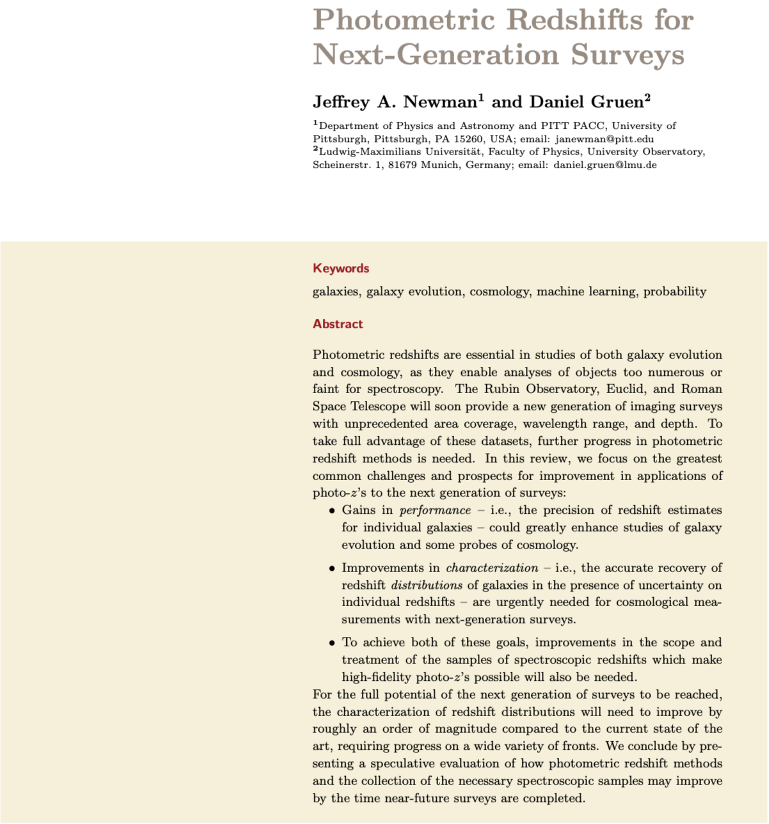
Spectroscopic Samples for Photometric Redshifts: Systematic Biases and Unknown Unknowns

Jeffrey Newman, U. Pittsburgh / PITT-PACC

Thanks to the US Department of Energy Office of Science,
Office of High Energy Physics for support!

Achieving the full potential of next-generation surveys via photometric redshifts presents many challenges...

- **In our recent *Annual Reviews* article, Daniel Gruen and I present key challenges for photometric redshifts with Rubin Observatory, Euclid, Roman, etc., that affect galaxy evolution & cosmology studies**
  - **Journal version:** https://www.annualreviews.org/doi/abs/10.1146/annurev-astro-032122-014611
  - **ArXiv version (with some formatting advantages):** https://arxiv.org/abs/2206.13633
- **Our focus is on ways we need to improve both:**
  - **The performance of photo-z algorithms: how well we can predict the redshifts and other properties of individual objects**
  - **The calibration of redshift distributions, which must reach <0.1% levels to not dominate over random errors in cosmology**



Photometric Redshifts for
Next-Generation Surveys

Jeffrey A. Newman[1] and Daniel Gruen[2]

[1]Department of Physics and Astronomy and PITT PACC, University of Pittsburgh, Pittsburgh, PA 15260, USA; email: janewman@pitt.edu
[2]Ludwig-Maximilians Universität, Faculty of Physics, University Observatory, Scheinerstr. 1, 81679 Munich, Germany; email: daniel.gruen@lmu.de

Keywords

galaxies, galaxy evolution, cosmology, machine learning, probability

Abstract

Photometric redshifts are essential in studies of both galaxy evolution and cosmology, as they enable analyses of objects too numerous or faint for spectroscopy. The Rubin Observatory, Euclid, and Roman Space Telescope will soon provide a new generation of imaging surveys with unprecedented area coverage, wavelength range, and depth. To take full advantage of these datasets, further progress in photometric redshift methods is needed. In this review, we focus on the greatest common challenges and prospects for improvement in applications of photo-z's to the next generation of surveys:

- Gains in *performance* – i.e., the precision of redshift estimates for individual galaxies – could greatly enhance studies of galaxy evolution and some probes of cosmology.
- Improvements in *characterization* – i.e., the accurate recovery of redshift *distributions* of galaxies in the presence of uncertainty on individual redshifts – are urgently needed for cosmological measurements with next-generation surveys.
- To achieve both of these goals, improvements in the scope and treatment of the samples of spectroscopic redshifts which make high-fidelity photo-z's possible will also be needed.

For the full potential of the next generation of surveys to be reached, the characterization of redshift distributions will need to improve by roughly an order of magnitude compared to the current state of the art, requiring progress on a wide variety of fronts. We conclude by presenting a speculative evaluation of how photometric redshift methods and the collection of the necessary spectroscopic samples may improve by the time near-future surveys are completed.

Many of these challenges are connected to the spectroscopic datasets we use to train and calibrate photo-z's

- **For machine learning-based methods, we use objects with secure redshift measurements from spectroscopy to train algorithms**
- **For template-based methods, they are still needed to refine models of galaxy spectral energy distributions**
- **However, obtaining large numbers of spectra is expensive: we are typically limited to only brighter objects**
  - **The range of colors / SEDs of bright galaxies does not span the range that faint galaxies cover: an example of domain shift**

