Digging Deeper (and More Greedily) in Imaging Surveys

Eric Suchyta
University of Pennsylvania
(In collaboration with Jelena Aleksić, Eric Huff, Niall MacCrann)

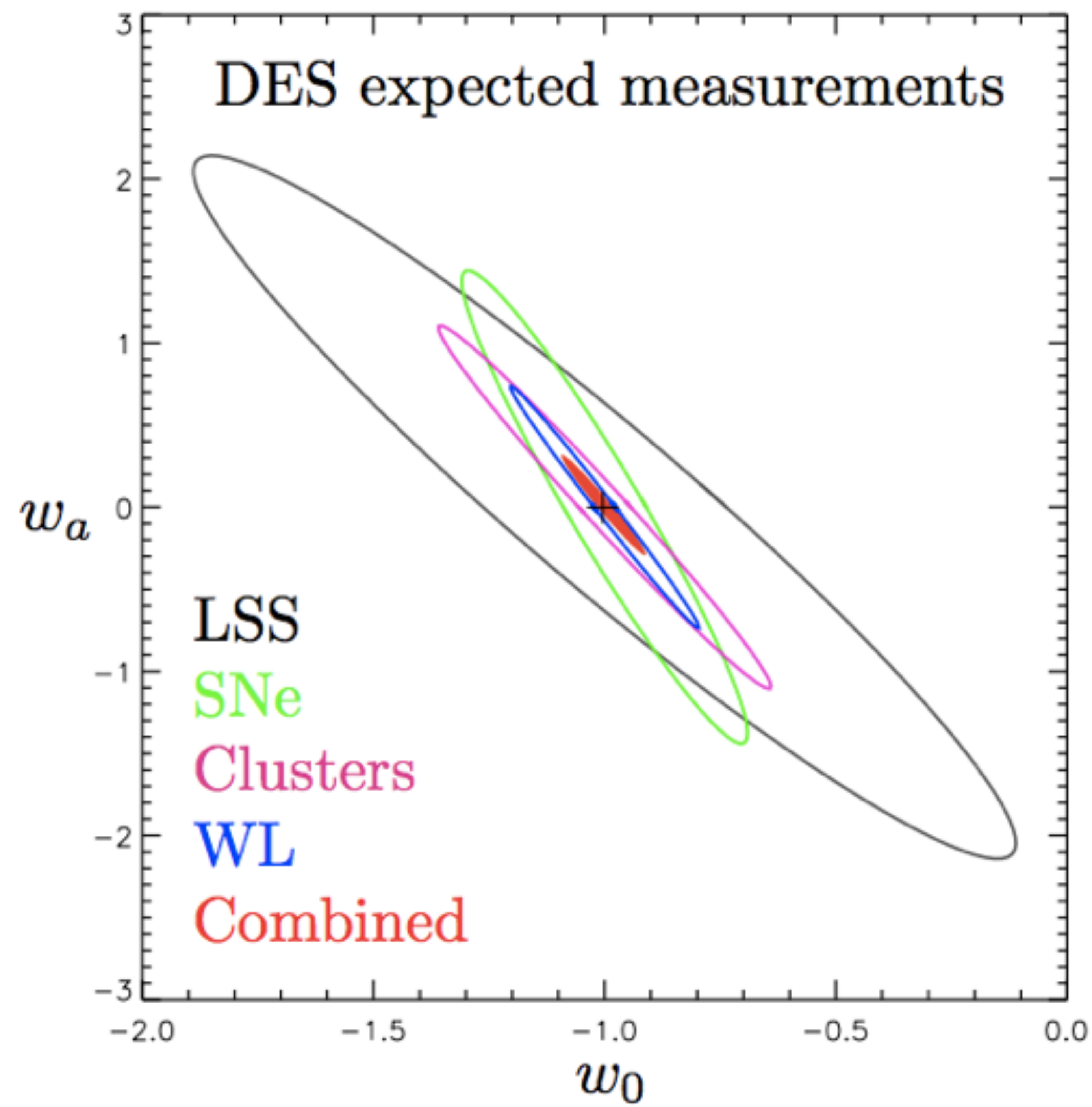
IFAE Pizza Seminar: 14 October 2015

The era of survey astronomy

- Sloan Digital Sky Survey (SDSS): nearly 6,000 publications
- Surveys are getting larger and larger:
  - Dark Energy Survey (DES) / Kilo-Degree Survey (KiDS): ~300 million galaxies
  - Large Synoptic Survey Telescope (LSST) / Euclid: ~10 billion galaxies

"With great statistical power comes great systematic error responsibility"

- DES statistical limits: ~percent-level uncertainty for constant dark energy equation of state
- Must limit systematic errors well below that
- Maximizing the return requires working with low S/N objects



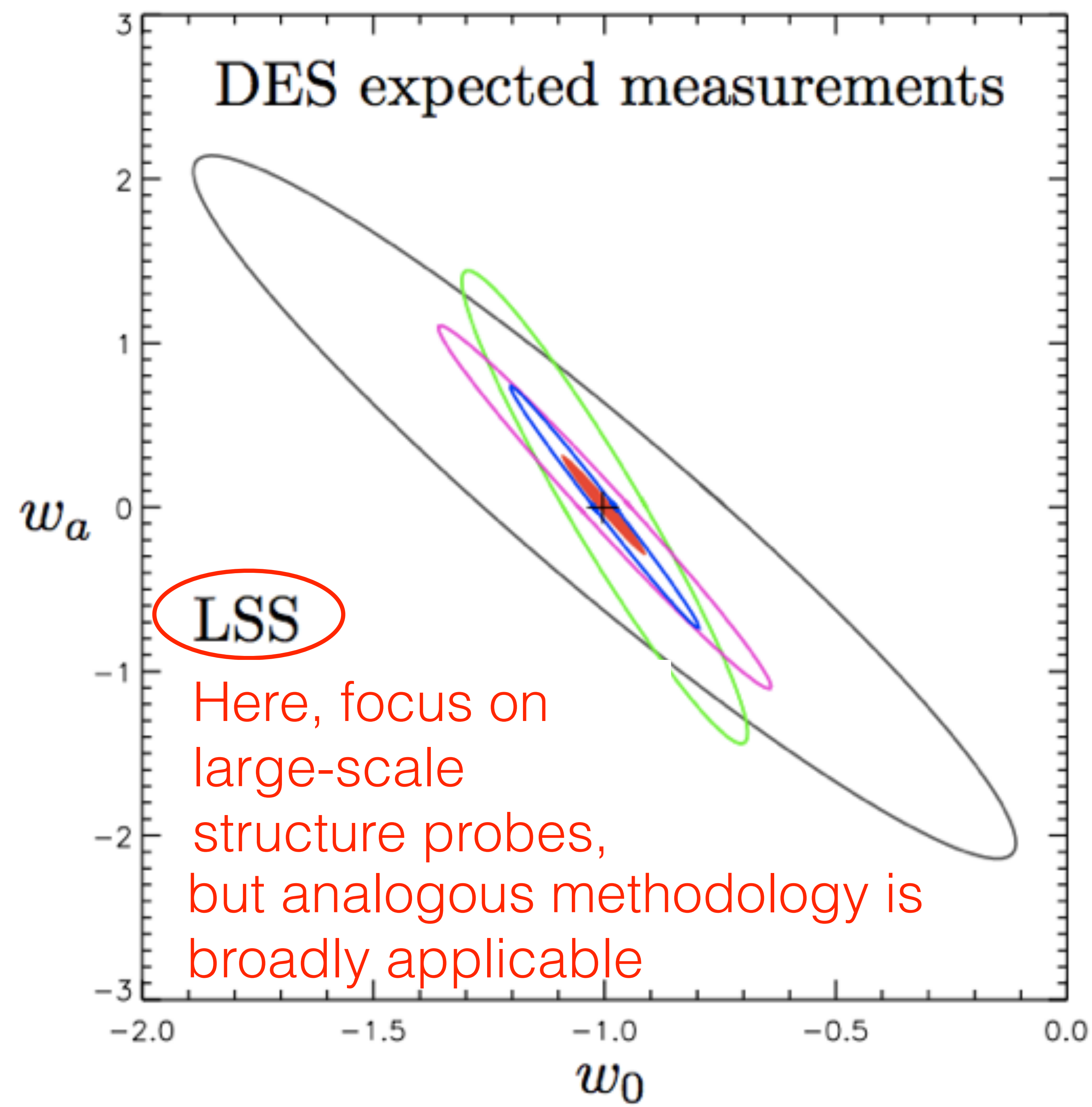
[Credit: DES collaboration]

“With great statistical power comes great systematic error responsibility”

- DES statistical limits: ~percent-level uncertainty for constant dark energy equation of state

- Must limit systematic errors well below that
- Maximizing the return requires working with low S/N objects

