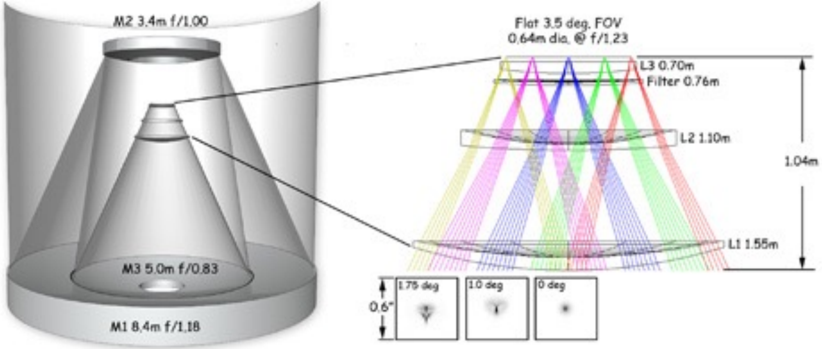
Improving 3x2pt Cosmology Constraints: Training Sample Augmentation, Optimal Binning, and Neural Network Classifiers

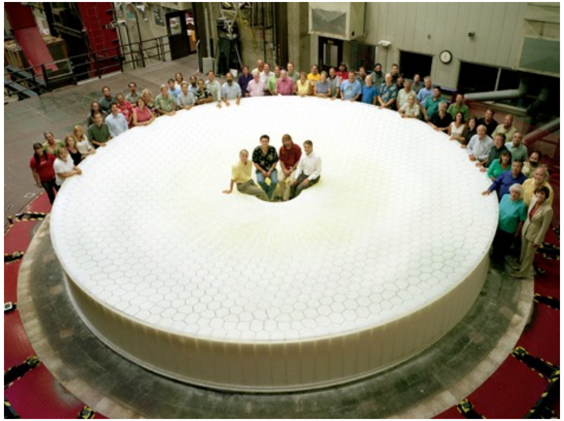
Eric Gawiser
(Rutgers University)

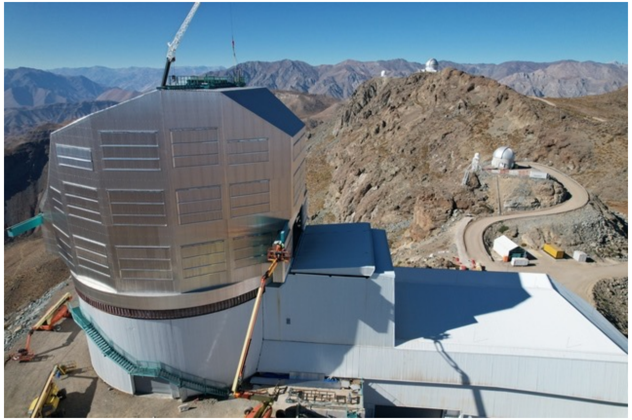
with Irene Moskowitz (Rutgers) and
the LSST Dark Energy Science Collaboration

Broussard & Gawiser 21 ApJ 922, 153
Moskowitz+23 ApJ 950, 49
Moskowitz+24 arXiv:2402.15551

The Vera C. Rubin Observatory Legacy Survey of Space and Time (LSST)







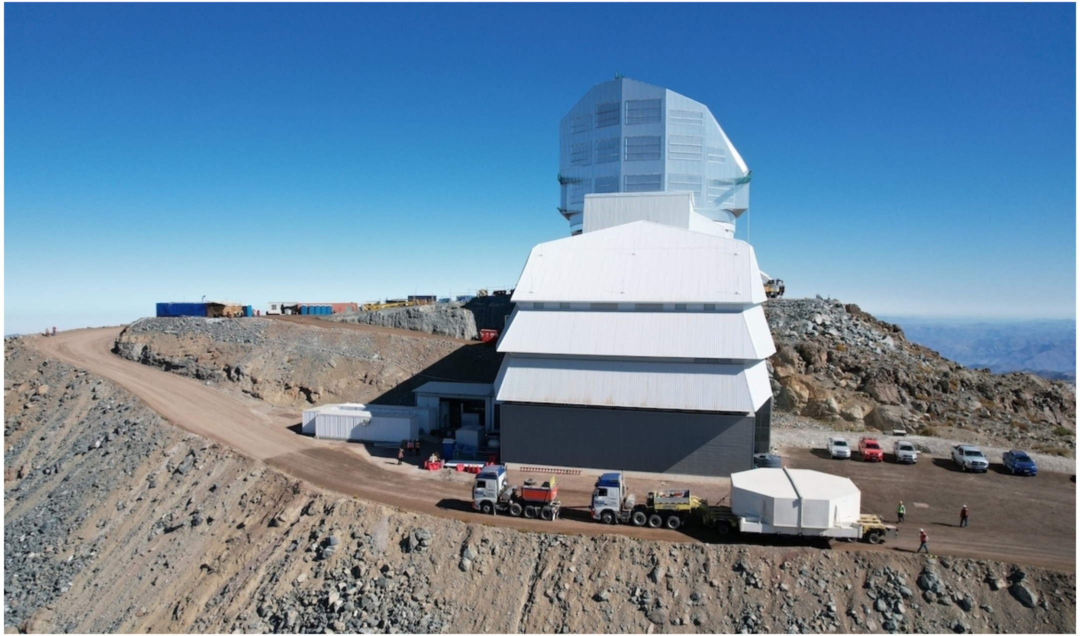
TIMELINE: 2025-2035

Director: Zejlko Ivezic

PRICETAG: $700M

(DOE+NSF+donors)

Vera C. Rubin Observatory



8.4 m primary/tertiary moving into the observatory on March 7, 2024.

Moved onto the mirror cell on March 20.
Will now be prepared for coating.

Rubin Obs/NSF/AURA/O.Rivera.

Slide courtesy of Elisa Chisari (DESC Deputy Analysis Coordinator)

The LSST Camera

will soon be shipped to Chile!

Wednesday's Press Release:

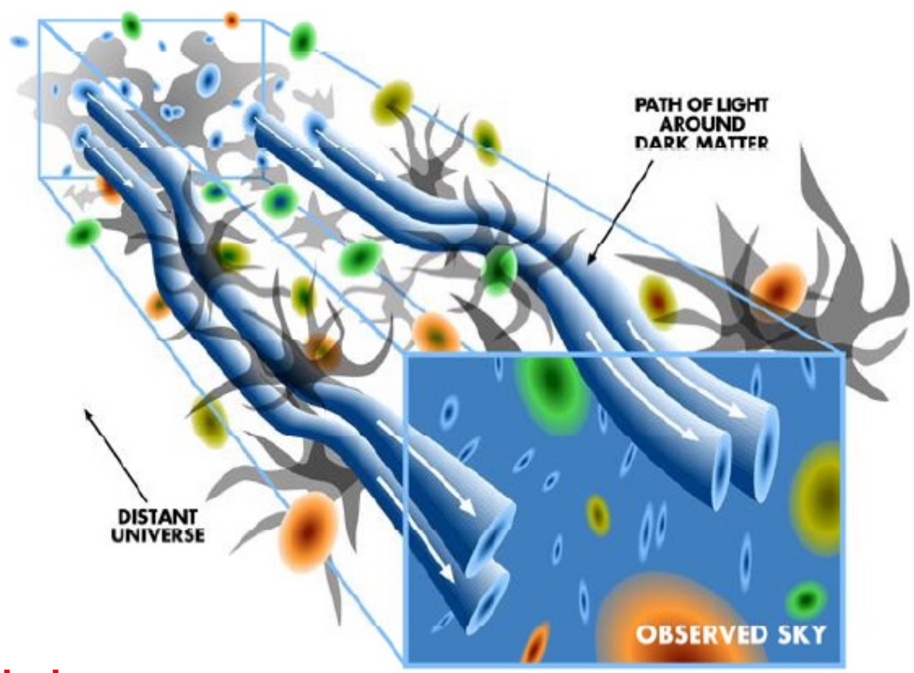
APRIL 3, 2024

SLAC completes construction of the largest digital camera ever built for astronomy

Once set in place atop a telescope in Chile, the 3,200-megapixel LSST Camera will help researchers better understand dark matter, dark energy and other mysteries of our universe.