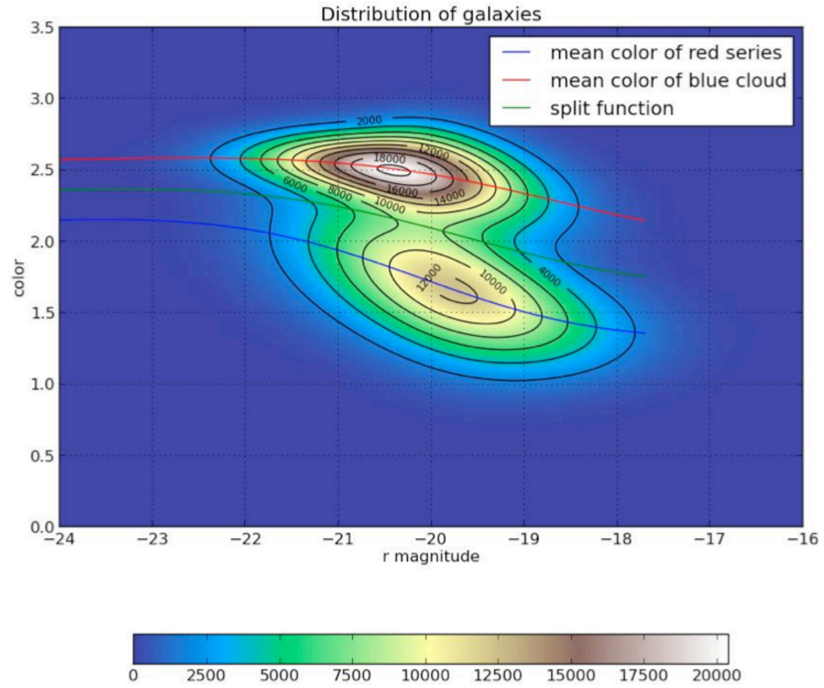


Figure: Zach Pace

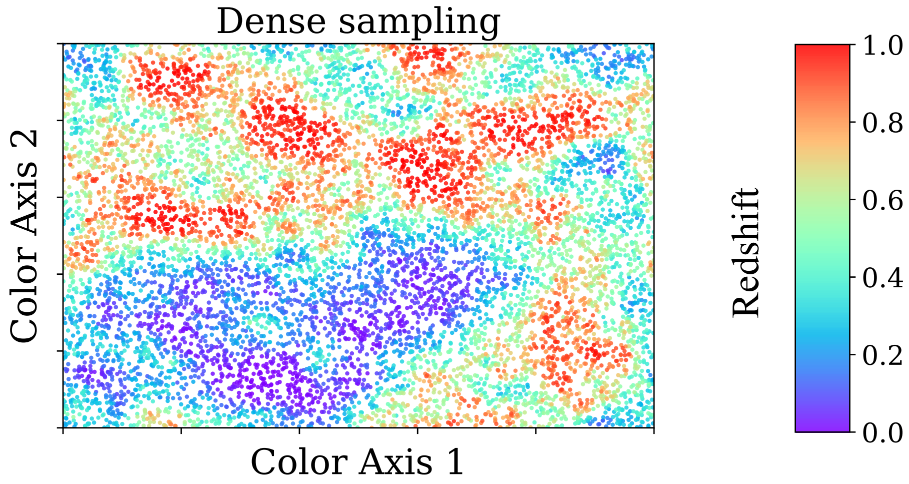
Alexandra Ciprijanovic:

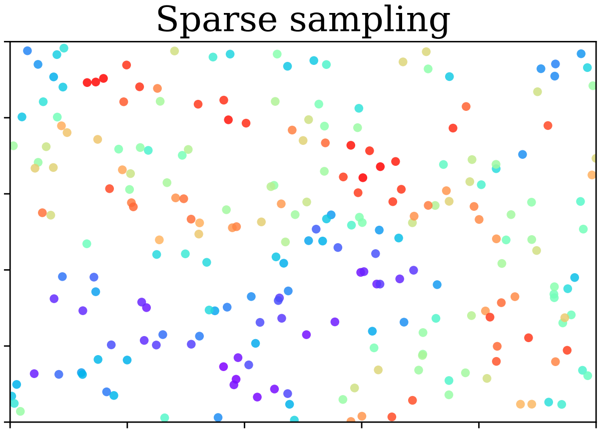
- How do we **fight against unknown unknowns causing the domain shift**?

Some of the ways that real-world spectroscopic datasets fall short of the ideal: 1) sparse sampling

- **Ideal case: we obtain redshifts for objects densely and evenly sampling the distribution of galaxies in SED space**
  - **Here, a toy model: e.g., what you would get dimensionality-reducing SED space to 2D**
  - **Can easily determine redshift at any point from redshifts of objects in the local neighborhood**

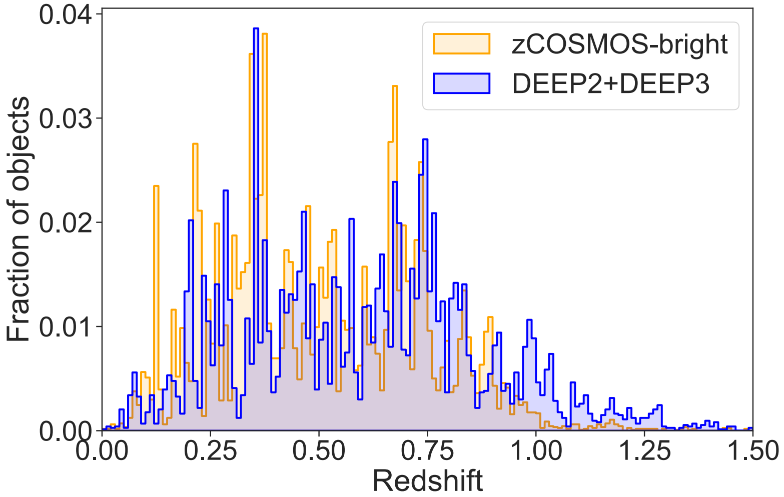
- **Real world: if we want spectroscopy of faint galaxies, sample sizes will be small and will only sparsely cover SED space**
  - **The objects with spectroscopy in the same neighborhood may not be all that close...**



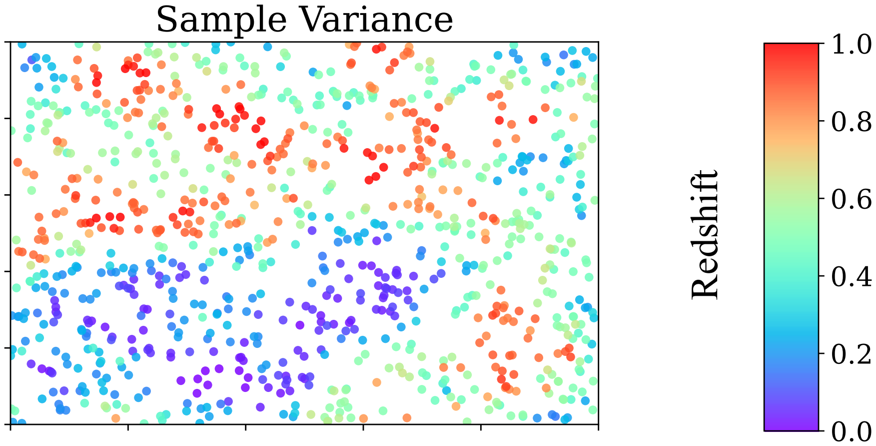


Some of the ways that real-world spectroscopic datasets fall short of the ideal: 2) sample / cosmic variance

- **Ideal case: the redshifts in your spectroscopic training set have a redshift distribution matching the overall average across the sky**