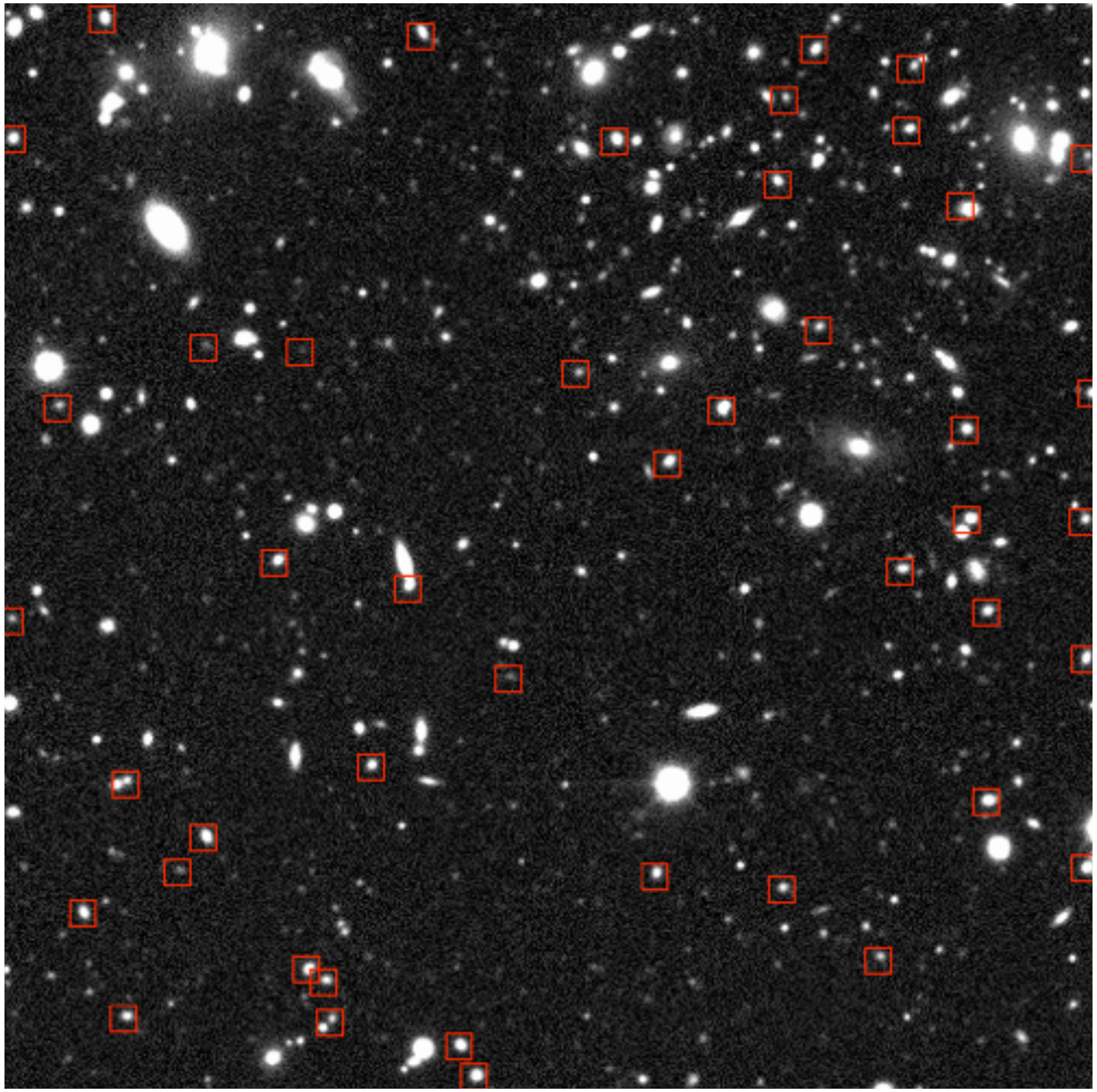
Designed new toolkit to help, called BALROG



[Credit: DES collaboration]

BALROG

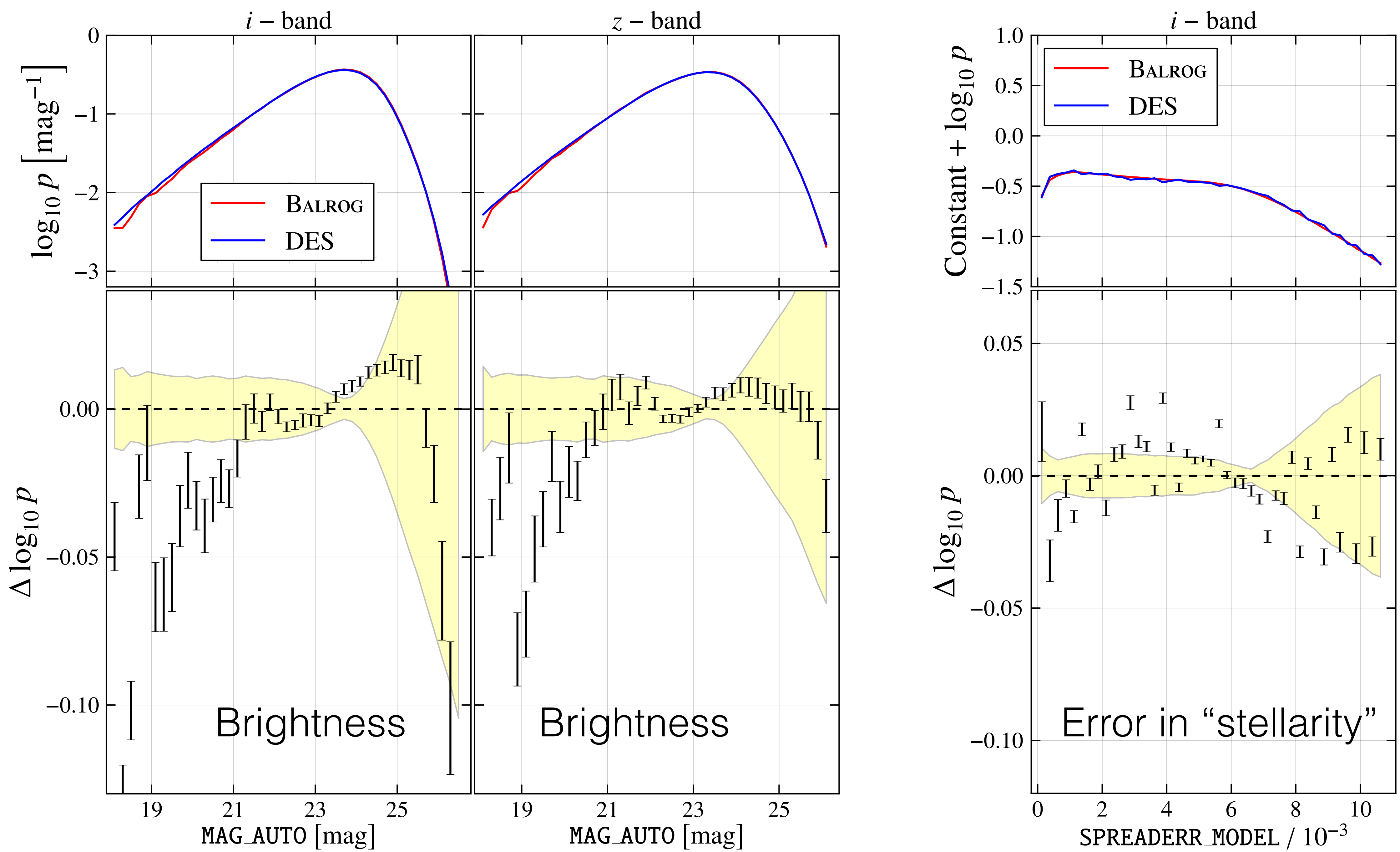
- Insert simulated objects into real DES images
- Apply measured detector responses to objects (e.g. flux calibration, optical distortions)
- Sample simulated ensemble from space-based measurements
- Run full measurement pipeline
- Investigate systematic biases in the output catalog



Red = simulated

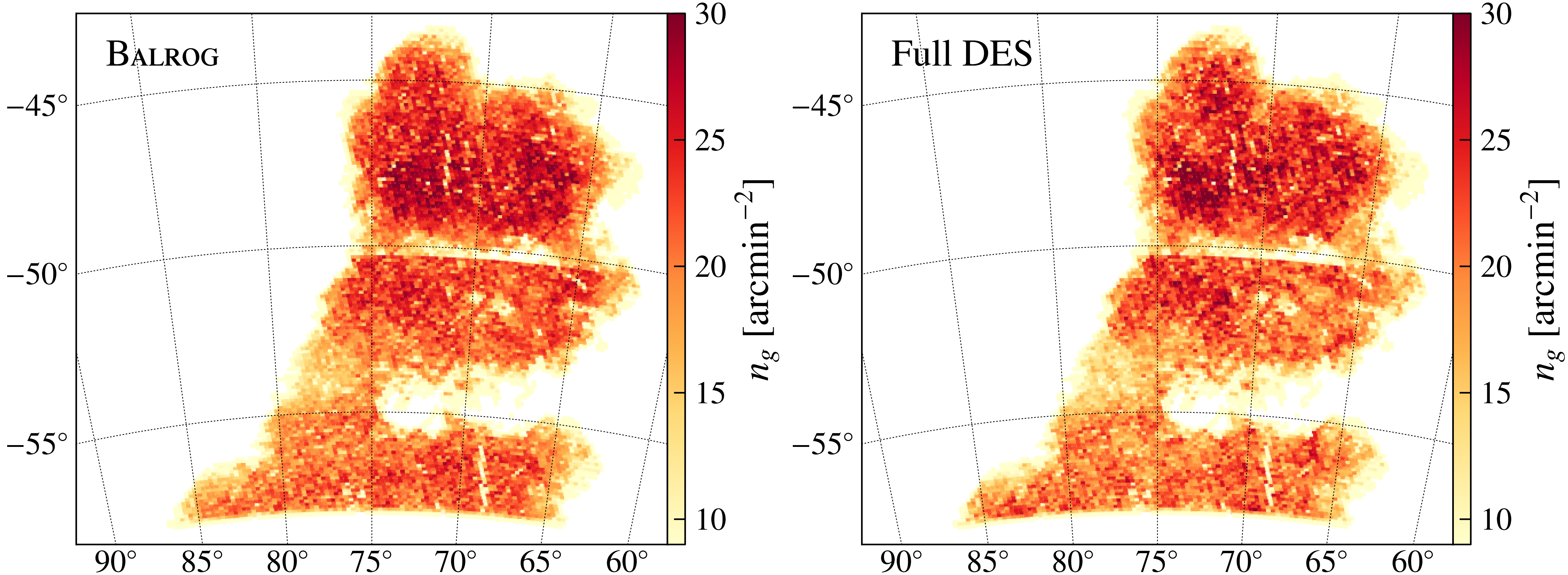
BALROG well captures DES behavior

- Agreement holds across bands and different measurements



[Suchyta et al. 2015]