Motivating LSST

- Time domain science
  - Nova, supernova, GRBs
  - Source characterization
  - Instantaneous discovery
- Census of the Solar System
  - NEOs, MBAs, Comets
  - KBOs, Oort Cloud
- Mapping the Milky Way
  - Tidal streams
  - Galactic structure
- **Dark energy and dark matter**
  - **Strong lensing**
  - **Weak Lensing**
  - **Constraining the nature of dark energy**



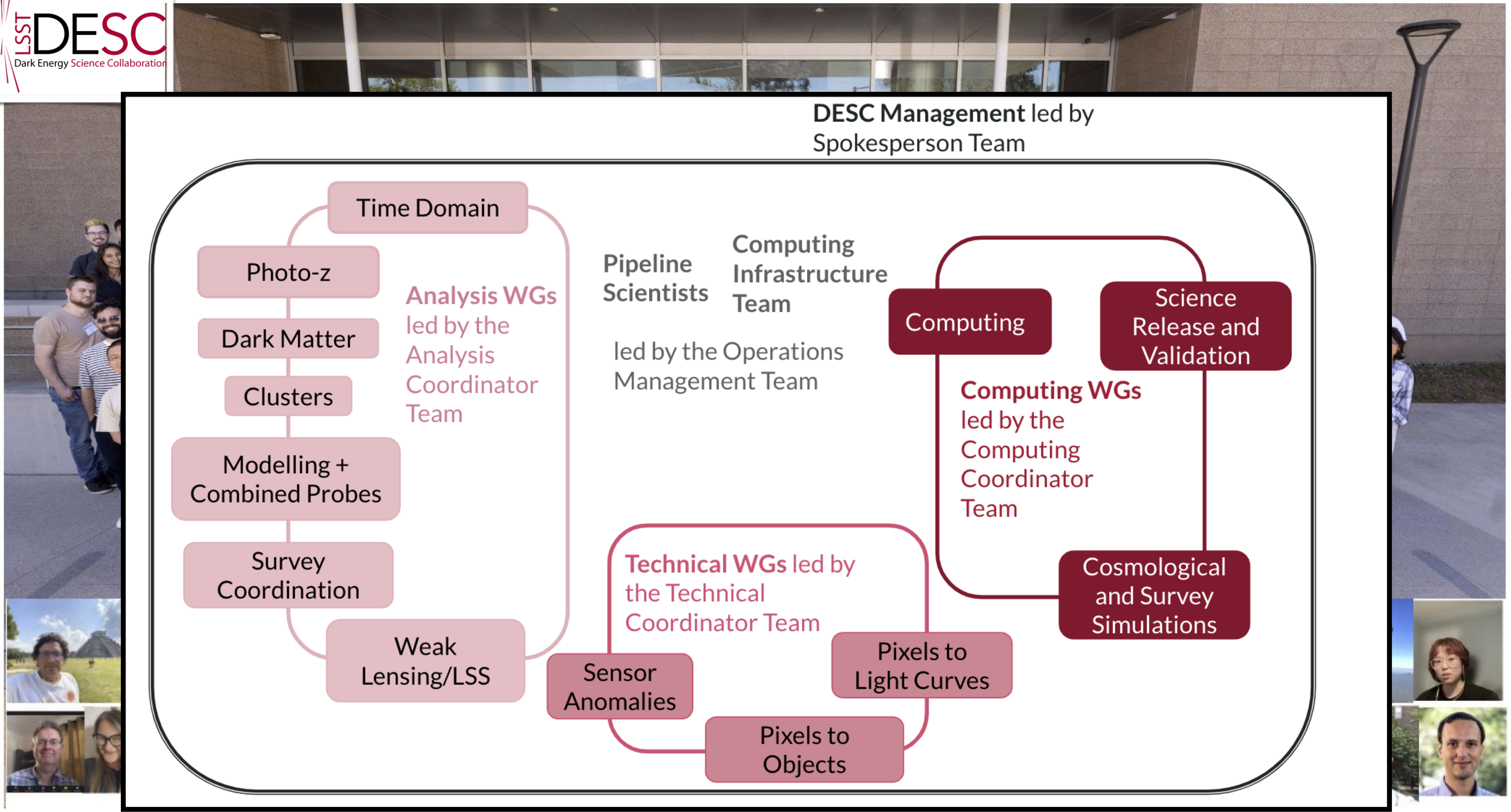
slide from Mario Juric

LSST Dark Energy Science Collaboration (DESC)

LSST DESC
Dark Energy Science Collaboration

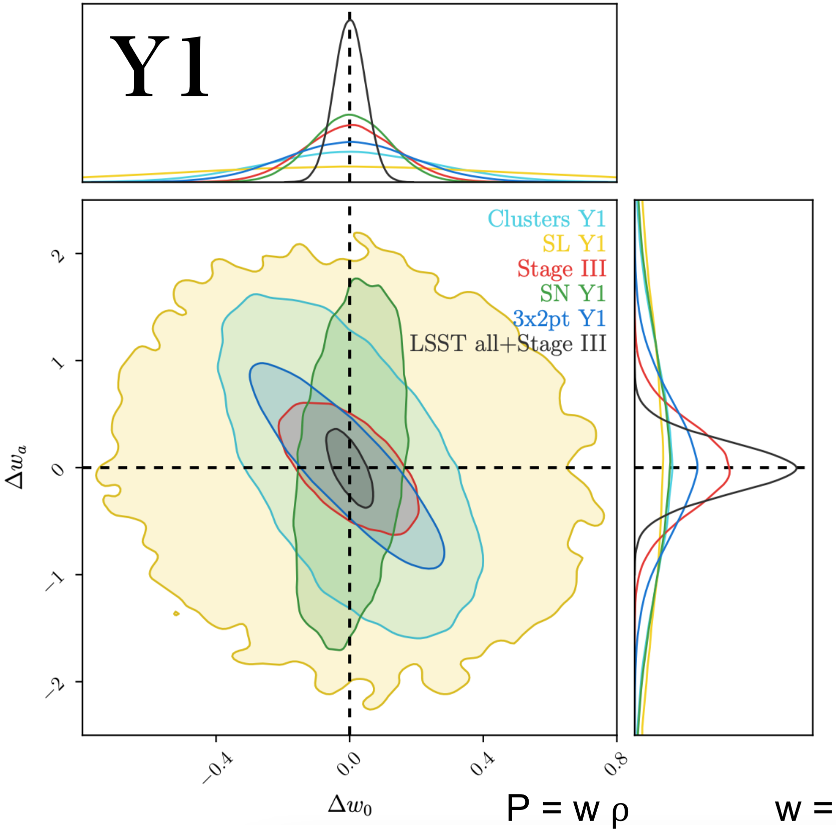
- 1200 MEMBERS
- 250 FULL (VOTING) MEMBERS
- 273 INSTITUTIONS (102 IN US)
- 31 COUNTRIES