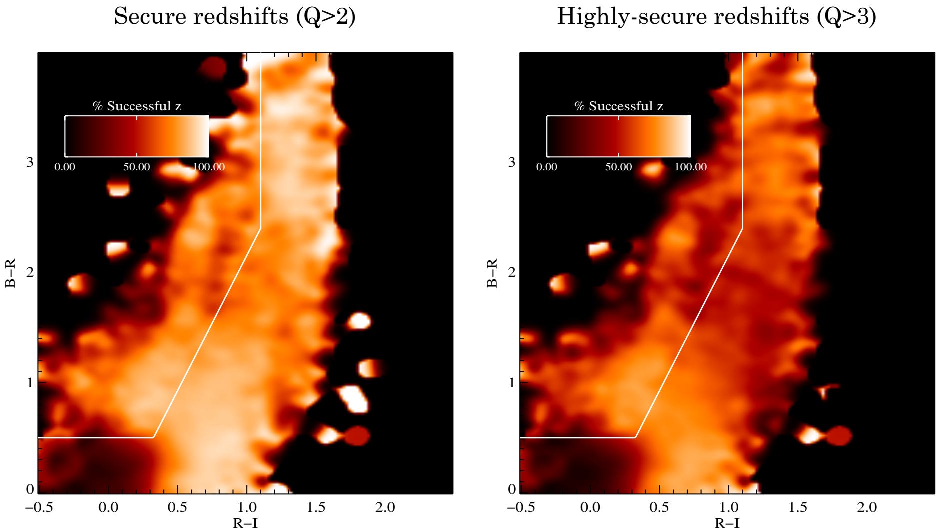


- **Real world: Deep training sets are obtained over only small areas of sky**
  - **The selected regions will be overdense or underdense at some redshifts due to large-scale structure**
  - **This can easily imprint on redshift distributions across the sky with ML methods**

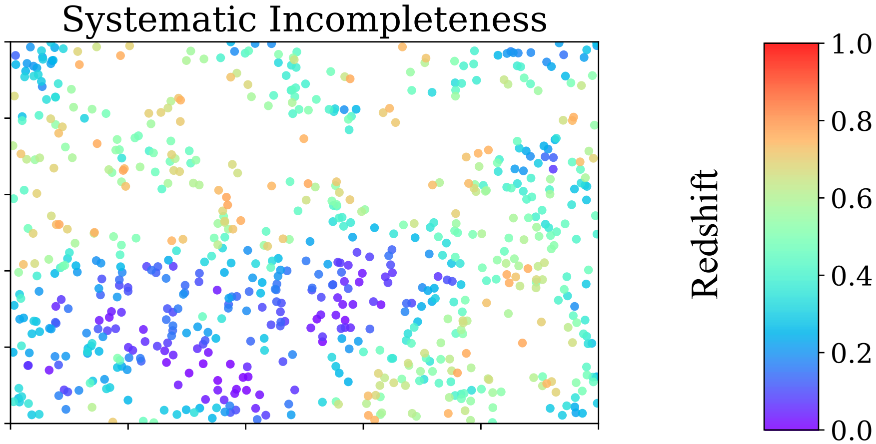


Some of the ways that real-world spectroscopic datasets fall short of the ideal: 3) systematic incompleteness

- **Ideal case: every galaxy you target for spectroscopy provides a secure measurement of its redshift**



- **Real world: When we target faint samples, we fail to measure the redshift ~30% or more of the time**
  - **The objects we do get redshifts for are systematically different in properties (including redshift) than the things we succeed for**



Some of the ways that real-world spectroscopic datasets fall short of the ideal: 4) incorrect redshifts

- **Ideal case: every time you measure the redshift spectroscopically you get the correct *z***

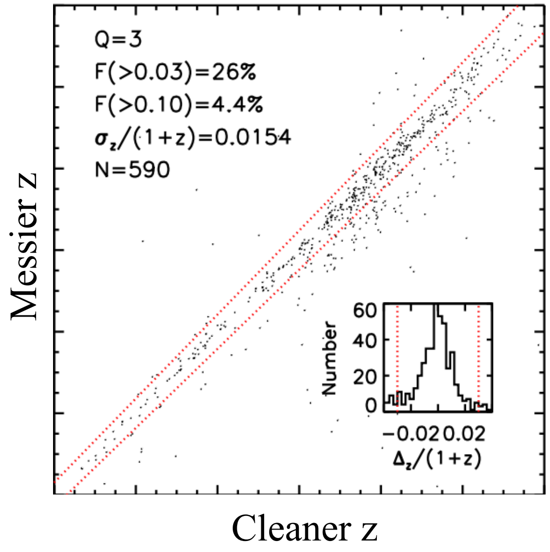
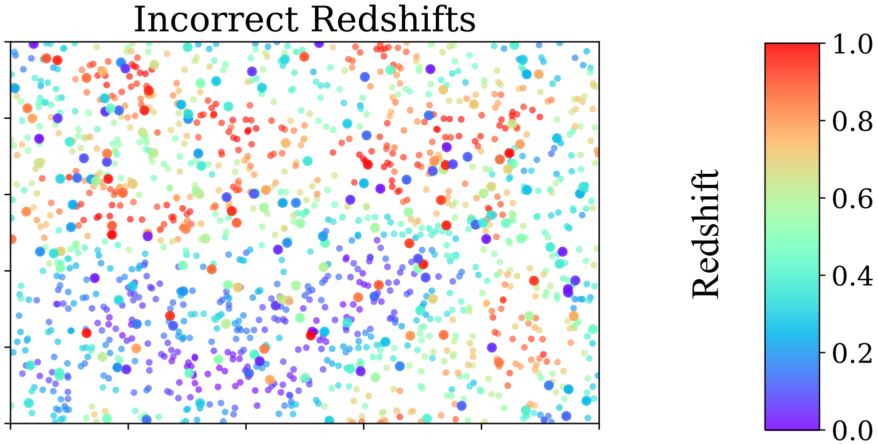


Figure: Coil et al. 2010

- **Real world: Depending upon the sample, 0.5%-10% of redshift measurements will be incorrect**
  - **E.g.: misidentified a single emission line, or mistook sky subtraction residuals for lines**
  - **Need robust ML methods for photometric redshifts**



Some of the ways that real-world spectroscopic datasets fall short of the ideal: 5) color selections

- **Ideal case: you can just use redshifts from pre-existing spectroscopic surveys and don't need to obtain any new measurements**

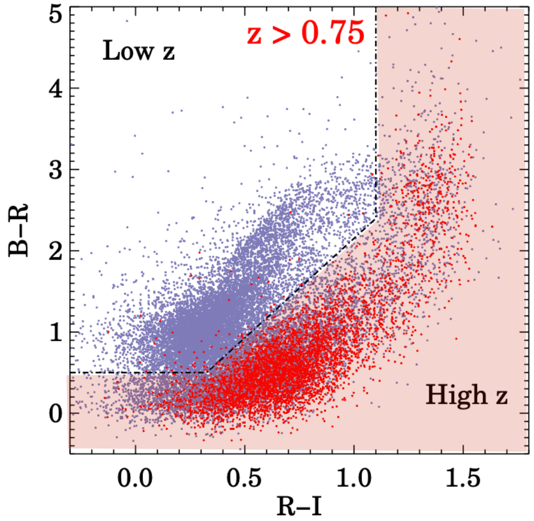
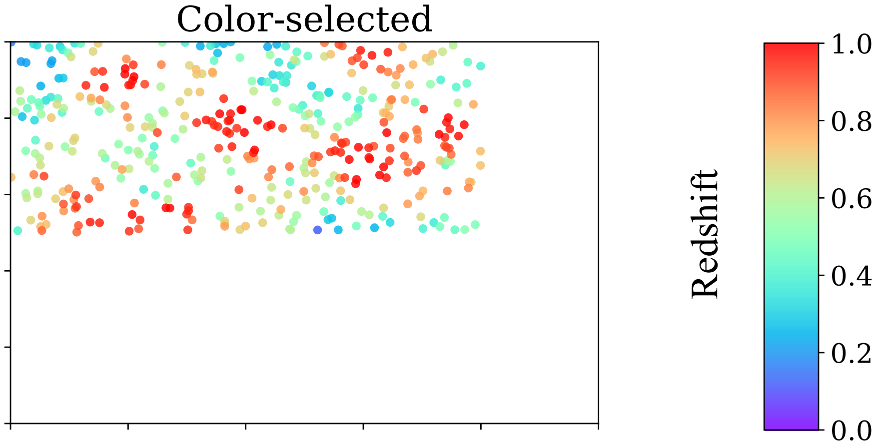