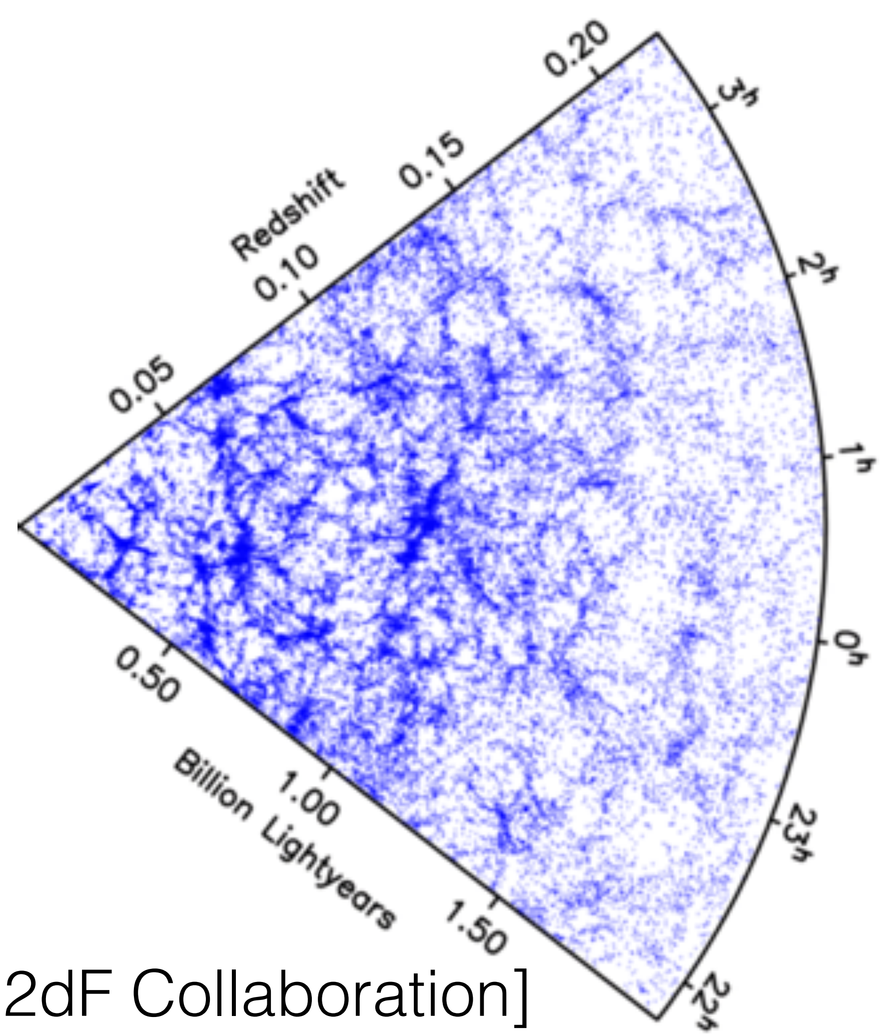
Use BALROG detections as a Monte-Carlo sampling of the survey detection probability, to remove systematic bias in large-scale structure angular clustering measurements



[Suchyta et al. 2015]

Looking to recover clustering — tendency for galaxies to clump together

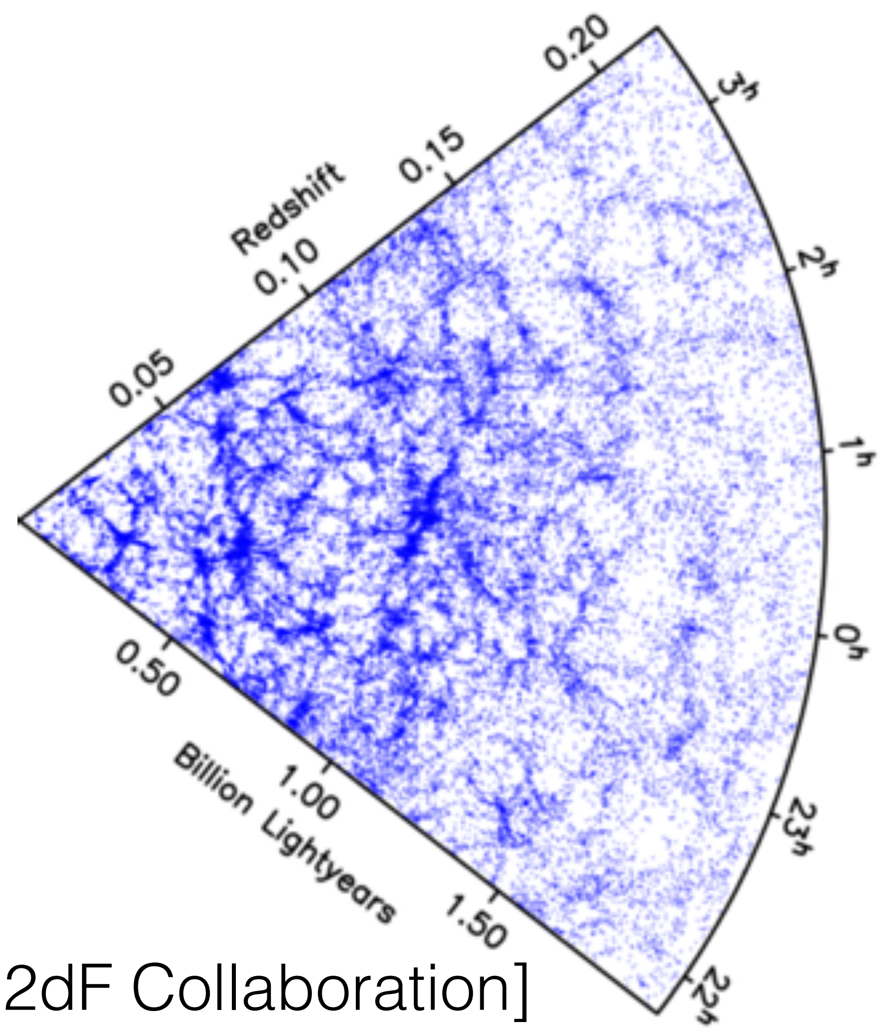
- Measures variability in number density: $w(\theta = |\boldsymbol{\vartheta} - \boldsymbol{\vartheta}'|) = \dfrac{\langle n(\boldsymbol{\vartheta})\, n(\boldsymbol{\vartheta}')\rangle}{\bar{n}_g^2} - 1$



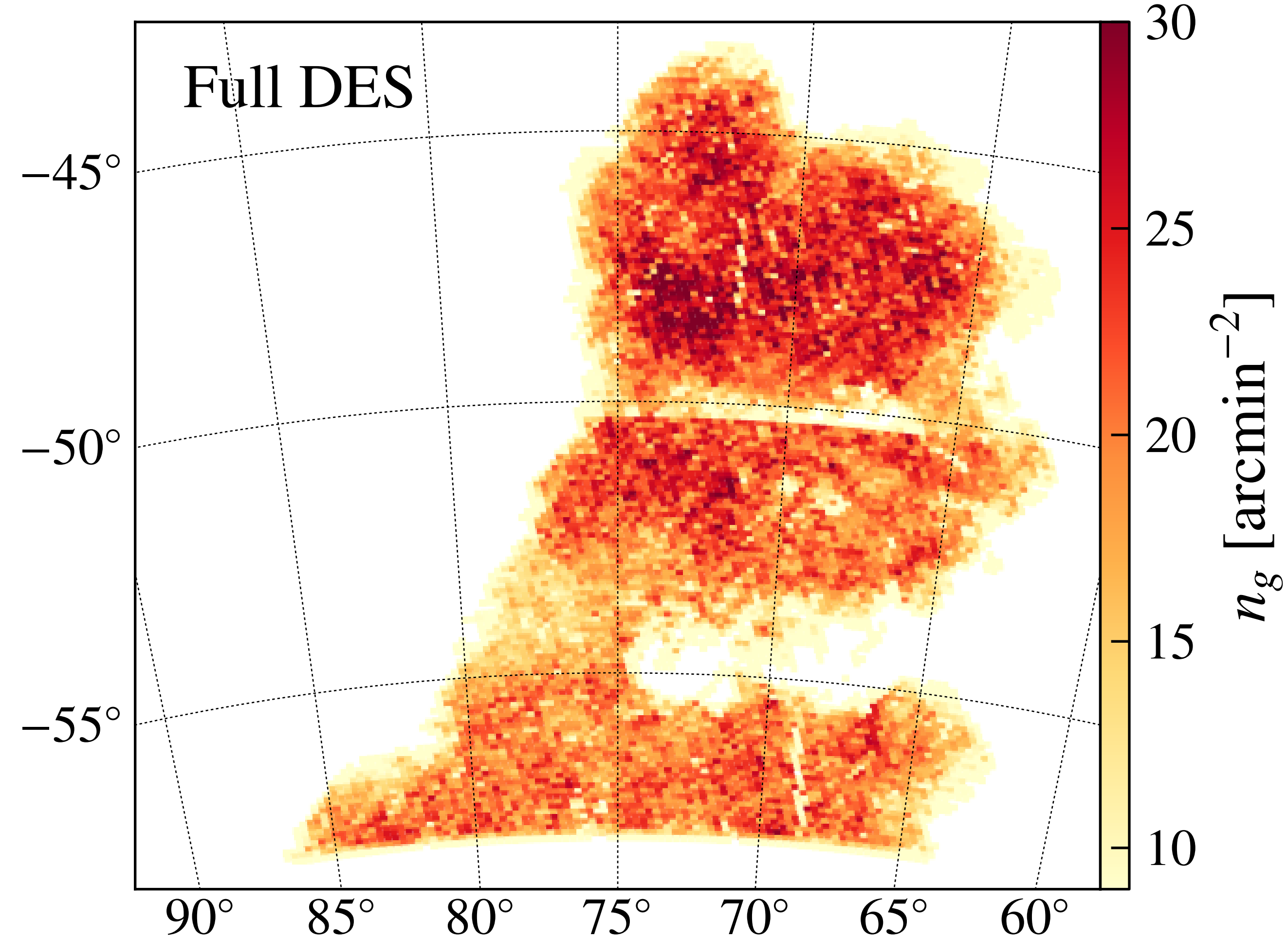
[Credit: 2dF Collaboration]

Looking to recover clustering — tendency for galaxies to clump together

- Measures variability in number density: $w(\theta = |\boldsymbol{\vartheta} - \boldsymbol{\vartheta}'|) = \frac{\langle n(\boldsymbol{\vartheta})\, n(\boldsymbol{\vartheta}')\rangle}{\bar{n}_g^2} - 1$
- $w(\theta)$ ~ (0.01-0.1), and only appreciable on variation scales smaller the variation scale seen in DES image; image should look nearly random



