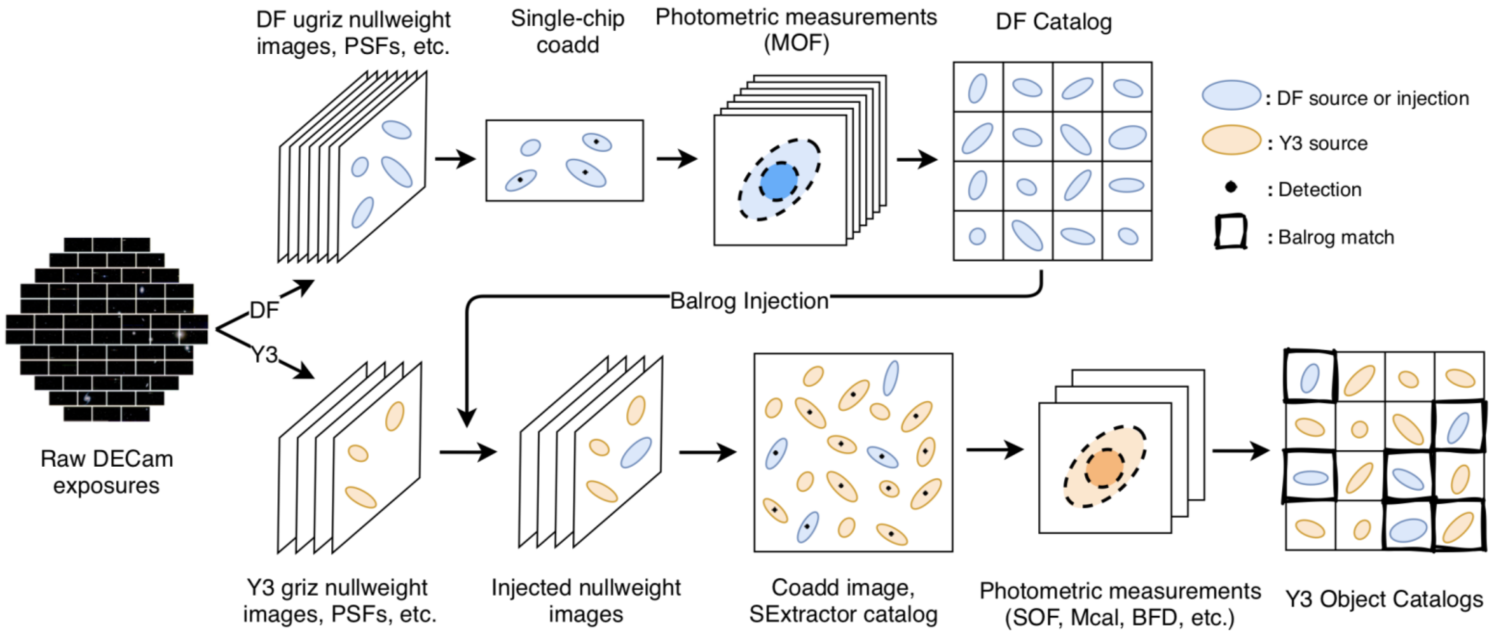
SYSNet (Rezaie et al. 2020, Rezaie et al. 2021)

Model the relationship between galaxy density and survey properties through a neural network.

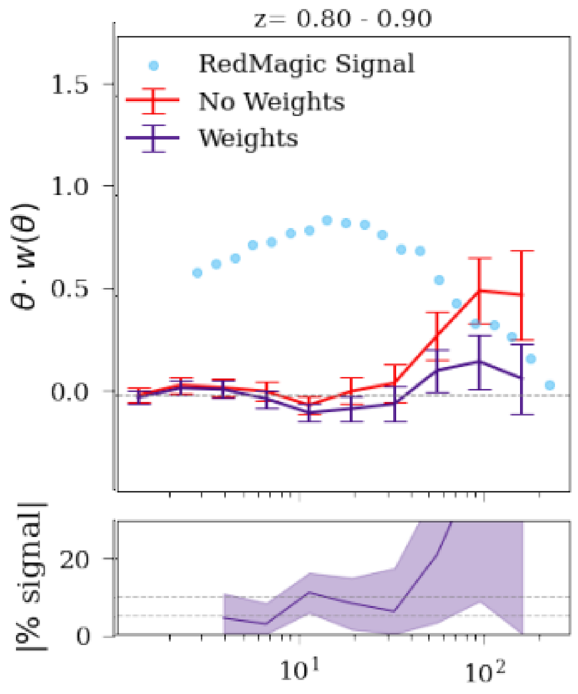
Create weights based on this trained relationship, without considering positional information (so no clustering is added)