Figure: Newman et al. 2013

- **Real world: Most large high-*z* surveys rely on color cuts to target a limited redshift range of interest**
- **Heterogeneous coverage of color space is a major problem in photo-z training and calibration**



Some of the ways that real-world spectroscopic datasets fall short of the ideal: 6) difficulties training at very low *z*

- **Ideal case: Training samples provide good coverage across all possible redshifts**

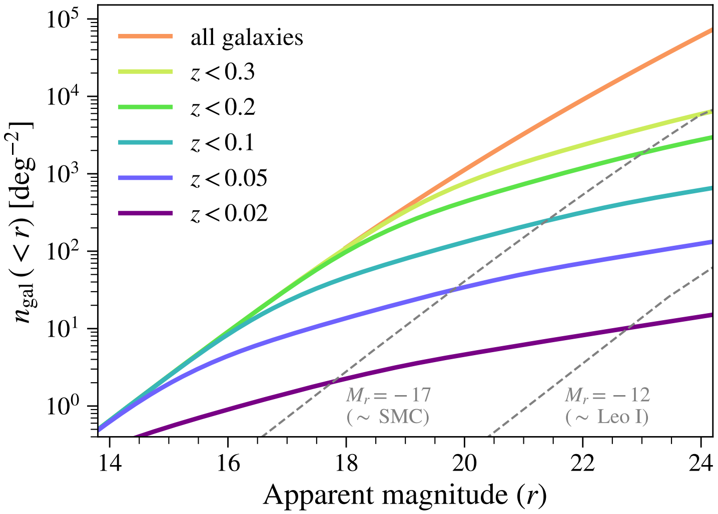
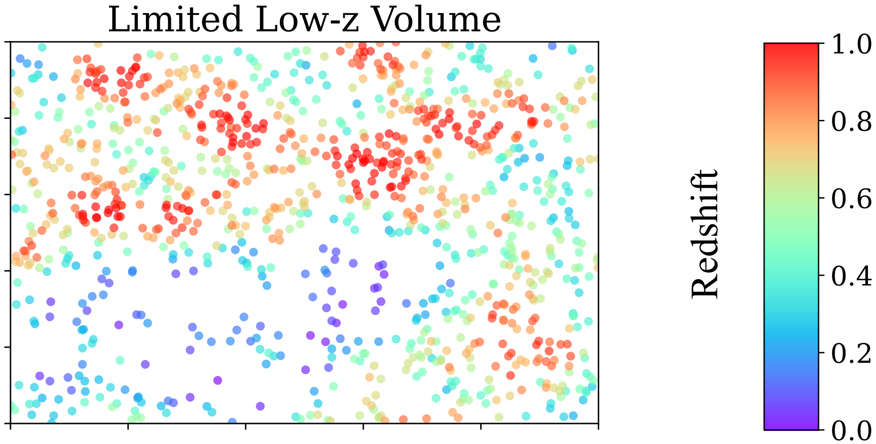


Figure: MSE team / Yao-Yuan Mao

- **Real world: The universe has little volume at low redshifts so low-z galaxies are rare in magnitude-limited samples**
- **Since they are poorly represented in training sets photo-*z* algorithms tend to disfavor low z solutions**



Do template-based photo-z's solve these problems by being less dependent on training sets?

- **Nope...**
- **Kodra et al. 2022 tested many template-based methods applied to CANDELS data**
- **Methods that all agree well with spec-z's where we have them predict very different redshift distributions vs. magnitude, from the same data**

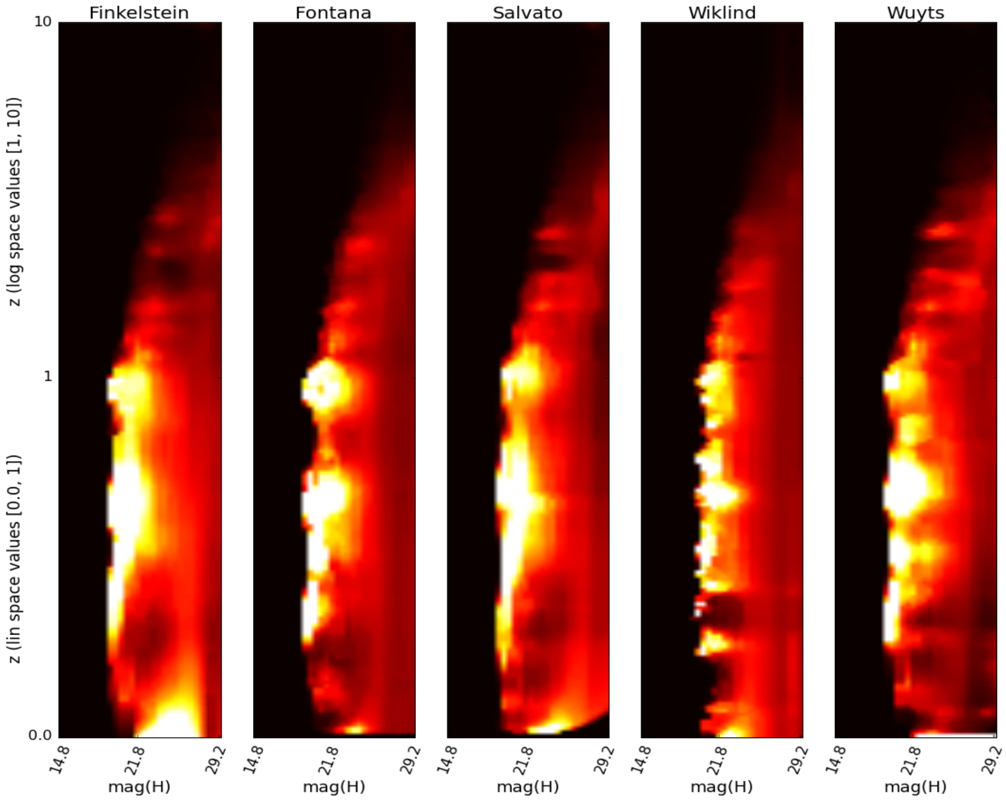
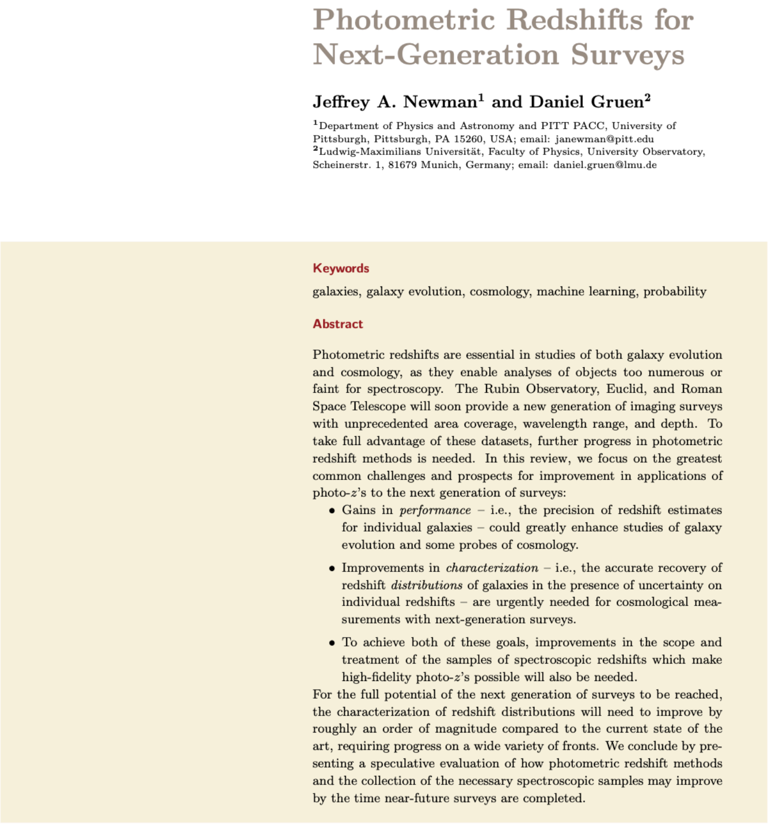