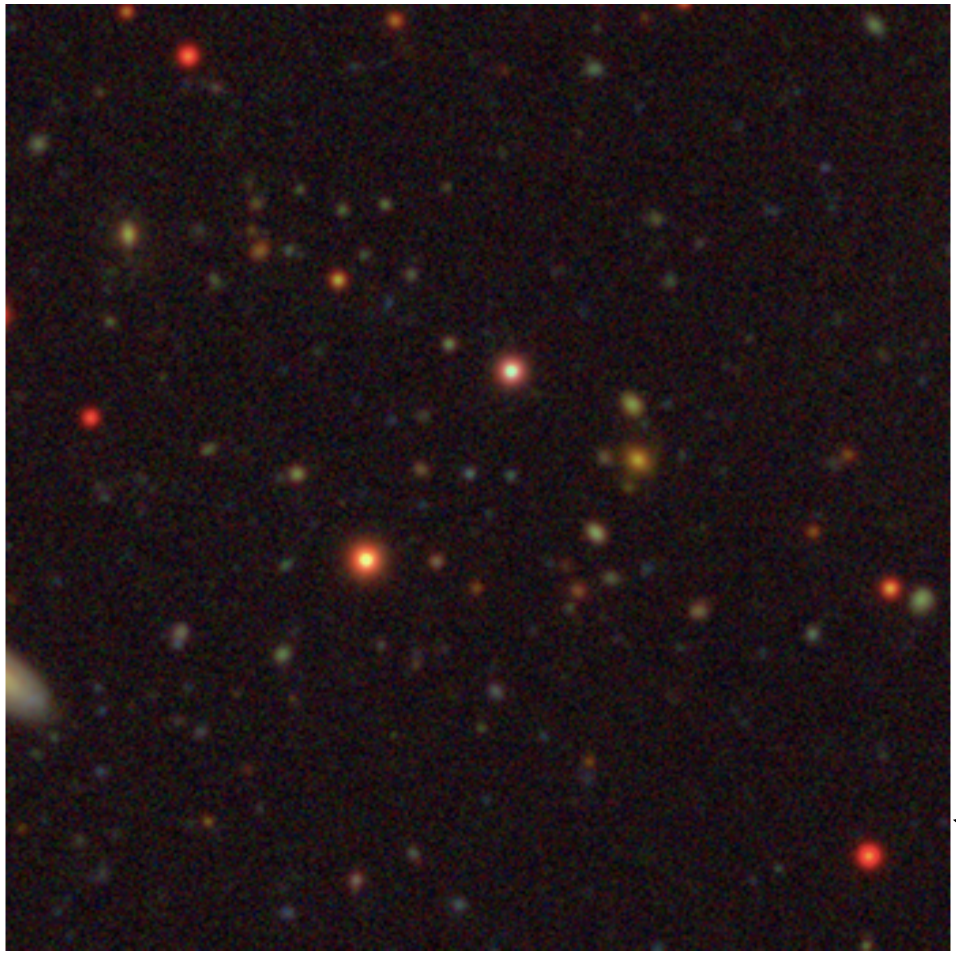
[Credit: 2dF Collaboration]

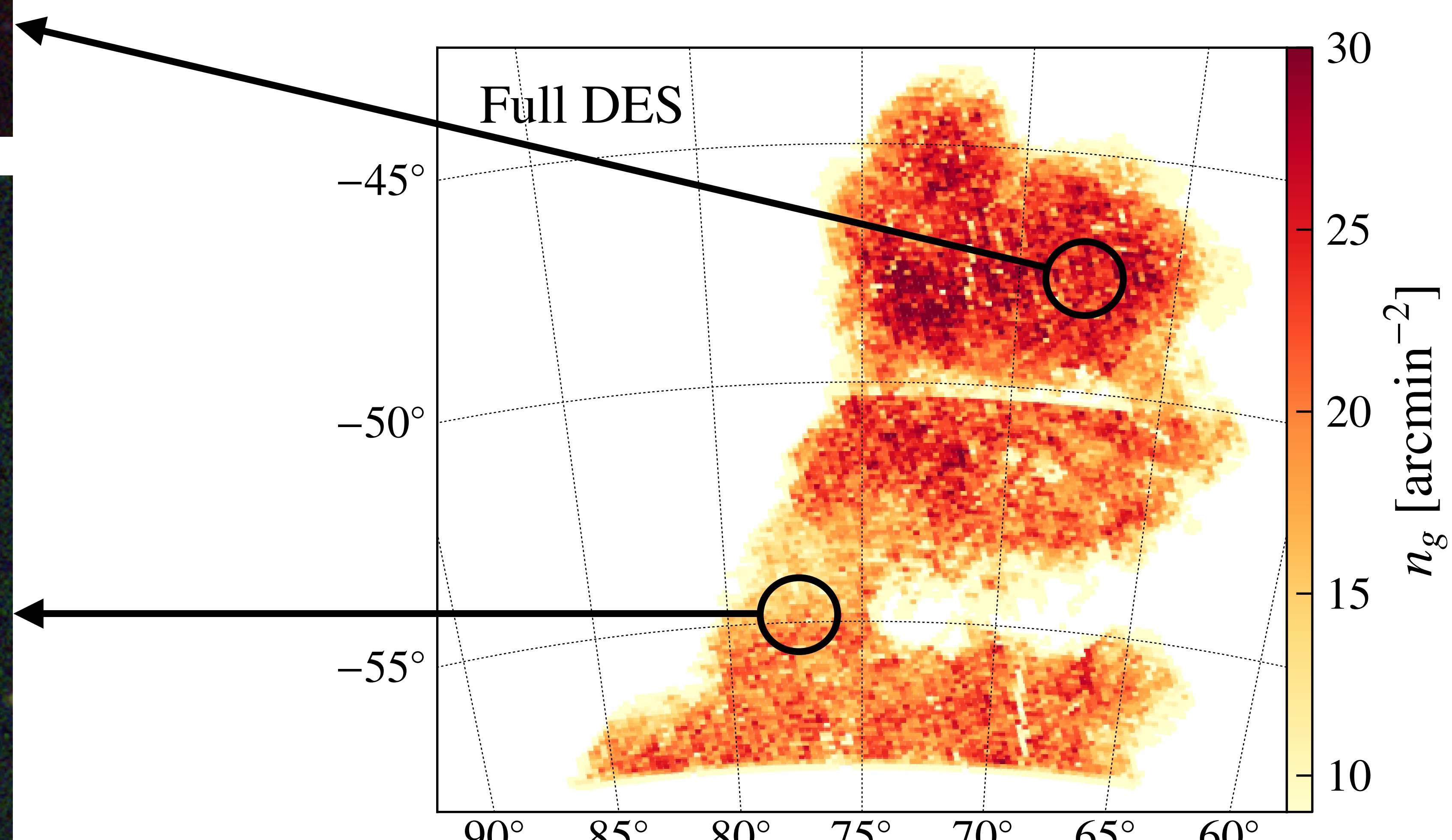


[Suchyta et al. 2015]

Variation seen is not signal; just survey inhomogeneities

- Changes in sky brightness, atmospheric turbulence, etc. change detection rate

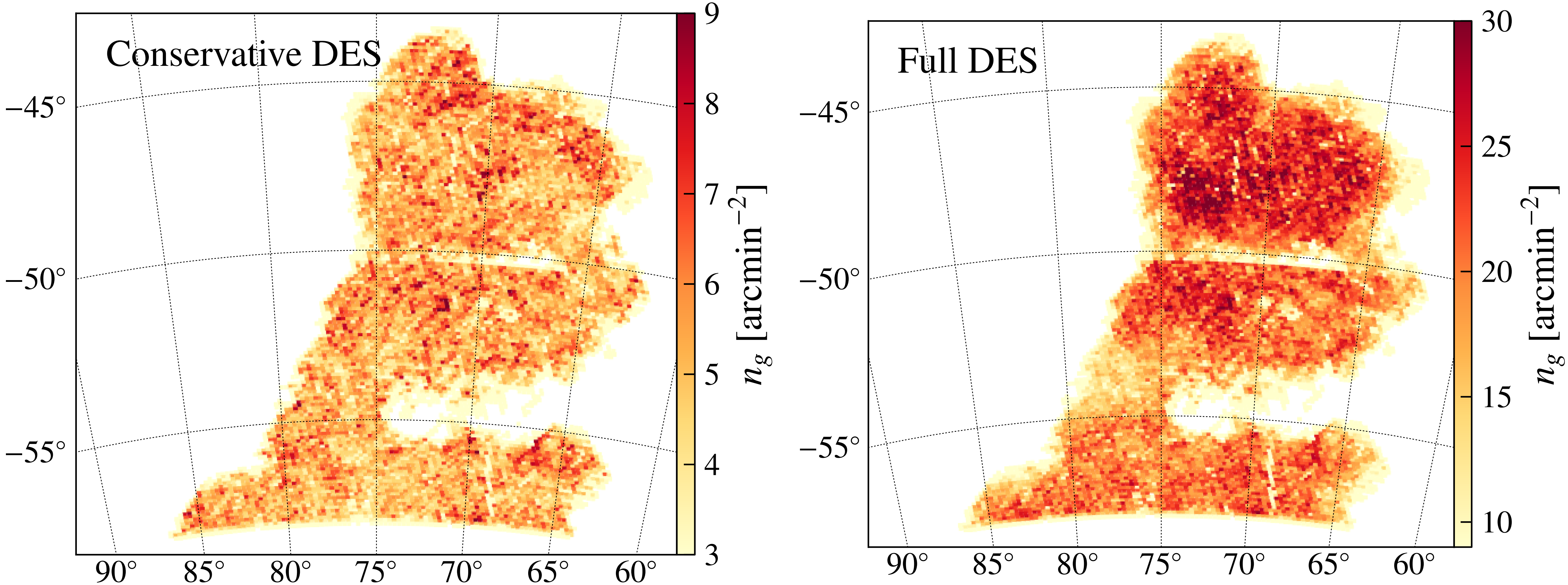




[Suchyta et al. 2015]

The usual way to recover the signal is to use complete sample — keep only galaxies bright enough that they are nearly always detected

- Eliminates need to model spatially varying completeness
- But removes much of the sample, degrading statistical precision