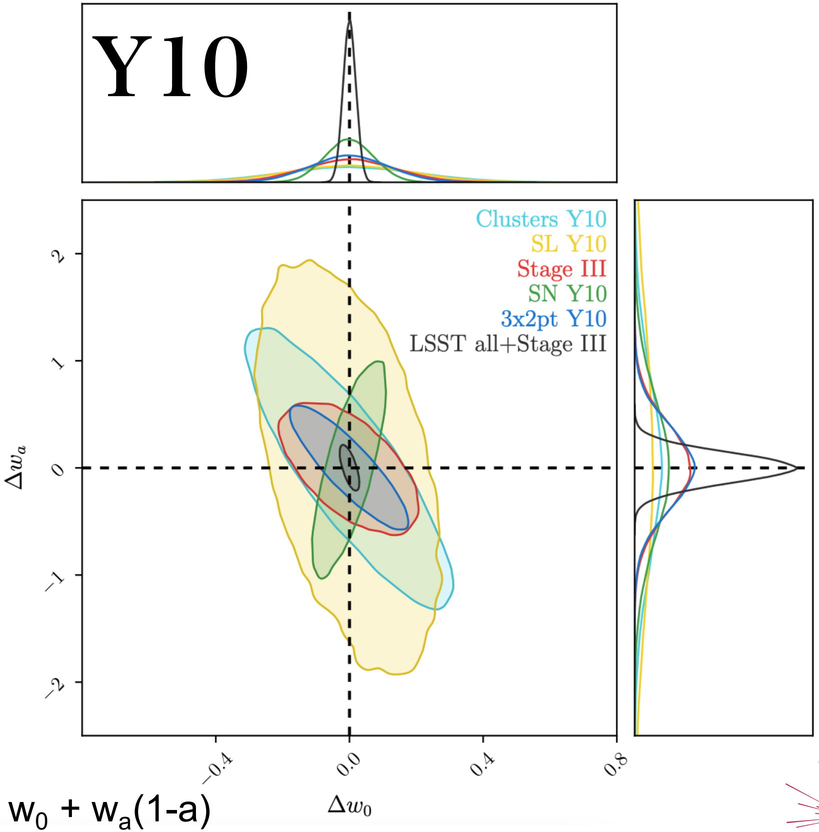
LSST Dark Energy Science Collaboration (DESC)



The Goal

Forecasted 68% credible regions on (w_0, w_a) for individual probes and their combination after all LSST data is analyzed.





$P = w\,\rho$ $\quad$ $w = w_0 + w_a(1-a)$

Mandelbaum+, incl. EG, LSST DESC Science Requirements Document (2018), 1809.01669

Introduction

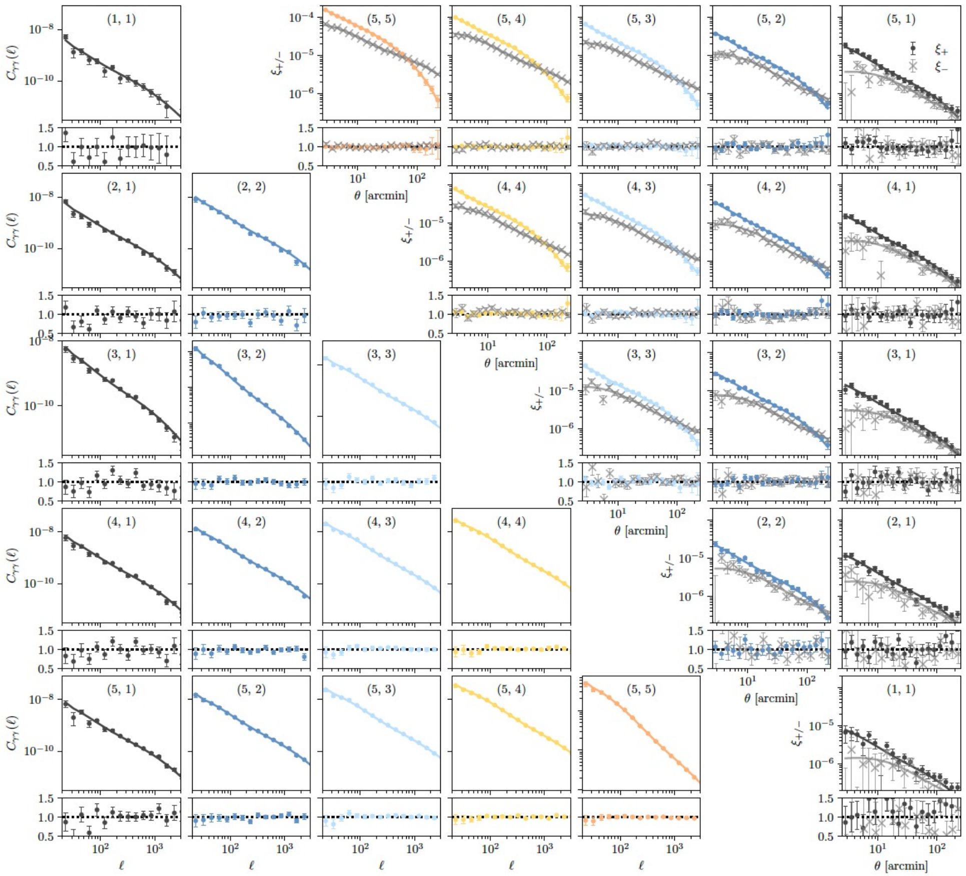
- The 3x2pt method consists of 3 two-point correlation functions combining weak lensing and large scale structure
  - Shear-shear: auto-correlation between shapes of galaxies
  - Galaxy clustering: auto-correlation between galaxy locations
  - Galaxy-galaxy lensing: cross-correlation between shapes and locations
- LSST will observe too many galaxies to get spectroscopic redshifts and will instead rely on less precise photometric redshifts
  - Galaxies will be sorted into tomographic redshift bins
  - 2D correlation functions are computed within and between bins