Figure: Kodra et al. 2022

Conclusions

- One of the first problems where machine learning was applied to astronomy was for finding photometric redshifts... but we're still not done
- Upcoming datasets are sufficiently powerful that we will need to continue to advance the state of the art to take full advantage of them
- Spectroscopic training sets pose many challenges
  - We will need large allocations of time on many of the largest telescopes to get datasets at all approaching the ideal
- For lots more details, see our ARAA article!
  - **Journal version:** https://www.annualreviews.org/doi/abs/10.1146/annurev-astro-032122-014611
  - **ArXiv version (with some formatting advantages):** https://arxiv.org/abs/2206.13633



Photometric Redshifts for Next-Generation Surveys

Jeffrey A. Newman[1] and Daniel Gruen[2]

[1]Department of Physics and Astronomy and PITT PACC, University of Pittsburgh, Pittsburgh, PA 15260, USA; email: janewman@pitt.edu
[2]Ludwig-Maximilians Universität, Faculty of Physics, University Observatory, Scheinerstr. 1, 81679 Munich, Germany; email: daniel.gruen@lmu.de

Keywords

galaxies, galaxy evolution, cosmology, machine learning, probability

Abstract

Photometric redshifts are essential in studies of both galaxy evolution and cosmology, as they enable analyses of objects too numerous or faint for spectroscopy. The Rubin Observatory, Euclid, and Roman Space Telescope will soon provide a new generation of imaging surveys with unprecedented area coverage, wavelength range, and depth. To take full advantage of these datasets, further progress in photometric redshift methods is needed. In this review, we focus on the greatest common challenges and prospects for improvement in applications of photo-z's to the next generation of surveys:

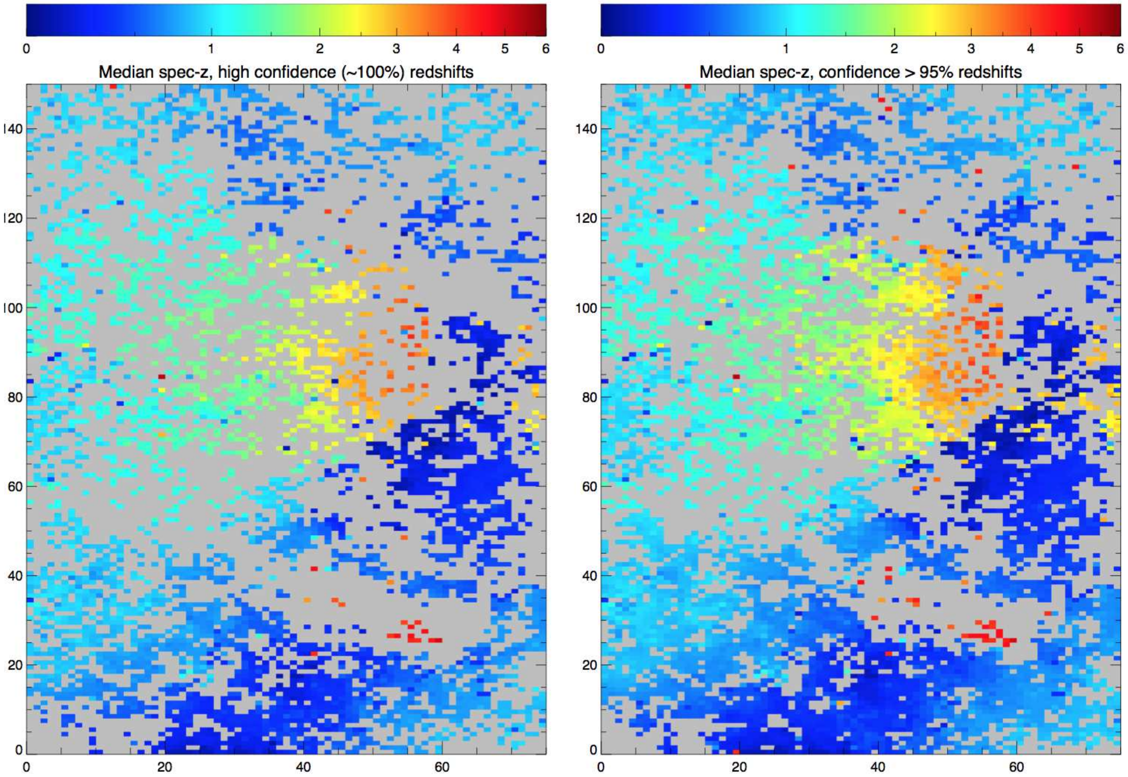
- Gains in *performance* – i.e., the precision of redshift estimates for individual galaxies – could greatly enhance studies of galaxy evolution and some probes of cosmology.
- Improvements in *characterization* – i.e., the accurate recovery of redshift *distributions* of galaxies in the presence of uncertainty on individual redshifts – are urgently needed for cosmological measurements with next-generation surveys.
- To achieve both of these goals, improvements in the scope and treatment of the samples of spectroscopic redshifts which make high-fidelity photo-z's possible will also be needed.