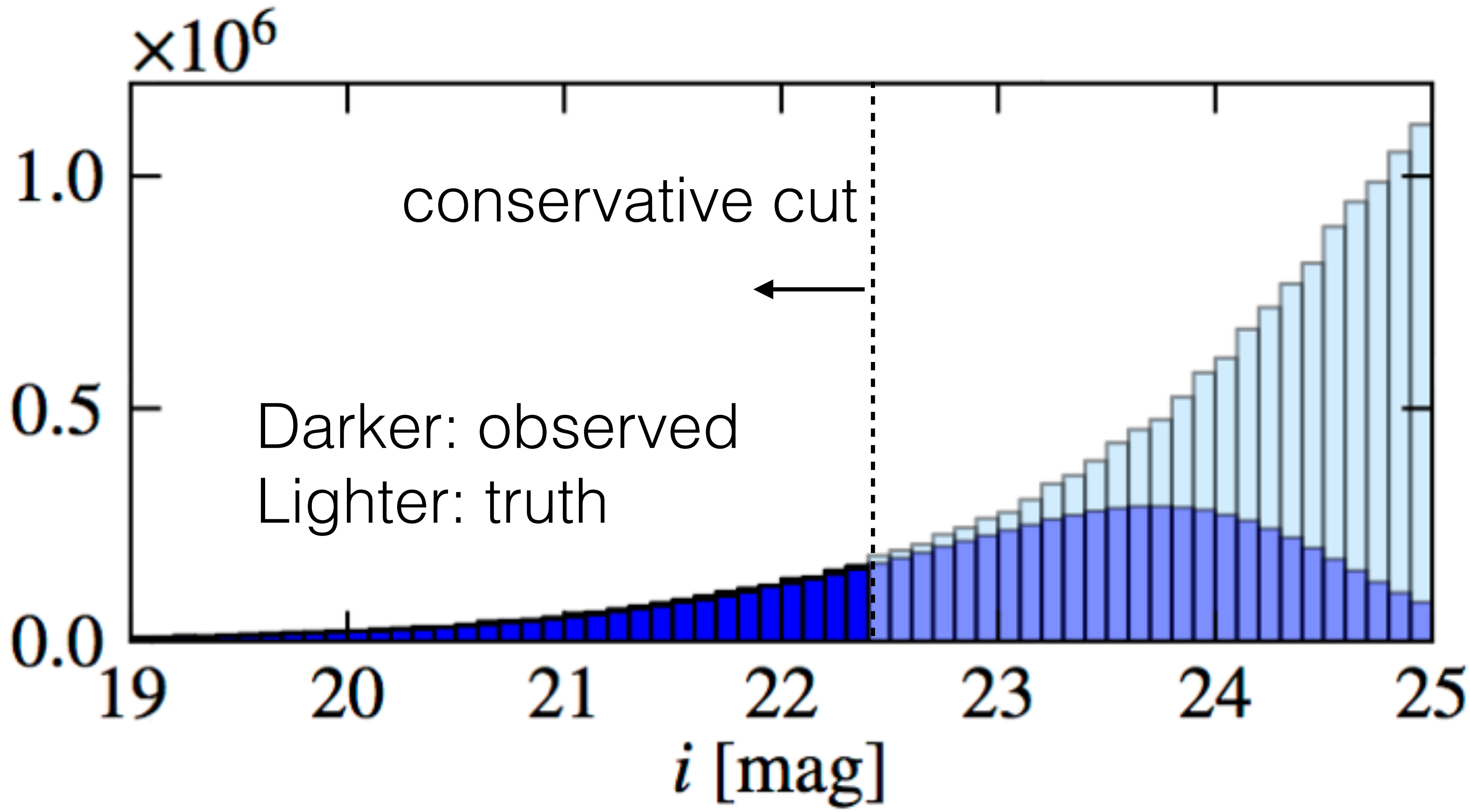


[Crocce et al. 2015]

[Suchyta et al. 2015]

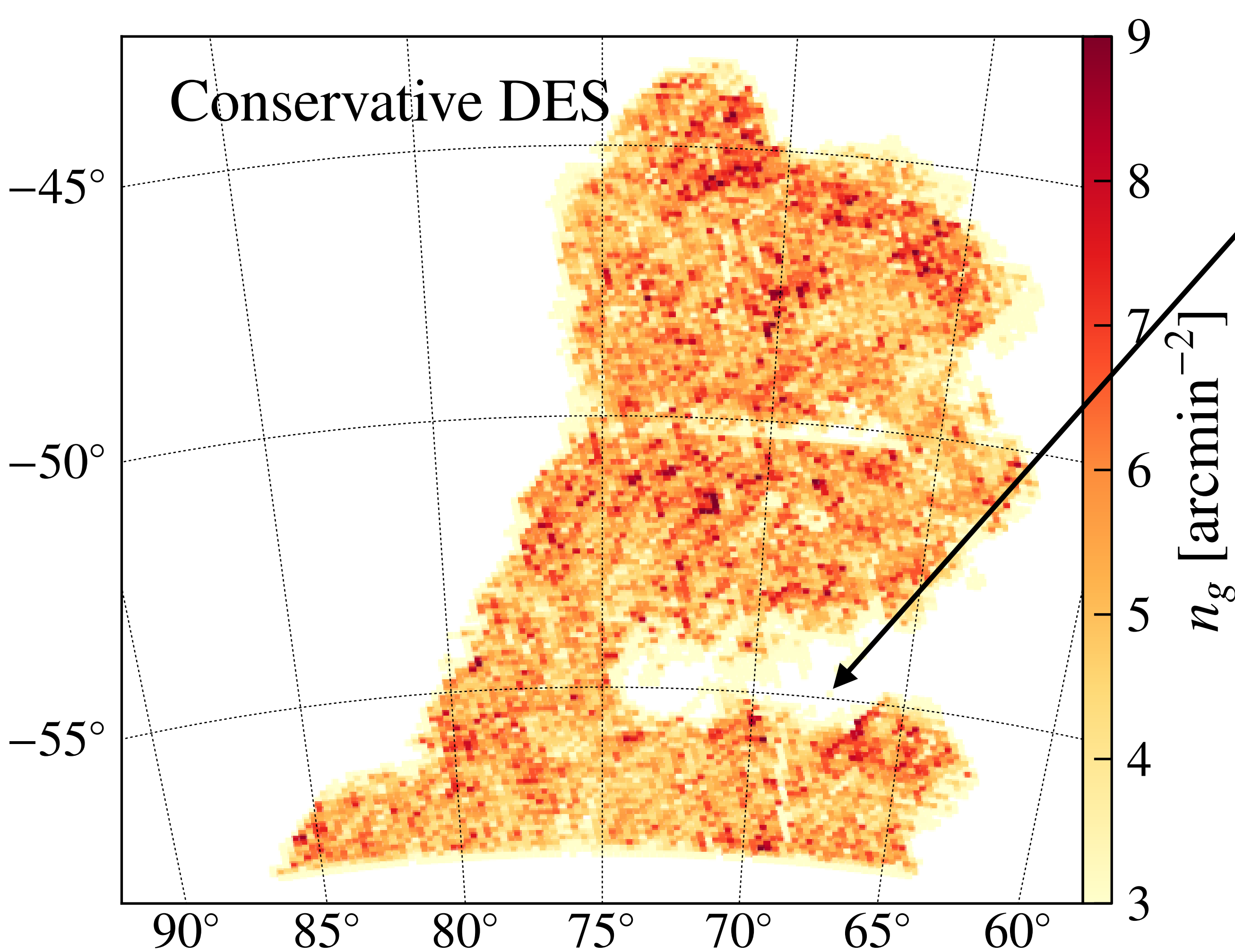
The usual way to recover the signal is to use complete sample — keep only galaxies bright enough that they are nearly always detected

- Eliminates need to model spatially varying completeness
- But removes much of the sample, degrading statistical precision



The usual way to recover the signal is to use complete sample — keep only galaxies bright enough that they are nearly always detected

- Eliminates need to model spatially varying completeness
- But removes much of the sample, degrading statistical precision



[Crocce et al. 2015]

- Area remaining incomplete after cut is masked — has imprinted variation; areas with no galaxies
- Measure $w(\theta)$ as excess counts relative to random points with same mask applied:

$$w(\theta) = \frac{DD - 2DR + RR}{RR}$$

D : data point
R : random point — uniform prior to masking

[Landy & Szalay 1993]

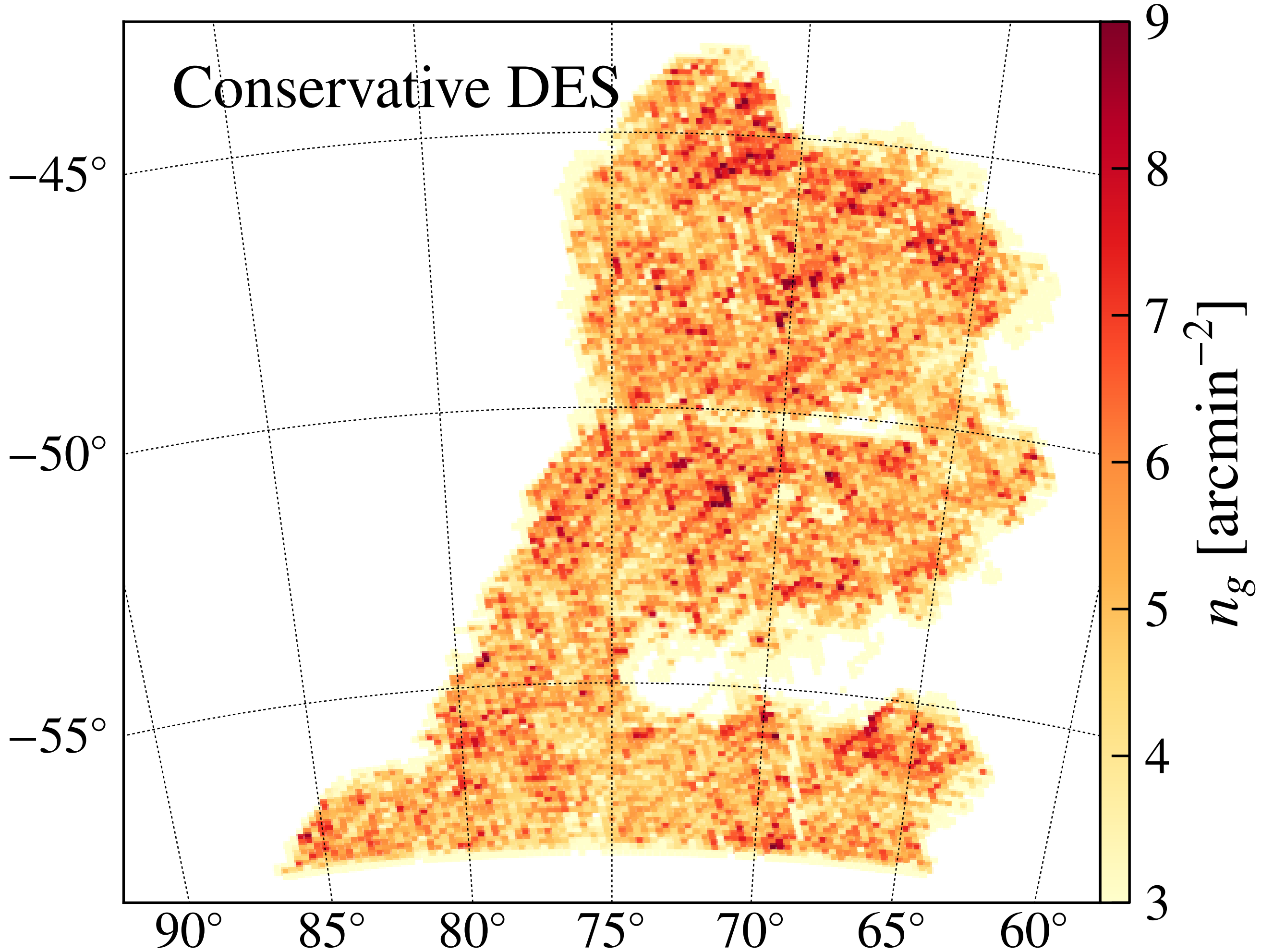
Our approach: use Balrog detections as randoms