Towards a 3x2pt analysis

Measurements: TXPipe

PRELIMINARY

Prat+, LSST DESC (2022), 2212.09345

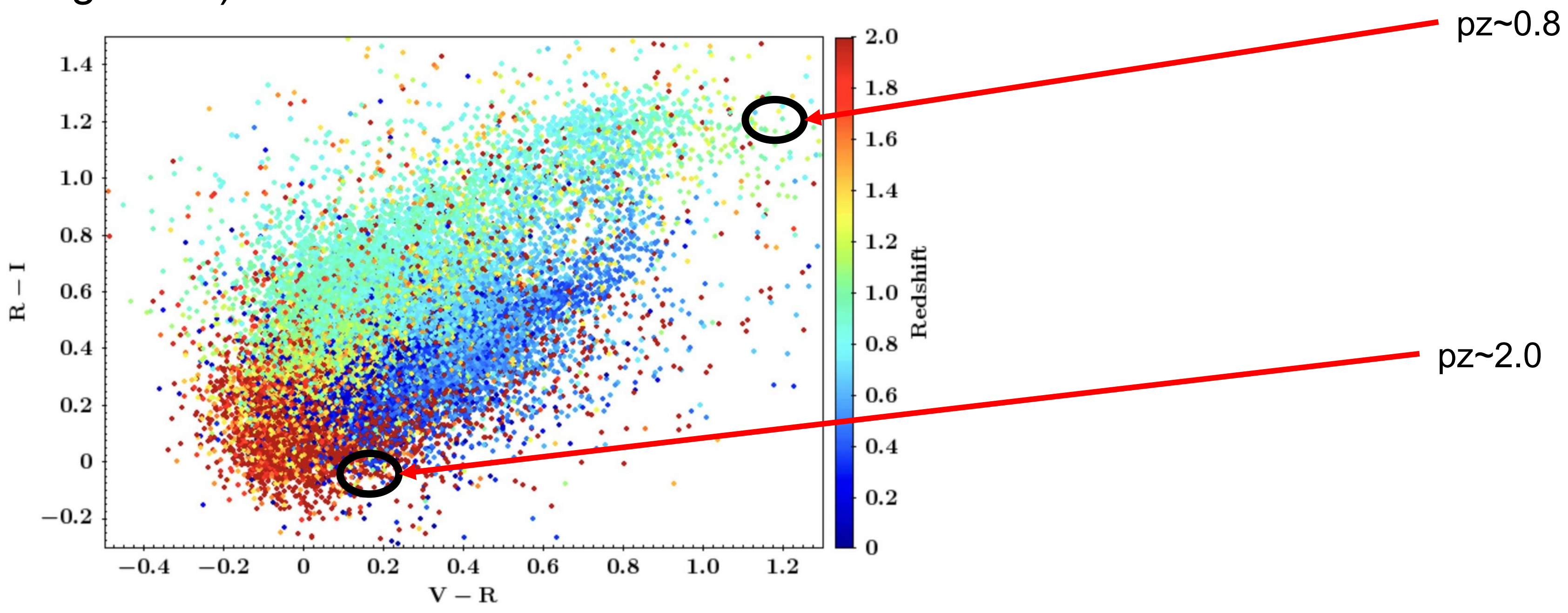


LSST DESC Dark Energy Science Collaboration

Slide courtesy of Elisa Chisari (DESC Deputy Analysis Coordinator)

Photo-z as a mapping of color space to redshift

Rubin ugrizy will give 6 dimensional space (6 magnitudes or 5 colors + 1 magnitude)



Slide stolen from Sam Schmidt’s excellent DE School lesson

Machine Learning for Photo-z's

A variety of machine learning methods can be used to estimate photometric redshifts (photo-z's) including:

- Self-organized maps
- Neural networks (e.g ANNz, NetZ)
- Random forests (e.g TPZ)
- Conditional density estimators (e.g FlexZBoost)

And many others!

These photo-z machine learning methods require two things:

- A training sample of galaxies with accurate (spectroscopic) redshifts and photometry
- An application sample of galaxies with only photometry

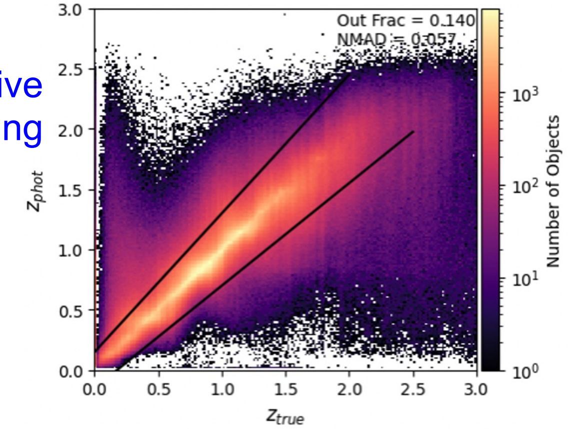
Motivation

- The DESC Tomographic Challenge (https://arxiv.org/abs/2108.13418) investigated methods of optimizing the source sample selection for cosmic shear
- We investigate optimizing the lens sample for galaxy clustering
  - Get to use the full range of photometric filters
  - We also add a realistic level of non-representativeness to the training sample
- We present a method for optimizing the tomographic binning strategy for the lens sample of galaxies used for galaxy clustering using a realistically non-representative training sample for estimating photo-z's
  - Pipeline is tested on 2 mock galaxy catalogs: Buzzard (DeRose et al. 2019) and CosmoDC2 (Korytov et al. 2019)
  - Training sample is redder and brighter than application sample, consistent with current spectroscopic samples
- Optimize choice of bin edges and selection of galaxies for binning

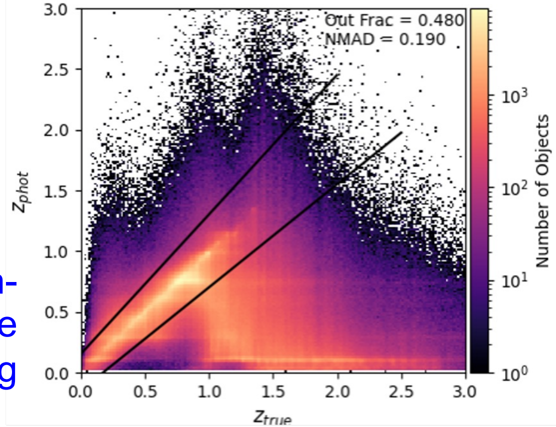
Challenge: Six-band photometric redshifts are difficult!

- LSST will rely on photometric redshifts
  - Current spectroscopic samples are brighter and redder (and lower redshift) than expected LSST data
  - DESI spectroscopy doesn't go deep enough to solve the problem
- Non-representative training samples lead to poor photo-z estimates for galaxies with features outside the training sample range
- Need to figure out a way to improve photo-z estimates without relying on new spectroscopic samples

Representative Training



Realistically Non-representative Training



Figures from Moskowitz+2024 [arXiv:2402.15551]

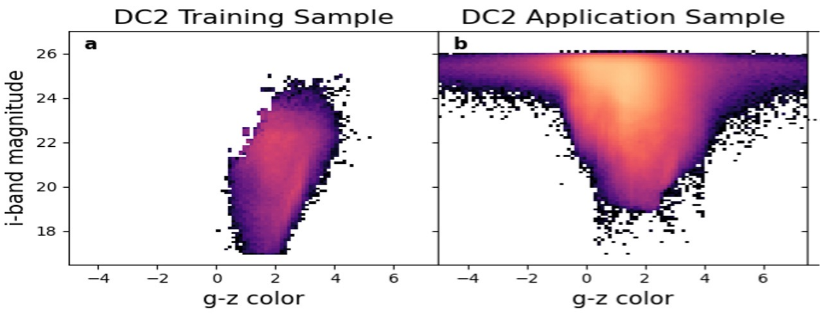
Training Sample Augmentation

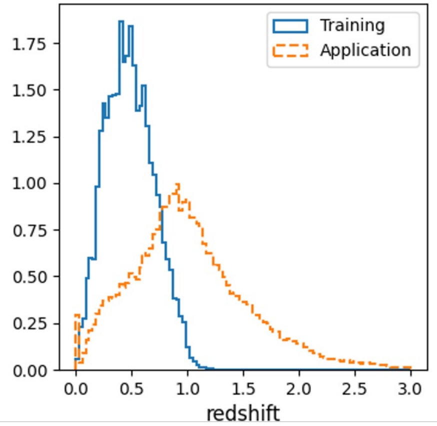
- What if we just add simulated galaxies to our training sample with otherwise unrepresented photometry/redshifts?
- No simulation is perfect: need two simulations that are different enough from each other to simulate the difference between real data and a simulation
  - Use LSST Dark Energy Science Collaboration Data Challenge 2 (DC2) as “real” data (galaxy SEDs constructed from FSPS)
  - Use Buzzard as simulated data (galaxy SEDs assigned to match the SED-luminosity-density relation measured in SDSS)
- In this work:
  - Application sample (i.e., “validation set”, simulating expected LSST data): all DC2 galaxies
  - Base training set (simulating spectroscopic samples): bright/red/low-redshift DC2 galaxies
  - Used to augment training set: dim/blue/high-redshift Buzzard galaxies