For the full potential of the next generation of surveys to be reached, the characterization of redshift distributions will need to improve by roughly an order of magnitude compared to the current state of the art, requiring progress on a wide variety of fronts. We conclude by presenting a speculative evaluation of how photometric redshift methods and the collection of the necessary spectroscopic samples may improve by the time near-future surveys are completed.

Thanks to the US Department of Energy Office of Science, Office of High Energy Physics for support!

If we restrict to the most-secure redshifts, much more of color space is untrained by current samples

- **Grey regions: cells in self-organized maps of galaxy color space that are not constrained by spectroscopic redshifts**



cells with <1% failure rate z's

with <5% failure rate z's

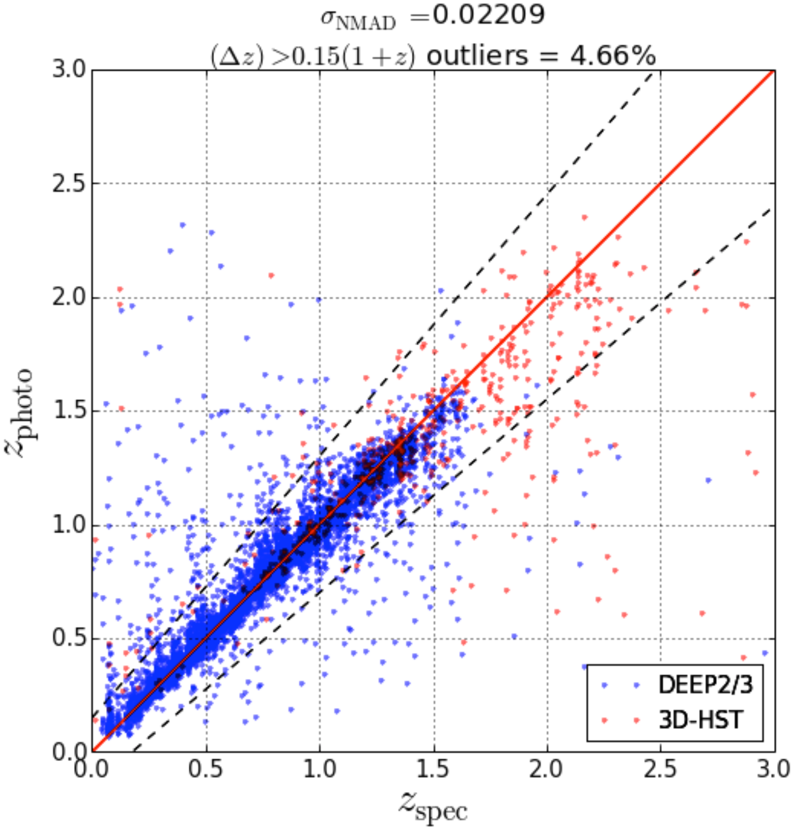
Masters et al. 2015

An additional issue: some photo-z/spec-*z* outliers are physical

- A few percent of DEEP2 spectroscopic targets correspond to multiple galaxies when you look at HST catalogs

- 1% of DEEP2 objects show spectral features from multiple redshifts

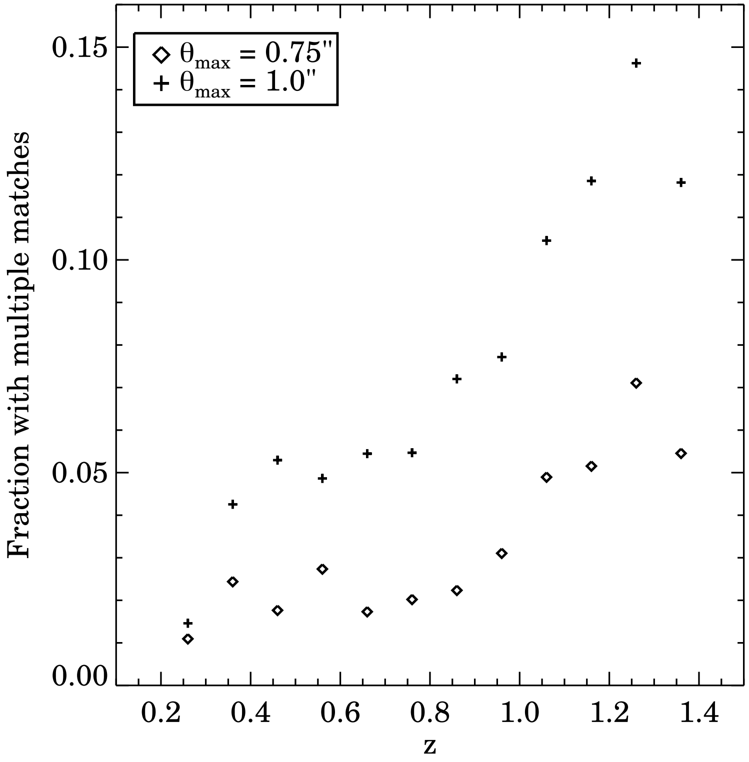
- Can identify many but NOT all of these blends with space-based imaging



Zhou, Cooper, JN et al. 2019

An additional issue: some photo-z/spec-*z* outliers are physical

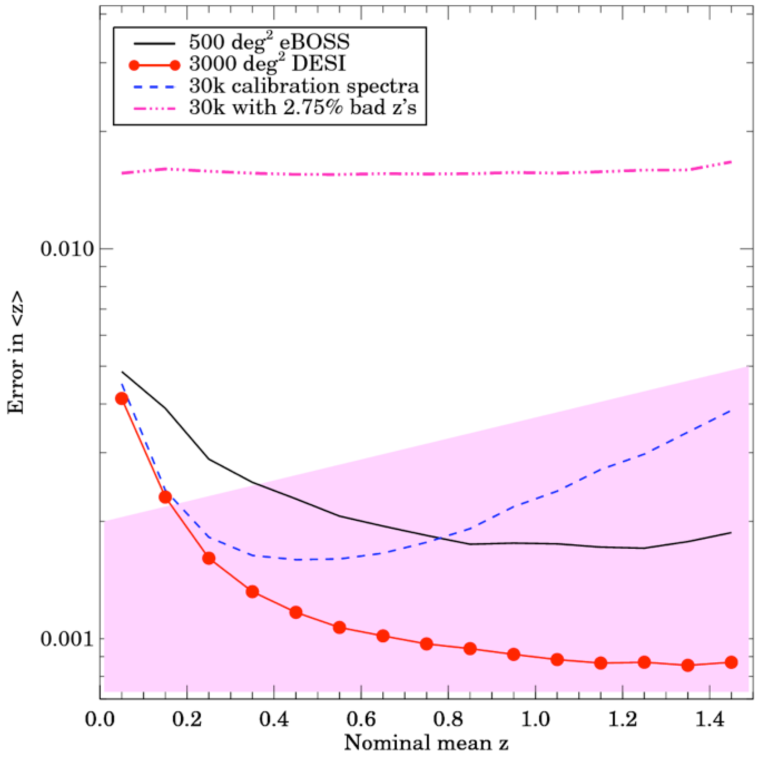
- **A few percent of DEEP2 spectroscopic targets correspond to multiple galaxies when you look at HST catalogs**
- **1% of DEEP2 objects show spectral features from multiple redshifts**
- **Can identify many but NOT all of these blends with space-based imaging**



Newman et al. 2013

If spectroscopy proves incomplete, calibration will probably need to come from cross-correlation methods...