$$w(\theta) = \frac{DD - 2DR + RR}{RR}$$

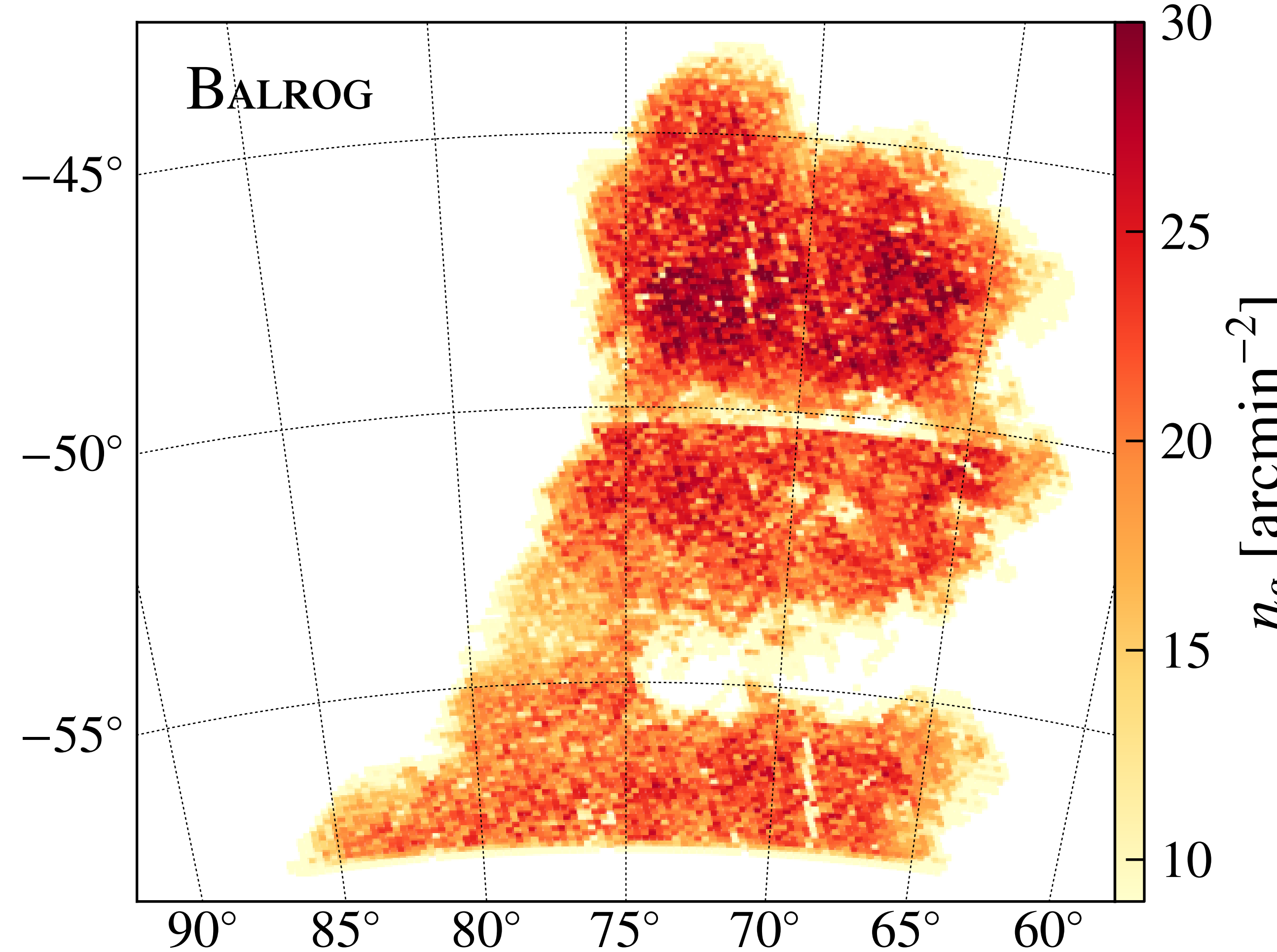
[Landy & Szalay 1993]

$\boldsymbol{D}$: data point

$\boldsymbol{R}$: random point — now accounts for spatially varying detection probability



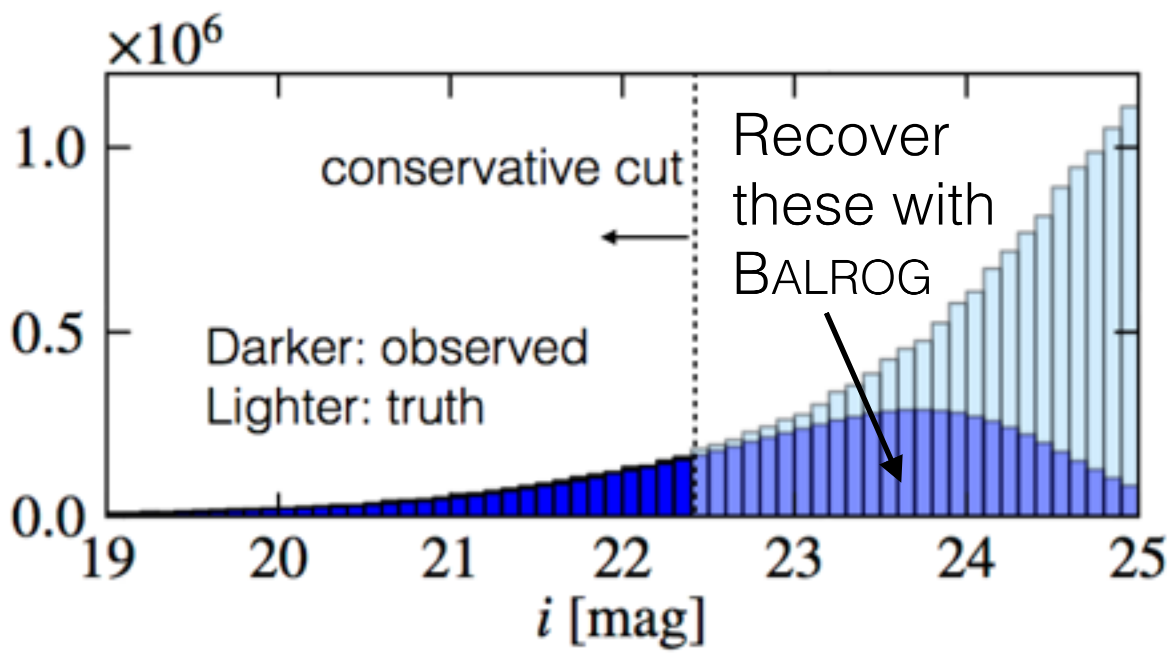
[Crocce et al. 2015]

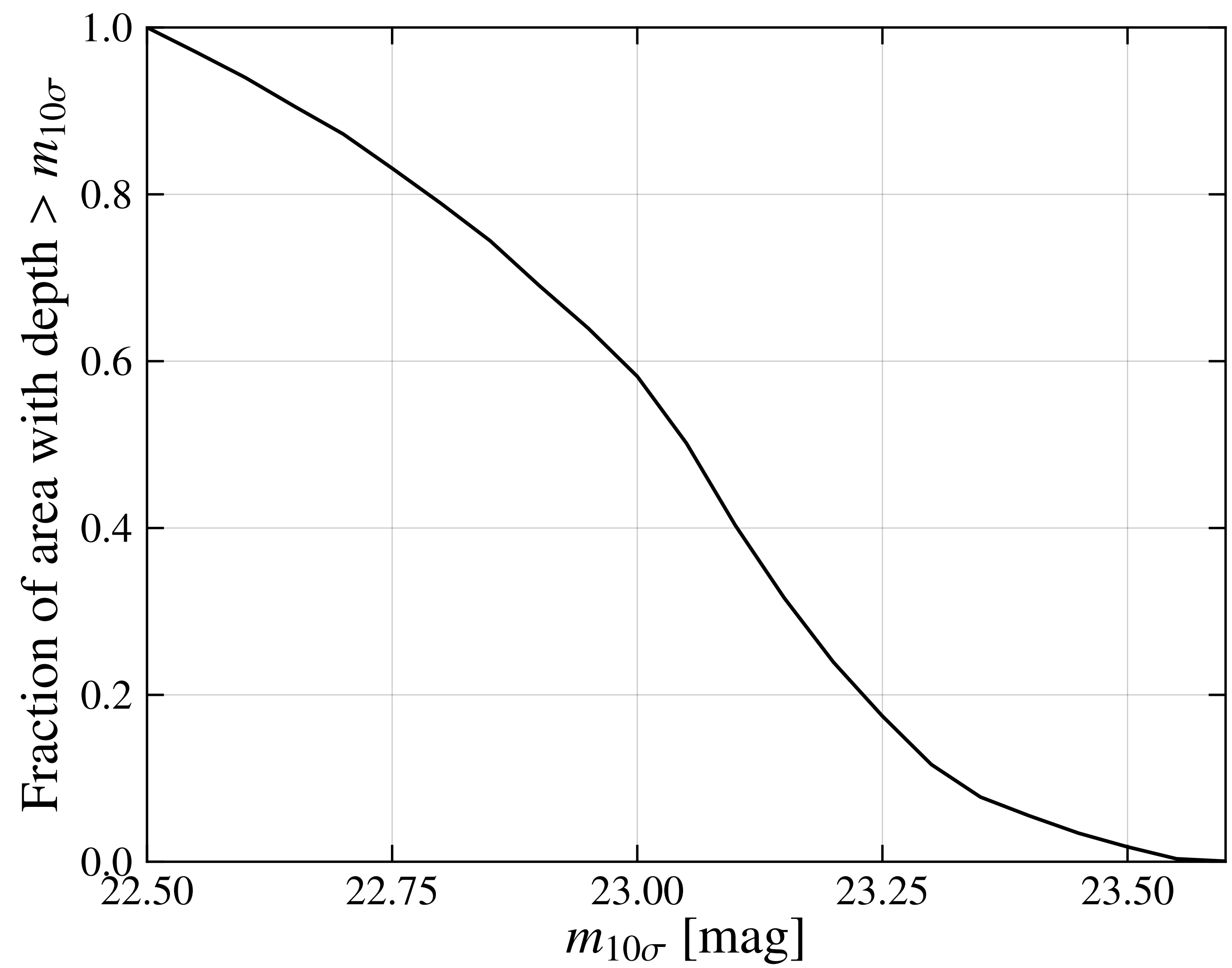


[Suchyta et al. 2015]