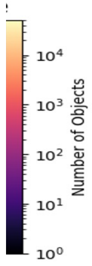
Methodology: non-representative Sample

Split DC2 catalog into a realistically non-representative training sample and application sample using the GridSelection degrader in [RAIL]

- Mimics HSC data with spectroscopic redshifts
- Training sample is redder and brighter than application sample
  Also tends towards lower redshift



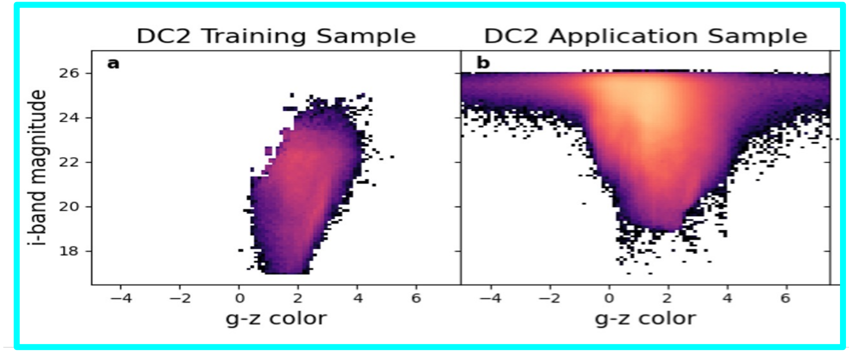


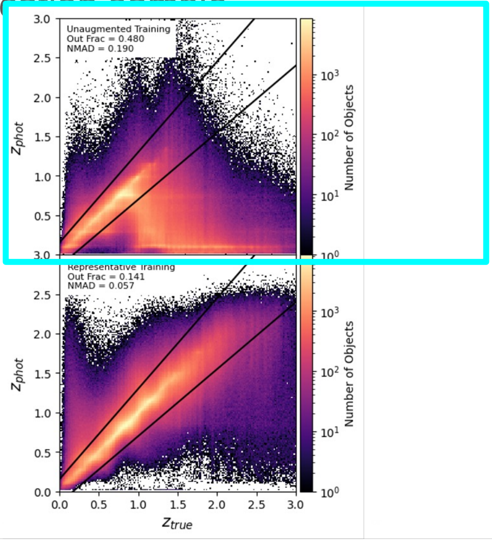


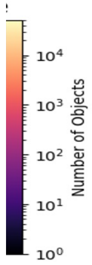
Methodology: non-representative Sample

Split DC2 catalog into a realistically non-representative training sample and application sample using the GridSelection degrader in [RAIL]

- Mimics HSC data with spectroscopic redshifts
- Training sample is redder and brighter than application sample
  Also tends towards lower redshift





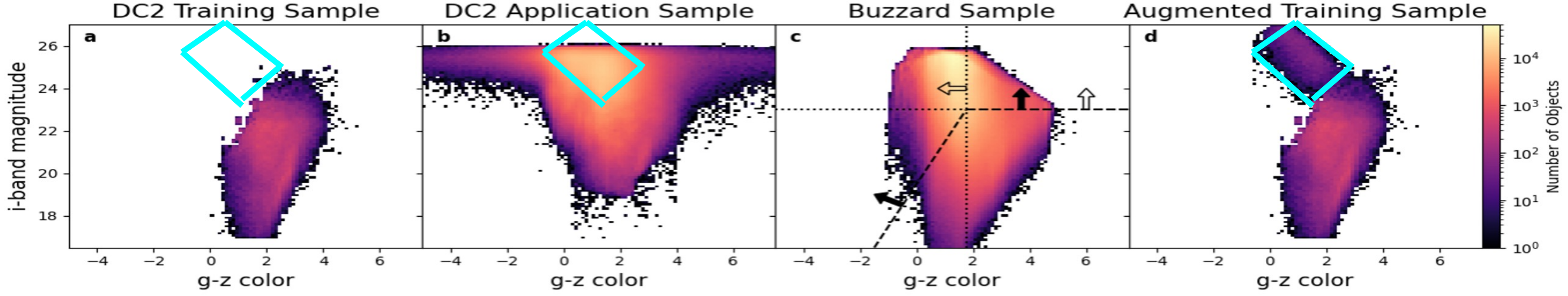


Methodology: Augmentation

- Select 10,000 Buzzard galaxies with features that are unrepresented in DC2 training sample
  - *i*-mag > 23
  - (*g*-*z*) color < 1.75
  - Z_{true} > 1.0
  - Combination of 2 or 3 features
    - Usually the intersection of both features, but matched to training sample shape for color+magnitude

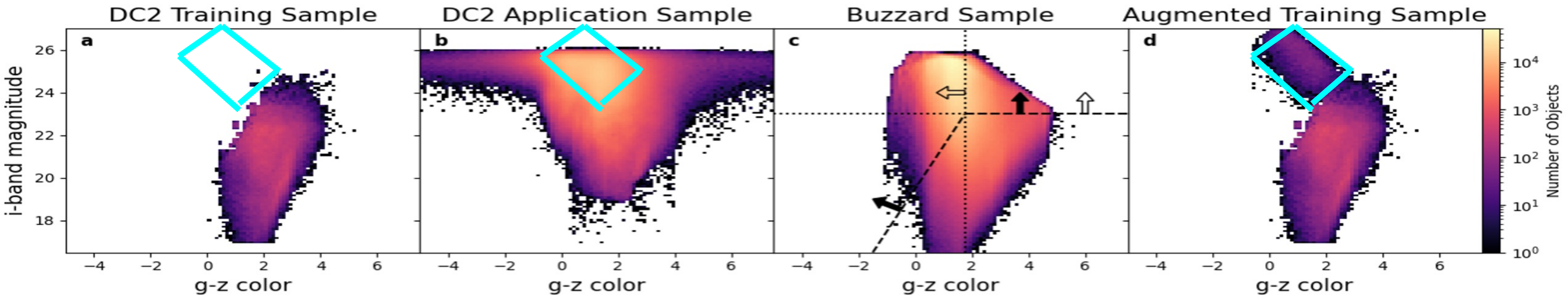
Matched to boundaries of DC2 training sample

Best case augmentation selection shown in panel d: using color+magnitude+redshift and the shifted magnitudes



Methodology: Photometry Shifting

- Buzzard has a different color-redshift relationship than DC2: can we shift the Buzzard photometry in some way so it looks more like the DC2 color-redshift relationship?
  - Magnitude shifting: shift magnitudes in all bands so the median magnitudes match. Generally works best
  - Normalizing flow: generate photometry that matches DC2, and use a conditional flow to generate Buzzard-like redshifts
- Best case augmentation selection shown in panel d: using color+magnitude+redshift and the shifted magnitudes