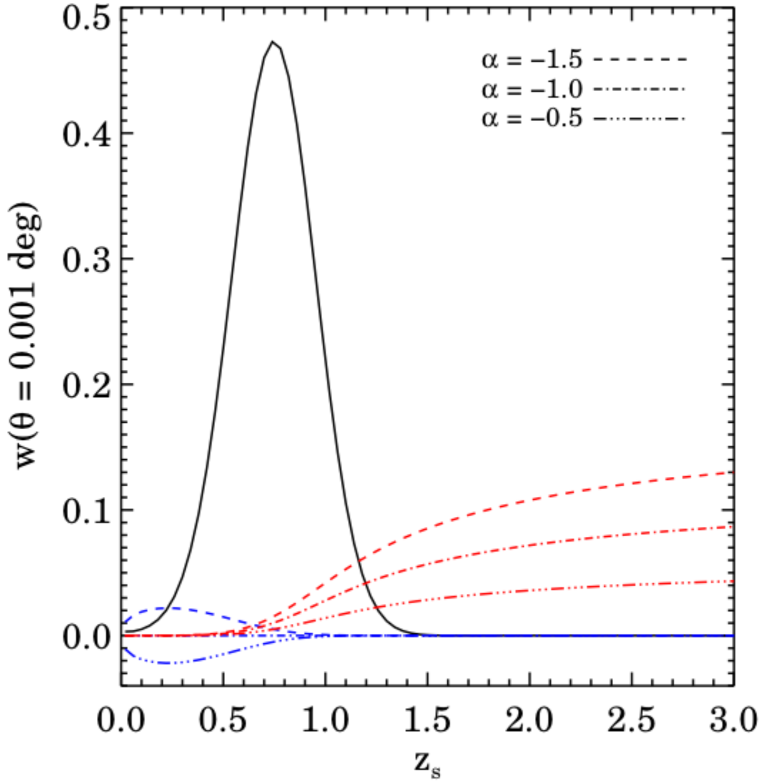
- Galaxies of all types cluster together: trace same dark matter distribution
- Enables reconstruction of *z* distributions via spectroscopic/photometric cross-correlations (Newman 2008)
- For LSST calibration, >500 degrees of overlap with DESI-like survey would meet LSST science requirements (>4000 sq deg of overlap expected)
  - ... IF LSST data is uniform (after calibration), as DESI is in North



Snowmass white paper: *Spectroscopic Needs for Imaging DE Experiments* (Newman et al. 2015, http://arxiv.org/abs/1309.5388)

Biggest concern: disentangling cross-correlations from clustering and lensing magnification

- *Black*: cross-correlations between photo-z objects (z=0.75 Gaussian) and spectroscopic sample as a function of *z*
- *Blue*: observed cross-correlation due to spectroscopic objects lensing photometric ones
- *Red*: observed cross-correlation due to photometric objects lensing spectroscopic ones
- Weak/CMB lensing could help us predict the red curves

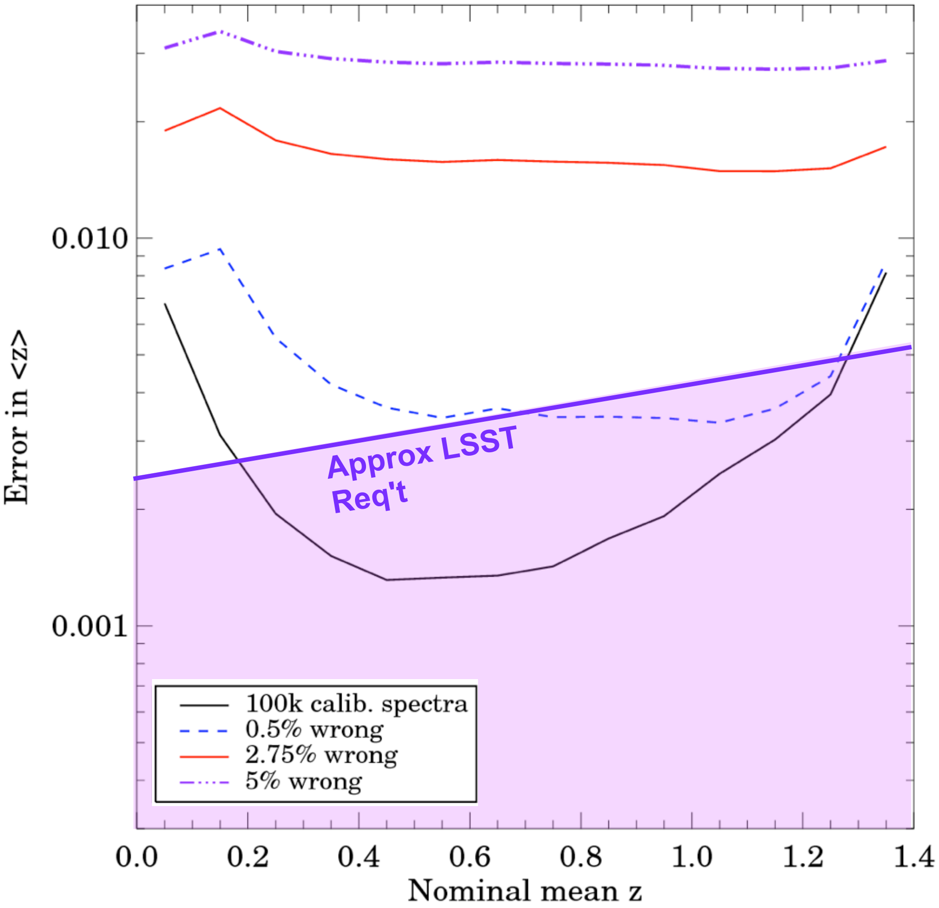


Daniel Matthews Ph. D. thesis, 2014

A dif

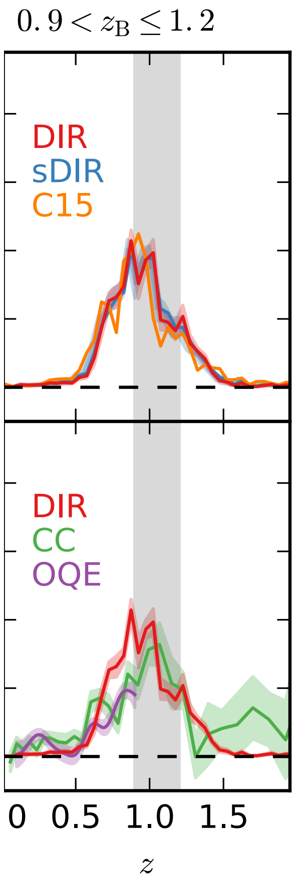
Note: even for 100% complete samples, current false-z rates would be a problem