The payoff

- Use faint galaxies too — higher statistics, more distant redshifts for cosmology
- Recover area to which no completeness cut would have been possible





Measurement strategy

- Compare 3 types of $w(\theta)$ measurements:
    1. DES data, using usual (uniform, masked) randoms
    2. DES data, using BALROG randoms
    3. COSMOS (Hubble Space Telescope based) data [Capak et al. 2007], using usual randoms

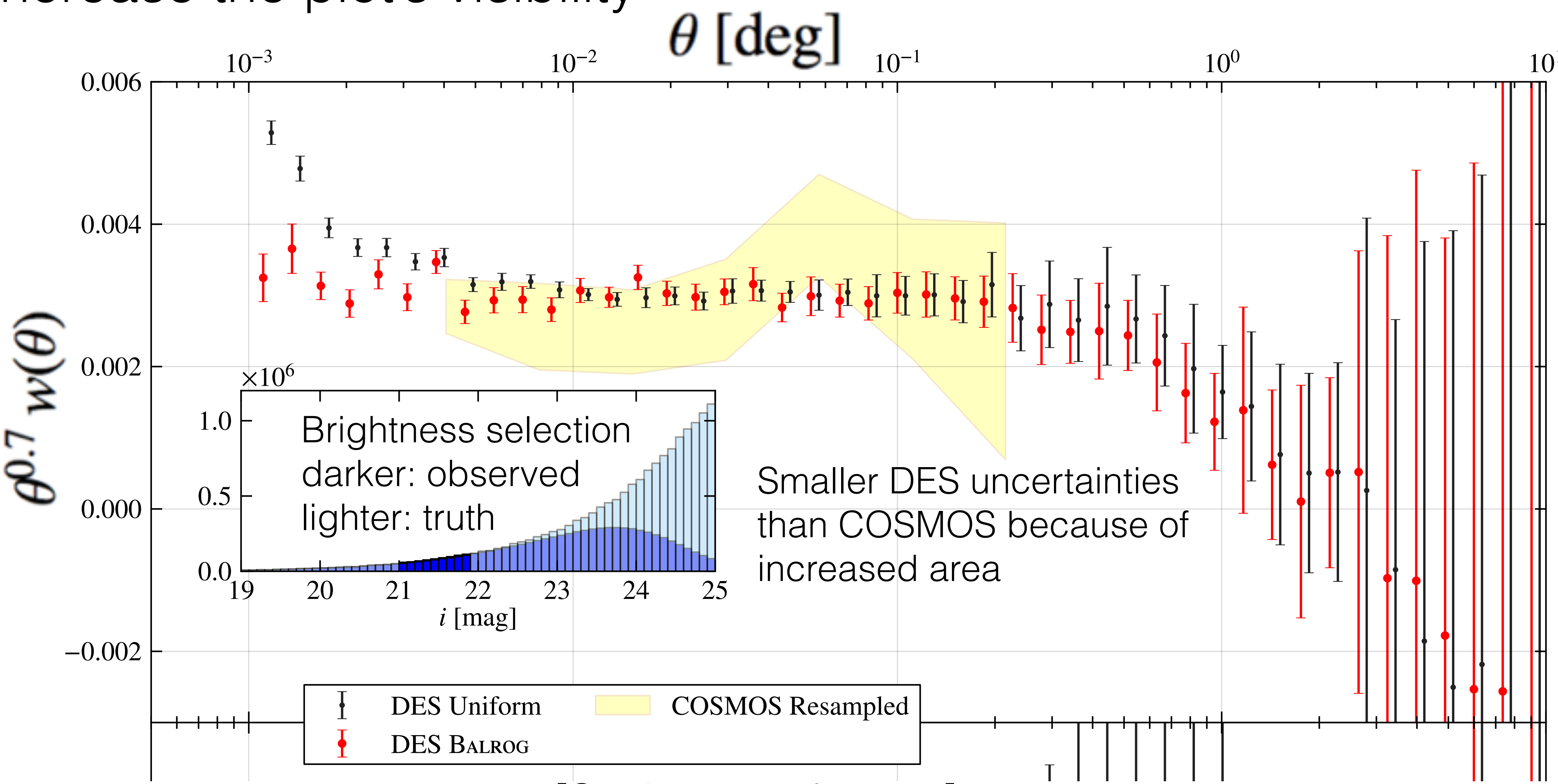
- Select two samples: one complete sample, one incomplete one
    1. Complete sample — sanity check
    2. Faint sample — Does BALROG agree? Compare to complete COSMOS measurement

Results: bright sample

- Signals are ~ power-laws, been multiplied by approx. exponent to increase the plot's visibility

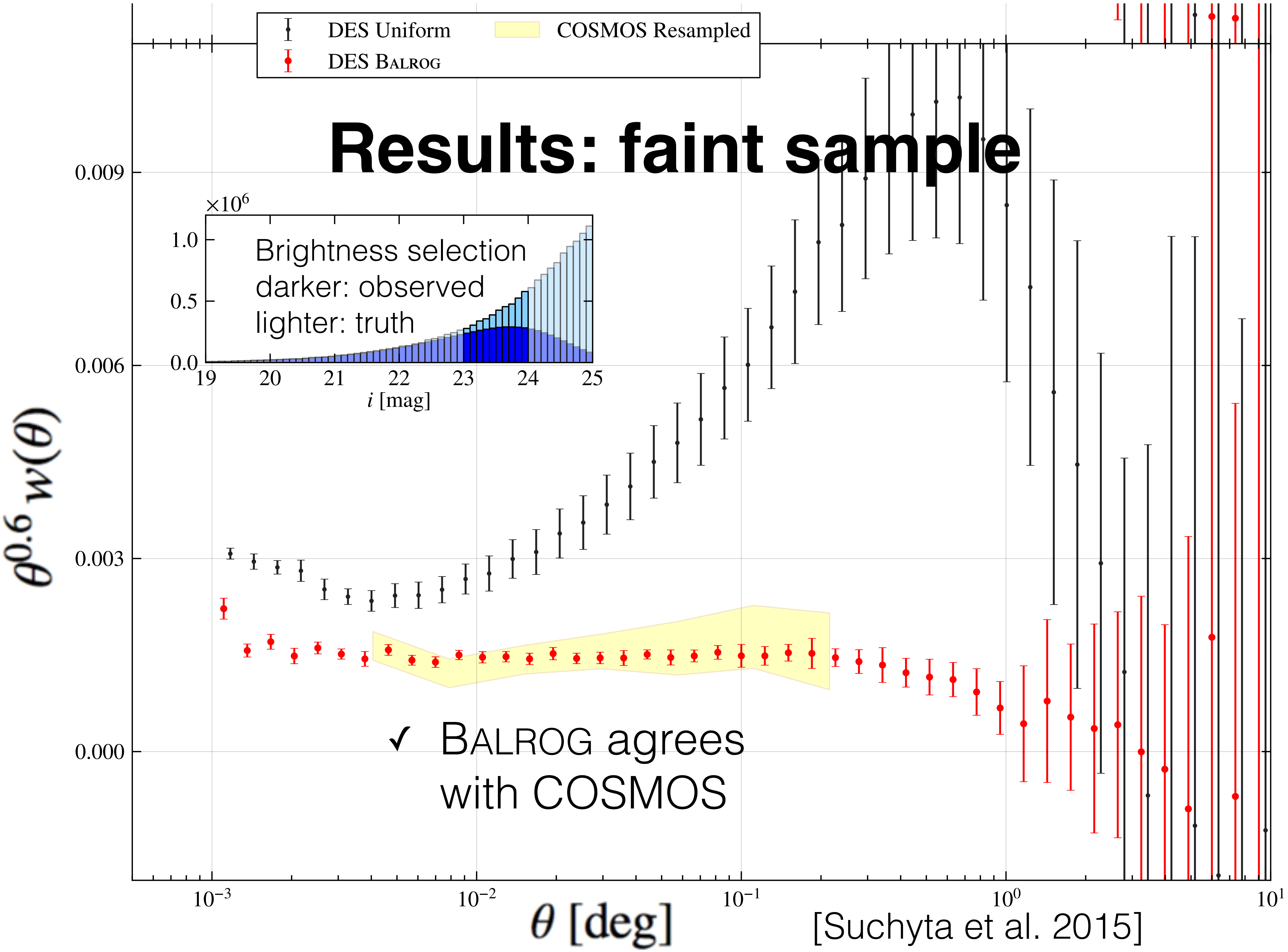
✓ All agree

✓ Sanity check passed



[Suchyta et al. 2015]