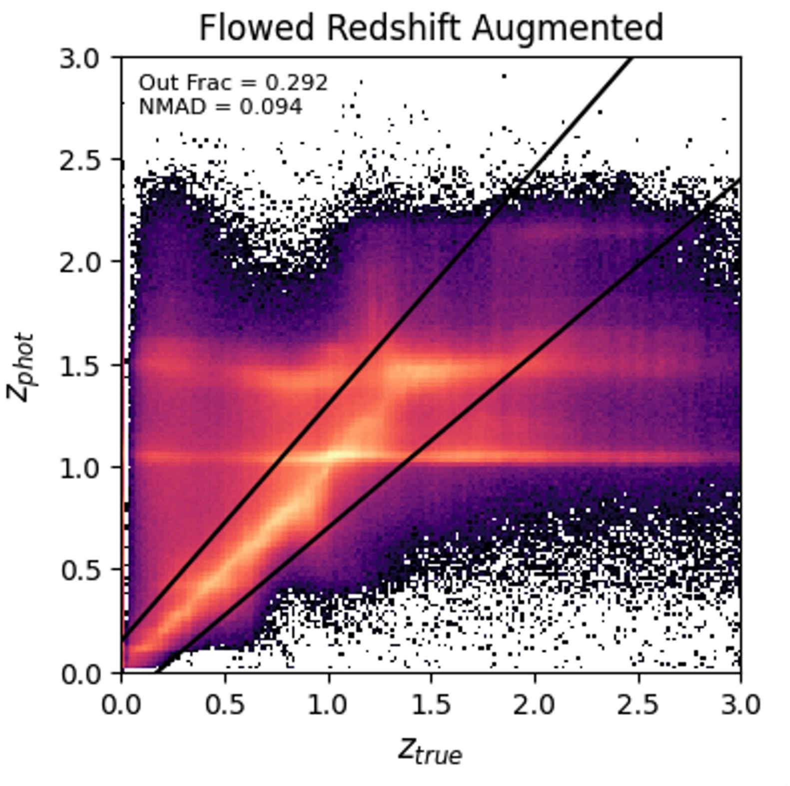


Results

Reported statistics:

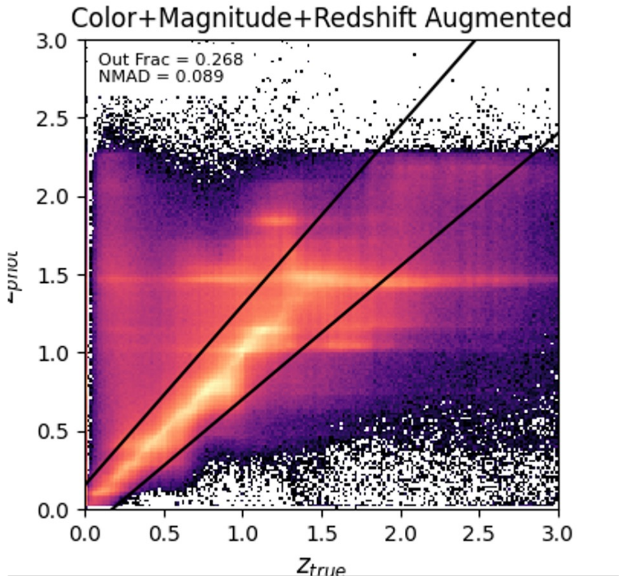
- Outlier fraction: define outliers as $|\Delta z|/(1+z) > 0.15$
- NMAD: normalized median absolute deviation, $1.4286 \times \mathrm{med}(|\Delta z|/(1+z))$
- Percent improvement: $(X_{unaug} - X_{aug})/X_{unaug}$, where X is outlier fraction or NMAD
- Percent recovery towards representative case: $(X_{unaug}-X_{aug})/(X_{unaug}-X_{rep})$
  - Even with a fully representative training sample, outlier fraction and NMAD are not 0, want to capture the recovery towards this best case scenario

Flowed Redshift Augmentation

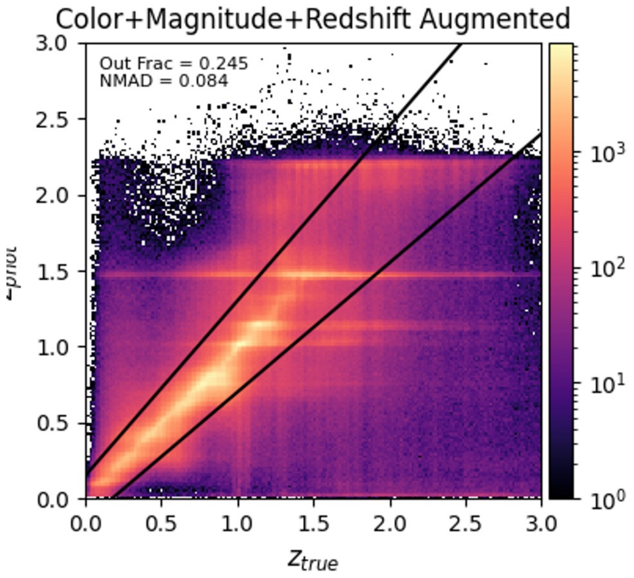


55% recovery outliers
70% recovery NMAD

Unshifted vs Shifted Magnitudes Augmentation: color+magnitude+redshift



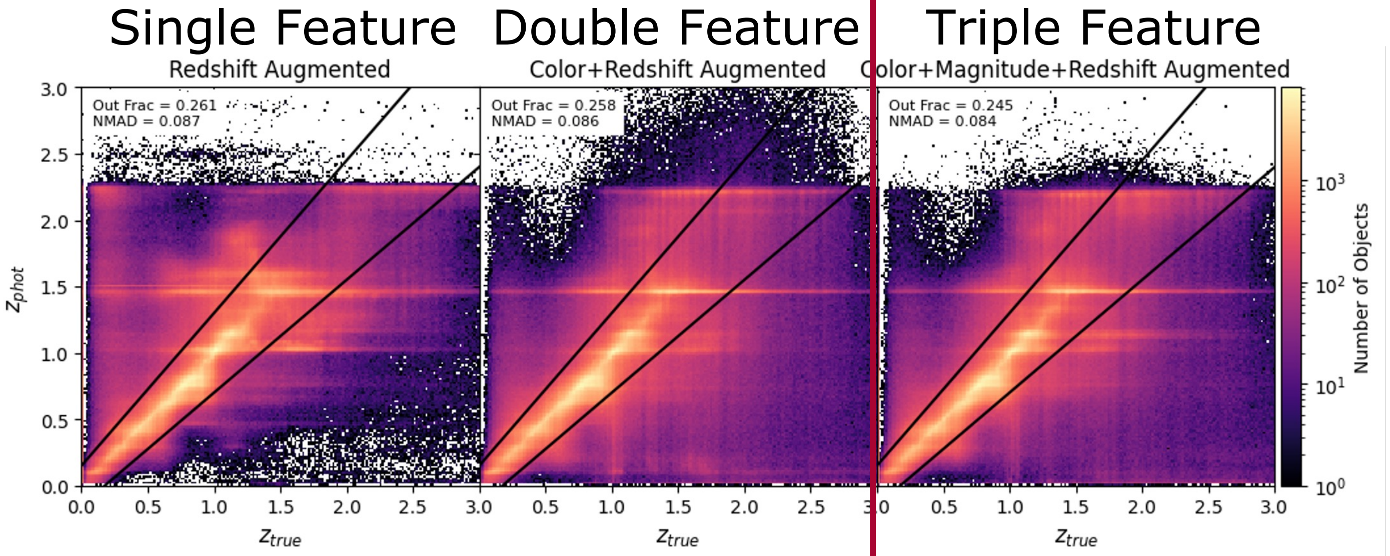
62% recovery outliers
76% recovery NMAD



69% recovery outliers
80% recovery NMAD

Best Results

Figures from Moskowitz+2024 [arXiv:2402.15551]