- Only the highest-confidence redshifts should be useful for precision calibration: lowers spectroscopic completeness further when restrict to only the best
- A major reason why getting highly secure redshifts is important

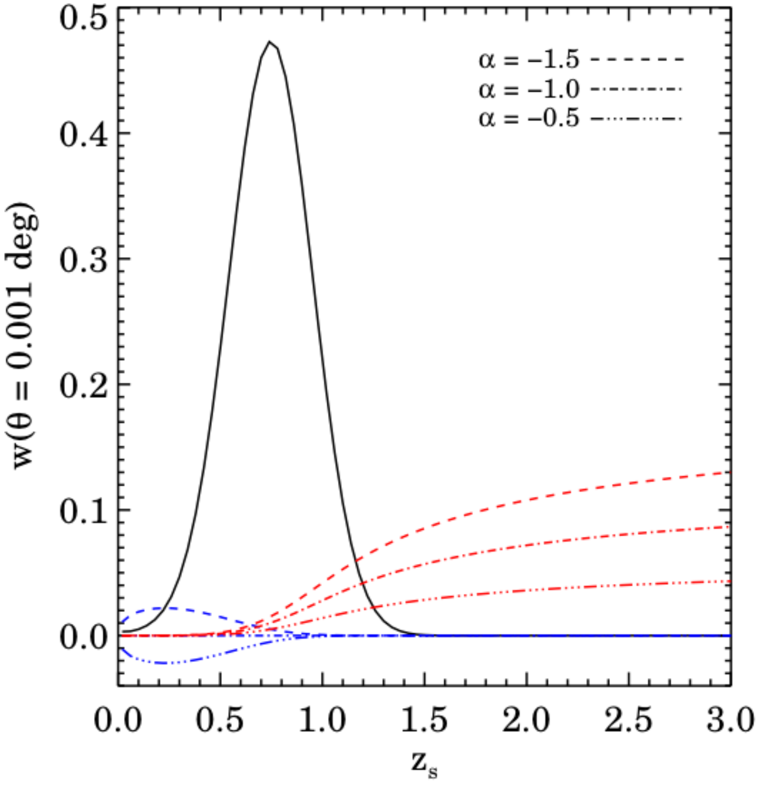


Based on simulated redshift distributions for ANNz-defined DES bins in mock catalog from Huan Lin, UCL & U Chicago, provided by Jim Annis

Biggest concern: disentangling cross-correlations from clustering and lensing magnification



Hildebrandt et al. 2018

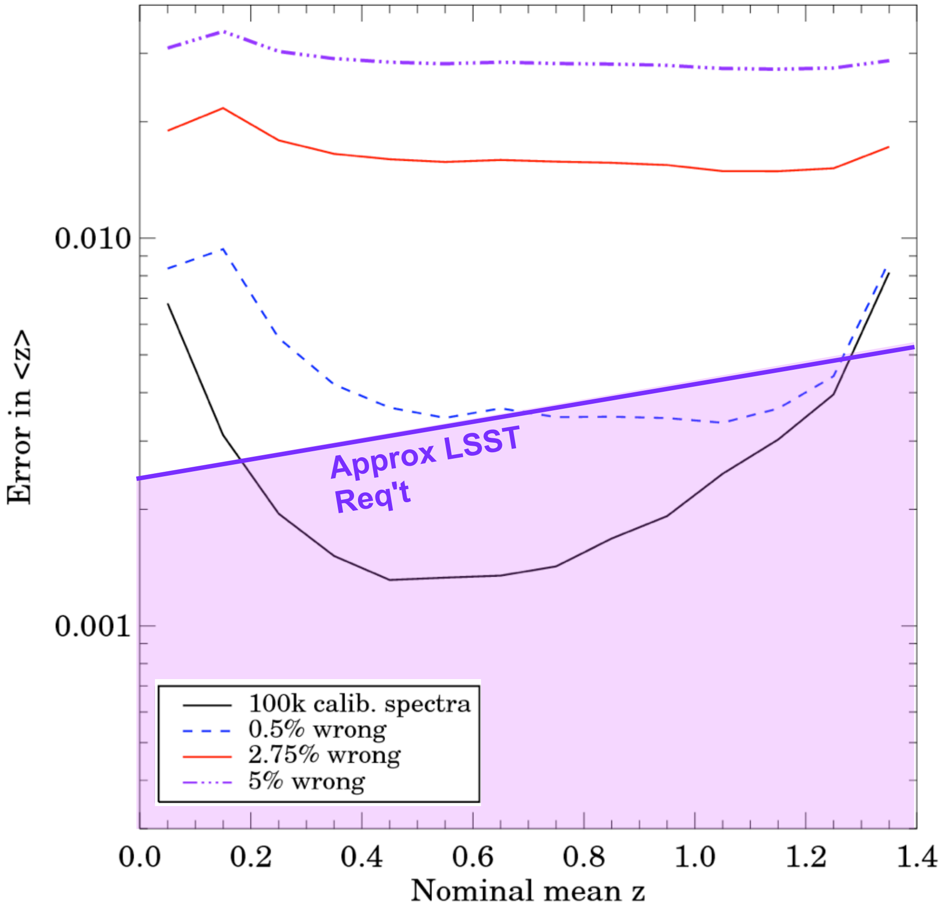


Daniel Matthews Ph. D. thesis, 2014

A dif

Note: even for 100% complete samples, current false-z rates would be a problem

- Only the highest-confidence redshifts should be useful for precision calibration: lowers spectroscopic completeness further when restrict to only the best
- A major reason why getting highly secure redshifts is important



Based on simulated redshift distributions for ANNz-defined DES bins in mock catalog from Huan Lin, UCL & U Chicago, provided by Jim Annis

What might an ideal photo-z algorithm look like?

- What might an ideal LSST/Euclid/Roman photo-z algorithm for the next decade look like?
  - Trained with >30,000 spectra spanning range of photometric objects
  - Develops priors & tweaks templates via hierarchical Bayesian hyperparameters, or via forward-modeling distributions
  - Incorporates variations in effective filter wavelengths due to observational conditions: requires applying algorithm to O(1000) measurements instead of O(6)
  - Incorporates AGN classification and AGN photo-z determination: colors are not constant with time for many objects!
  - Want algorithms to be fast: create ML-based emulators for template photo-z's?
  - For bright objects, may also be useful to compare template to ML techniques to identify potential outliers (different failure modes)