Last few years has seen an explosion of new methods

- Multilinear simultaneous regression (Bautista et al. 2018, M.Vakili et al. 2020)
- Map galaxy density to Self Organized Maps (H. Johnston et al. 2021)
- Machine learning predictions (E. Wagoner et al. 2020)
- Use of weights through randoms (C. Morrison & H. Hildebrandt, 2015; M. Rodríguez-Monroy & I.S-N 2018)
- Realistic injections of simulated data into real images to obtain the transfer function: Balrog (E.Suchyta et al. 2016, S.Everett et al. 2020; M.García-Fernández, E.Sánchez, I.S-N et al. 2017)



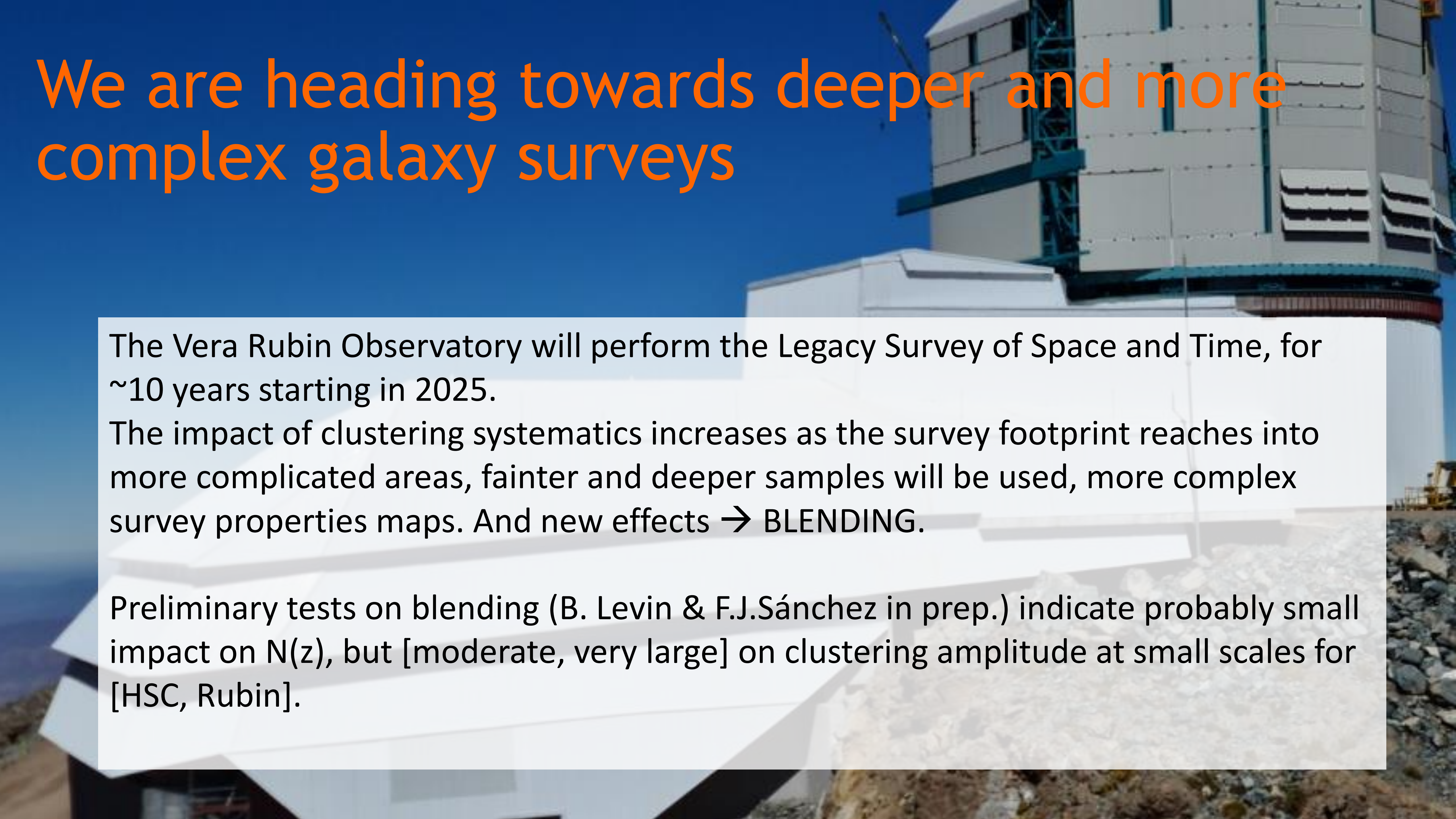
S.Everett et al. 2020



Recipe for success

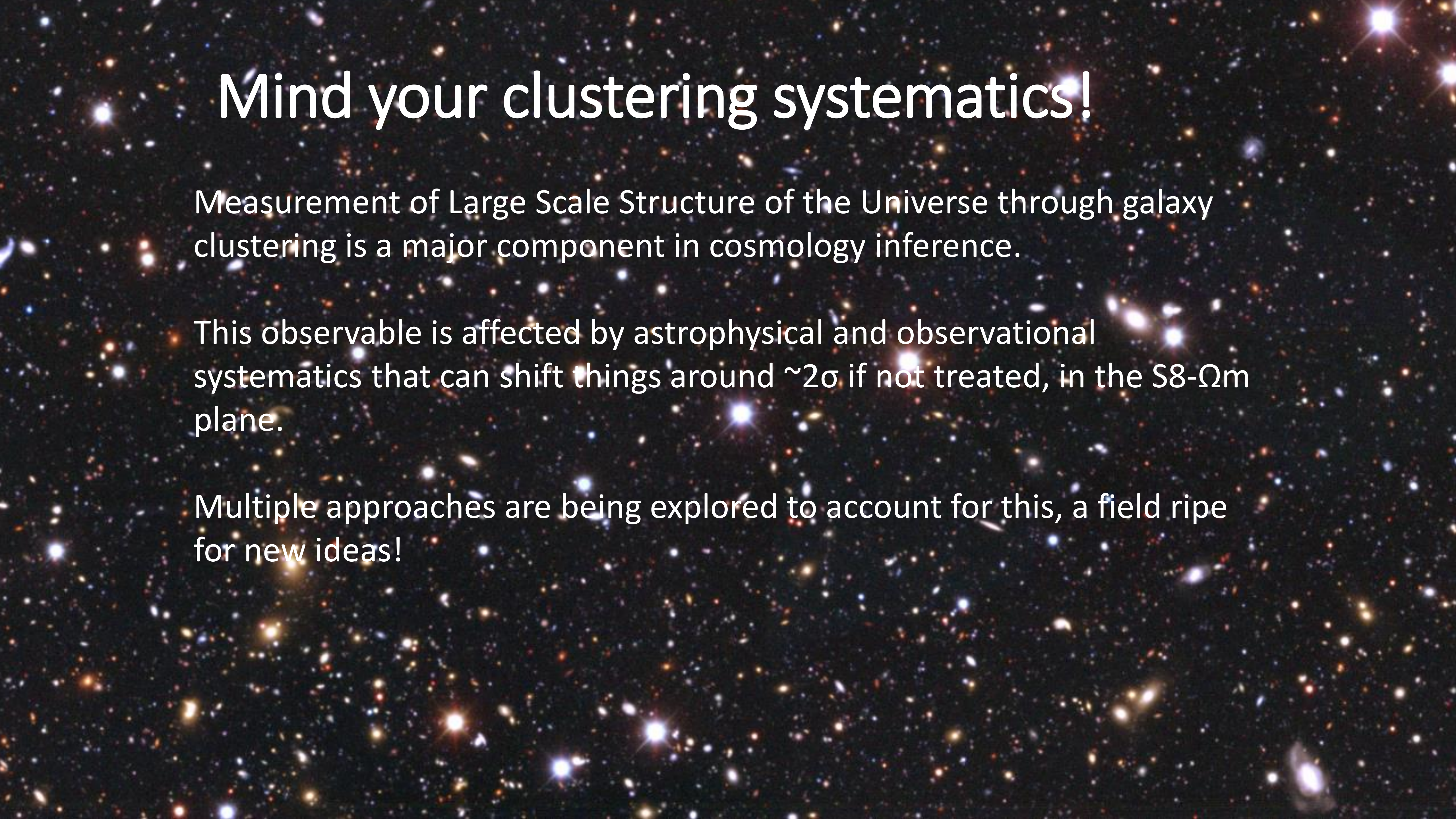
- Verify correlations **in different areas of the survey**
- Use **different methodologies** and verify consistency
- Incorporate systematic effects on **simulations** or mocks and then decontaminate them
- **Predict with other probes** (Posterior Predictive Distribution method, Gelman 1996)
- **Cross-correlation of weights with external mass estimates** (other surveys, convergence maps)
- **Bias predictions** of gg lensing vs w(theta)
- Use a variety of **samples**

We are heading towards deeper and more complex galaxy surveys



The Vera Rubin Observatory will perform the Legacy Survey of Space and Time, for ~10 years starting in 2025.
The impact of clustering systematics increases as the survey footprint reaches into more complicated areas, fainter and deeper samples will be used, more complex survey properties maps. And new effects → BLENDING.

Preliminary tests on blending (B. Levin & F.J.Sánchez in prep.) indicate probably small impact on N(z), but [moderate, very large] on clustering amplitude at small scales for [HSC, Rubin].



Mind your clustering systematics!

Measurement of Large Scale Structure of the Universe through galaxy clustering is a major component in cosmology inference.

This observable is affected by astrophysical and observational systematics that can shift things around ~2σ if not treated, in the S8-Ωm plane.

Multiple approaches are being explored to account for this, a field ripe for new ideas!