% improvement	46% (outlier frac) 54% (NMAD)	46% (outlier frac) 56% (NMAD)	49% (outlier frac) 56% (NMAD)
% recovery	65% (outlier frac) 77% (NMAD)	65% (outlier frac) 78% (NMAD)	69% (outlier frac) 80% (NMAD)

Summary of Training Sample Augmentation

- Can get over ⅔ of the way back to the photo-z quality achieved by a fully representative training sample with some simple augmentation
- Only added 10,000 Buzzard galaxies to our DC2 training sample (originally 180,000 objects)
- Simple selection criteria (redshift, one magnitude and one color)
- Expect even better results when using an updated, more realistic simulation on real data
- Full quantification of improvements will require an end-to-end cosmological parameter estimate to confirm that augmentation reduces parameter biases

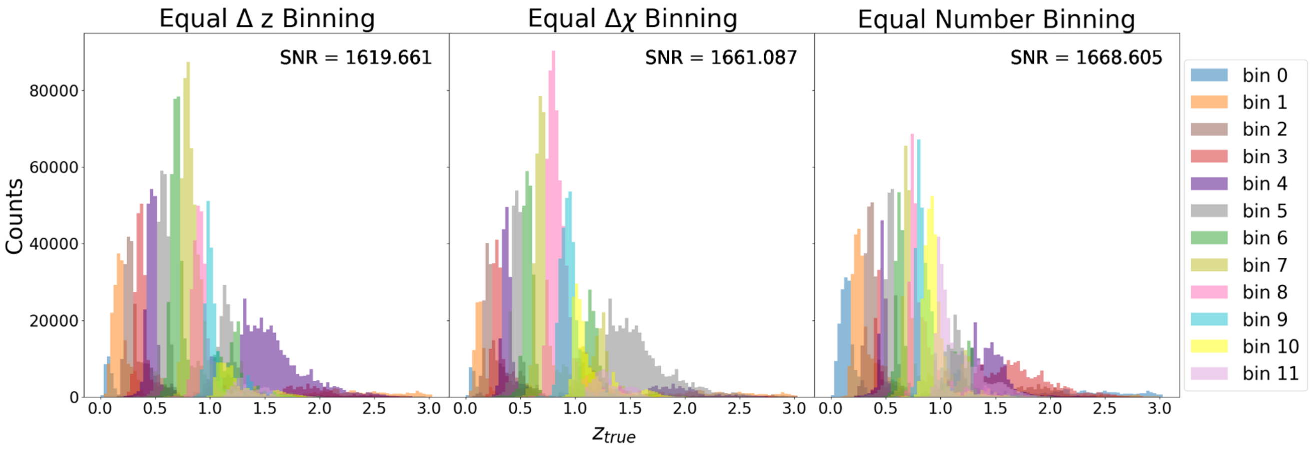
Optimizing Bin Edges

Figures from Moskowitz+2023 [ApJ 950, 49]

3 common, physically motivated choices:

- Equally spaced in redshift (equal Δz)
- Equally spaced in comoving distance (equal Δχ)
- Equal numbers in each bin

Equal number binning performs ~best out of these three choices



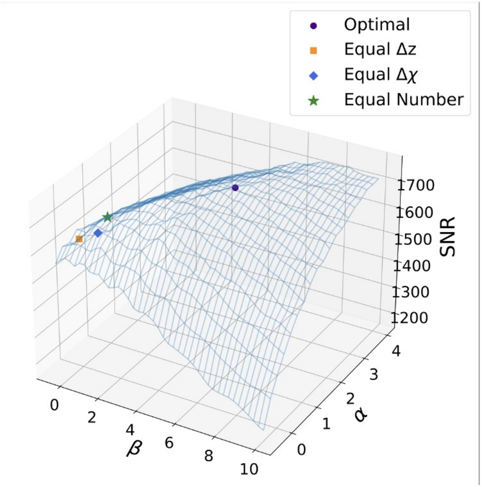
Optimizing Bin Edges

- Start with three base methods: equal Δz, equal $\Delta\chi$, and equal number binning
- Generalize by introducing the binning equation $\mathcal{M} = \int_0^{z_{max}} \left(\frac{dN}{dz}\right)^{\alpha} \left(\frac{d\chi}{dz}\right)^{\beta} dz$
- Divide $\mathcal{M}$ into 12 equal bins, interpolate back to redshift values
- Three special cases:
    - $\alpha=\beta=0$: $\mathcal{M}$ is z_{max}, recover equal Δz bins
    - $\alpha=0$, $\beta=1$: $\mathcal{M}$ is χ_{max}, recover equal $\Delta\chi$ bins
    - $\alpha=1$, $\beta=0$: $\mathcal{M}$ is the total number of galaxies, recover equal number bins

Optimizing Bin Edges

$$\mathcal{M} = \int_0^{z_{max}} \left(\frac{dN}{dz}\right)^{\alpha} \left(\frac{d\chi}{dz}\right)^{\beta} dz$$

- Compute bin edges for many values of α and β
- Sort galaxies into bins using their photo-z estimates
- Calculate SNR of the angular power spectra derived from the bins
- For CosmoDC2, we find the highest SNR at α=2.0 and β=5.25 - a 5% improvement versus equal-number binning

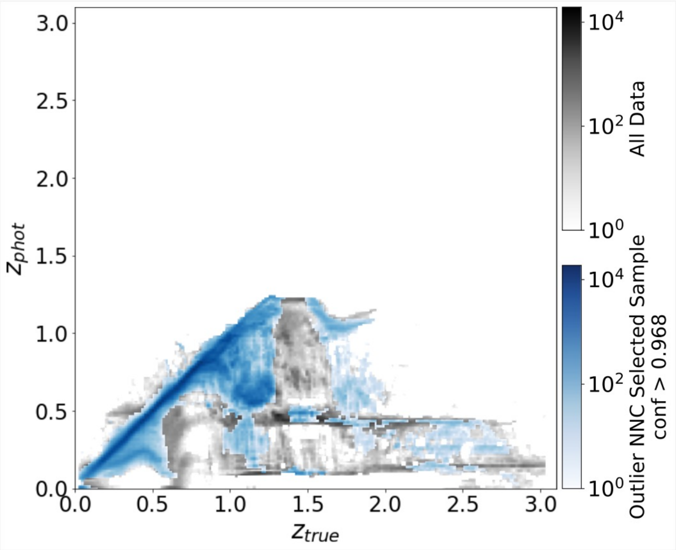


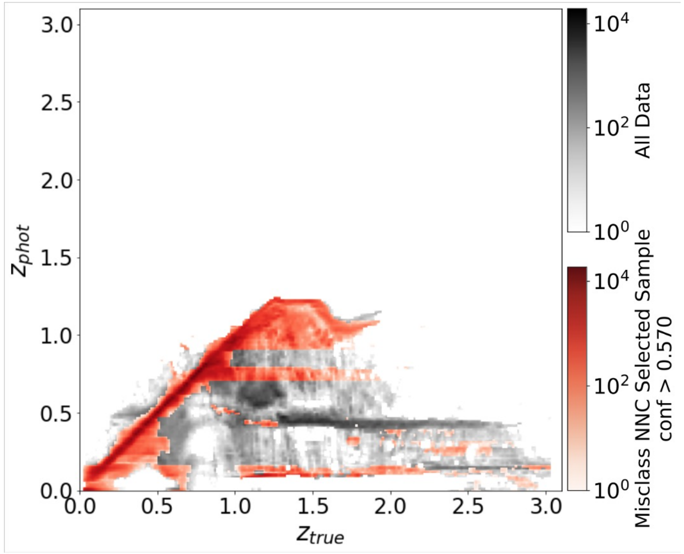
Optimizing Sample Selection

- LSST will not be shot-noise limited, so we we can improve the binning by removing galaxies with poor photo-z estimates
- Only care if the photo-z is good enough to place it in the correct bin
- Train two Neural Network Classifiers to make the sample selection
- Outlier NNC (Broussard & Gawiser 2021):
  - Estimates confidence that a given photo-z estimate is accurate
  - By excluding galaxies with low Outlier NNC confidence, we remove galaxies with high probability of being outliers
- Misclassification NNC:
  - Estimates confidence that a given photo-z estimate will result in the galaxy being sorted into the correct redshift bin
  - By excluding galaxies with low Misclassification NNC confidence, we remove galaxies with high probability of being misclassified
  - Must be retrained for each choice of bin edges