Results: faint sample

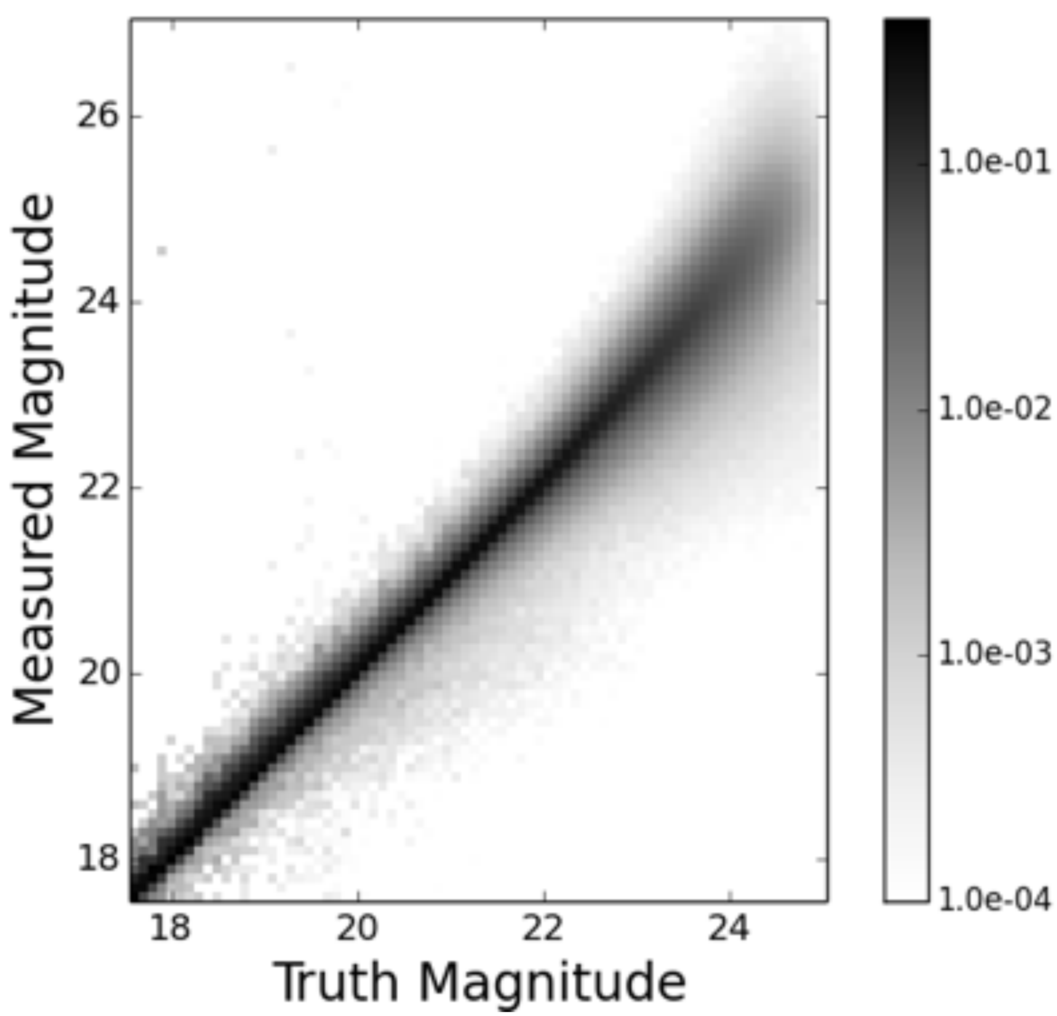


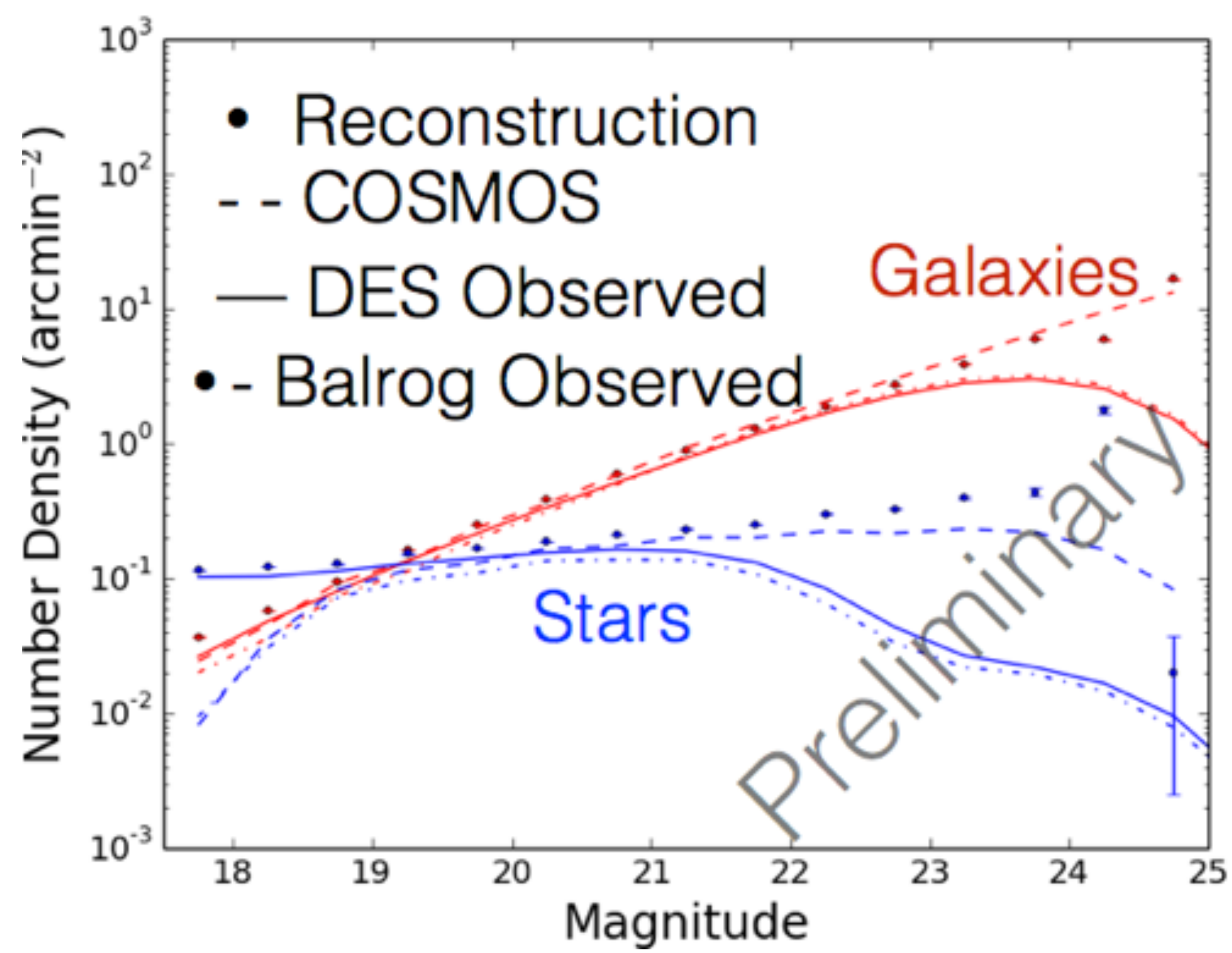
[Suchyta et al. 2015]

- Have removed about 2 orders of magnitude of excess systematic power
- First-ever $w(\theta)$ measurement for highly incomplete sample

BALROG work in progress

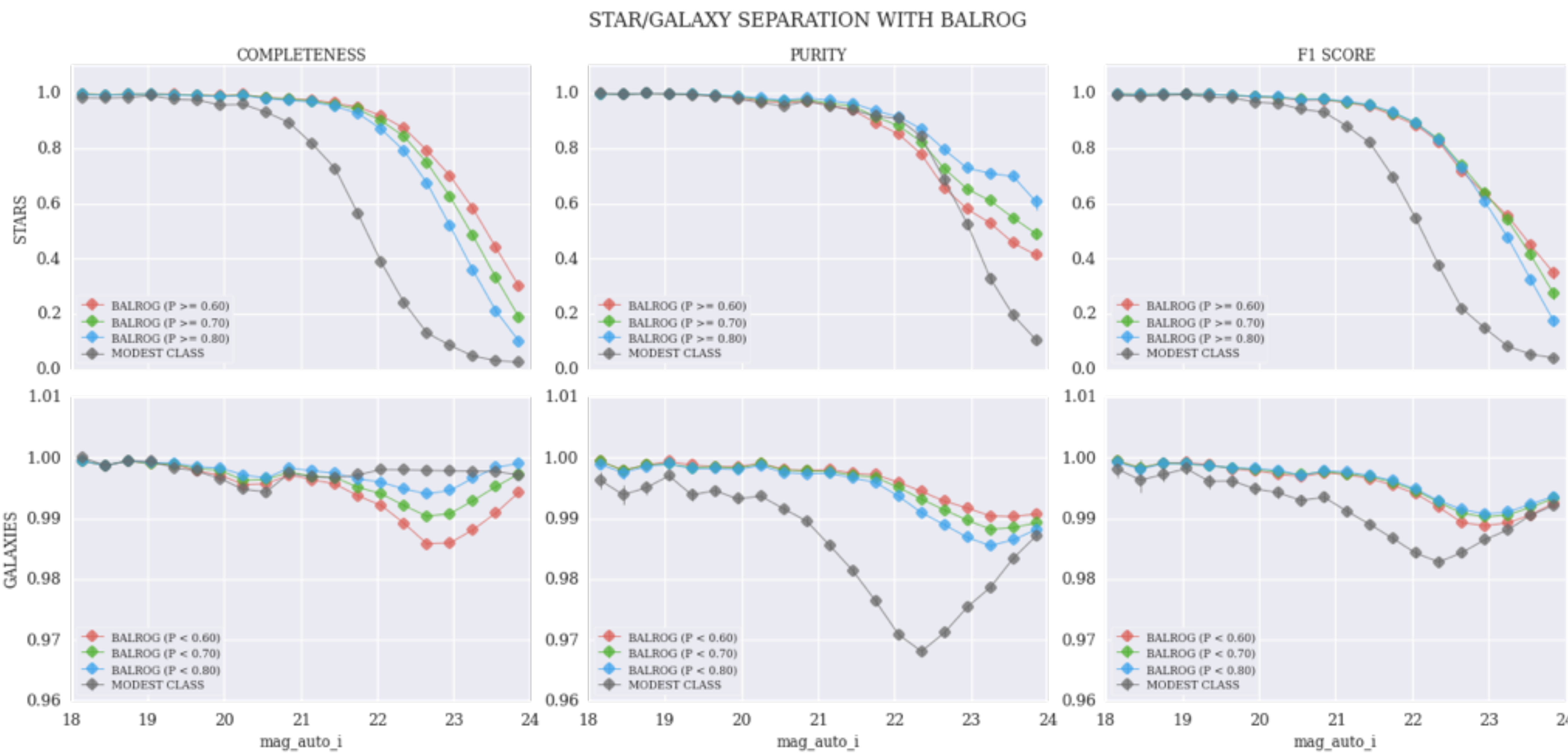
- Inverting survey likelihood function to recover underlying truth properties
- Ultimate goal: magnification





BALROG work in progress

- Machine learning: star/galaxy separation, photometric redshifts



Conclusion

- Developed a new tool called BALROG to carefully identify and then remove systematic effects from difficult science analyses
- Used the machinery to make the first-ever galaxy clustering measurement for a highly incomplete sample
- Expect the methodology to be broadly useful for extending the statistical reach of measurements in a wide variety of coming imaging surveys