Optimizing Sample Selection

Figures from Moskowitz+2023
[ApJ 950, 49]

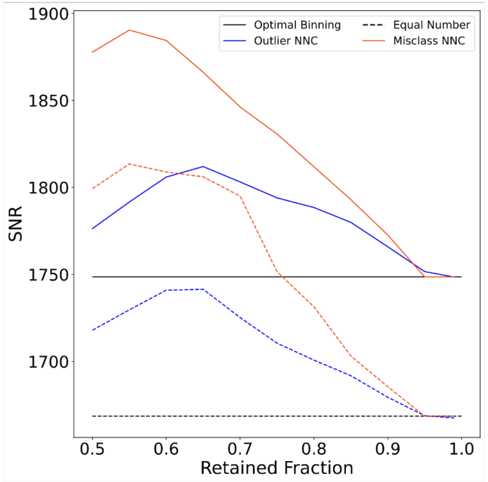




Outlier NNC (left) and Misclassification NNC (right) pick out different samples of galaxies when selecting the same overall number

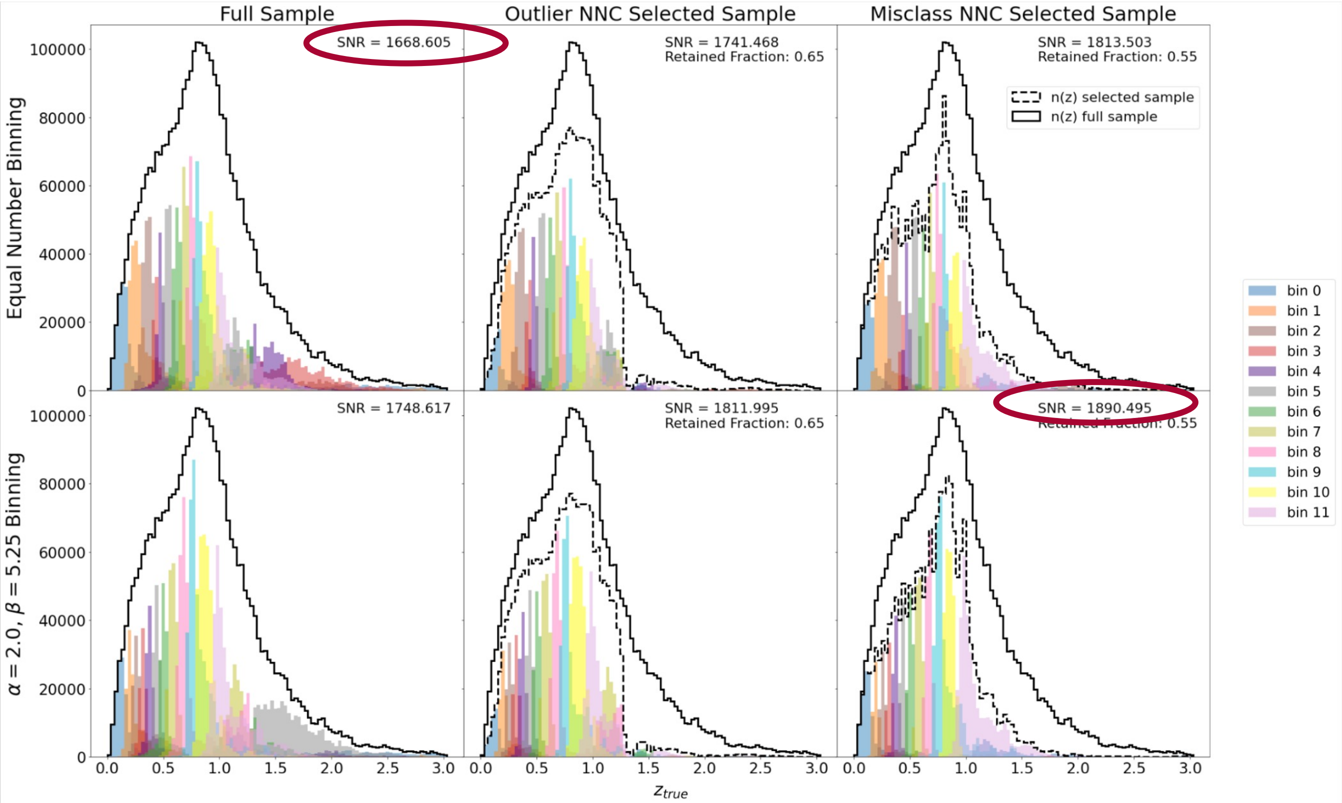
Optimizing Sample Selection

- Can also ask how much of the sample we should remove
- We compare the results for the optimized binning choice (α=2.0 and β=5.25) to the best of the three base choices (equal number binning)
- Misclassification NNC outperforms Outlier NNC in both cases
- Misclassification NNC prefers to remove a slightly larger fraction of the sample than Outlier NNC



Final Results

Figures from Moskowitz+2023 [ApJ 950, 49]



The optimized bin edge choice and the NNC sample selection improve SNR by ~13% over the base equal number binning choice

Other Research Projects You Can Ask Me About

- ODIN: Large Survey for Lyman-Alpha Emitting Galaxies on the Dark Energy Camera to study how galaxies like the Milky Way were formed
- JWST-CEERS: Reconstructing star formation histories of galaxies at $z>5$ – and maybe $z>10$
- Simons Observatory: Cosmic Microwave Background searching for primordial gravitational waves; Engagement, Mentoring & Climate Committee = EMC^2

Conclusions

- We built a realistically non-representative training sample from the LSST-DESC DC2 catalog and augmented it with simulated Buzzard galaxies
- Can get over ⅔ of the way back to the photo-z quality achieved by a fully representative training sample with some simple augmentation **Moskowitz+2024**
- We proposed a method for optimizing the tomographic binning strategy from two directions
  - Optimizing the choice of bin edges
  - Optimizing the selection of galaxies for binning

Moskowitz+2023

- We found that the optimized choice of bin edges is not one of 3 common choices
- Removing ~50% of the galaxies with our neural network classifier further improves 3x2pt SNR for a total improvement of ~13% over a standard choice of equal number binning
  - Misclassification NNC performs better than the Outlier NNC of Broussard & Gawiser 2021
- Full quantification of these improvements requires an end-to-end cosmological parameter estimate to see if augmentation can reduce biases
- Our method can optimize the tomographic binning strategy for any 3x2pt lens galaxy sample