Hunting down systematics in modern galaxy surveys

Mohammadjavad Vakili
Center for Cosmology and Particle Physics
New York University

Berkeley / Cosmology Seminar
2017 January

Outline

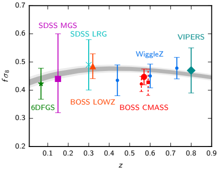
- Large-scale structure mocks for estimation of galaxy clustering covariance matrices
- Weak lensing systematics
  - Point Spread Function
  - Photometric redshifts

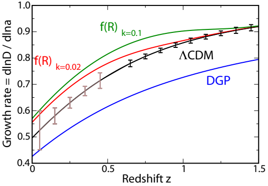
Accurate galaxy mocks for estimation of galaxy clustering covariance matrices

- Based on works in collaboration with: Francisco-Shu Kitaura (IAC), Yu Feng (Berkeley), Gustavo Yepes (UAM), Cheng Zhao (Tsinzua), Chia-Hsun Chuang (Leibniz), ChangHoon Hahn (NYU)

Future of spectroscopic galaxy surveys

- Measurement of growth rate and expansion history with sub-percent precision

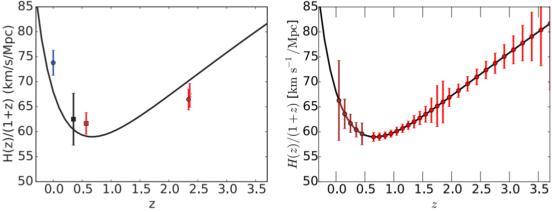




Right: Planck Collaboration XIII (2015), Left: DESI Collaboration (2016)

Future of spectroscopic galaxy surveys

- Measurement of growth rate and expansion history with sub-percent precision



DESI Collaboration (2016)

We need mocks for both precision and accuracy!

- Estimation of uncertainties (covariance matrix)
  - Need **a large number** of mocks ($N_{\mathrm{mock}} >> N_{\mathrm{data}}$)
  - Mocks need to be statistically consistent (1-point, 2-point, 3-point, ...) with the data!
- We live in the era of systematic-limited measurements
  - Need **accurate** end-to-end simulations of galaxy surveys to characterize sytematic uncertainties

How do we efficiently generate mocks for galaxy surveys?

- Requirements:
  - Need to simulate large volumes to sample the BAO signal
  - Need to accurately model nonlinear clustering (current $k_{\mathrm{max}} \sim 0.25\ h\mathrm{Mpc}^{-1}$)
  - Need to resolve low mass halos that host faint galaxies
  - Need to accurately describe two-point and higher-order statistics
- N-body simulations are expensive!

How do we efficiently generate mocks for galaxy surveys?

- Approximate Methods:
  - Approximate (DM-only) structure formation model + Empirical sampling of galaxies/halos from the dark matter field

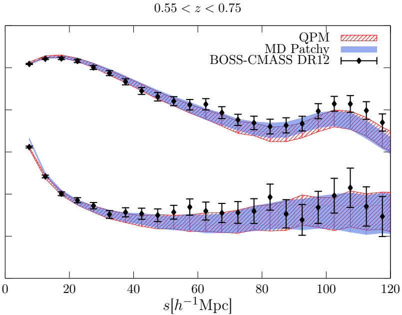
State-of-the-Art: **SDSS III-BOSS**

- QPM (**White *et al.* 2014**)
  - Low resolution N-body
  - Sample halos by matching the mass function and large scale bias
- ALPT-PATCHY (**Kitaura *et al.* 2016**)
  - perturbation theory
  - Sample halos by matching the $n-$point functions

State-of-the-Art Approximate Methods: **SDSS III-BOSS**

Two-point statistics:

$$\xi_0(s), \xi_2(s)$$

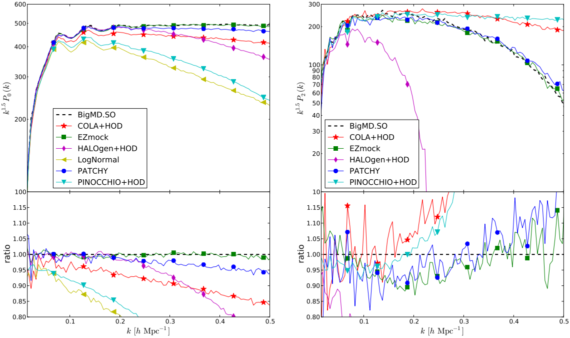


Percentage-level accuracy galaxy mocks

- Precision large-scale structure cosmology requires mocks with percentage-level accuracy!
- Main challenges:
  - Nonlinear Scales
  - RSD
  - Higher order Statistics

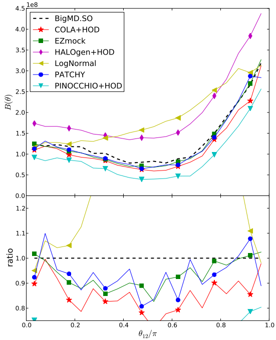
Goal: Percent-level accuracy

- Main Challenges: **(Quasi)Nonlinear Scales**, **RSD** (Chuang *et al.* 2015):



Goal: Percentage-level accuracy

- Main Challenges: **high-order statistics!**
  - BAO detection (Slepian *et al.* 2015)
  - Breaking the degeneracy between f, σ_8 (Gill-Marín *et al.* 2014)



PATCHY : Nonlinear Stochastic Biasing

For a given dark matter density field ρ_m, halos/galaxies are generated from a nonlinear stochastic bias model: (1) Empirical nonlinear bias

$$\langle\rho_g\rangle(\rho_m) = f_g \underbrace{\theta\big(\rho_m - \rho_{th}\big)}_{\text{threshold bias}} \times \underbrace{\rho_m^{\alpha}}_{\text{nonlinear bias}} \times \underbrace{\exp\big(-(\rho/\rho_\epsilon)^{\epsilon}\big)}_{\text{exponential cutoff}}$$

(2) stochastic bias (deviation from Poissoinity):

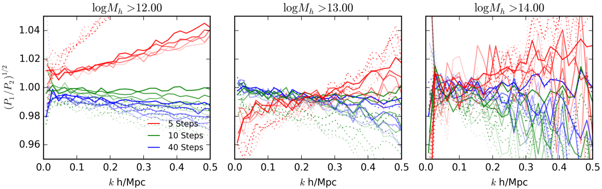
$$\rho_g \sim NB\big(\langle\rho_g\rangle;\beta\big)$$

How can we improve Patchy?

- **Limitations of PATCHY**:
  - Brute-force estimation of bias parameters
  - Limited accuracy of ALPT as a gravity solver
    ALPT = LPT (on large Scales) + SC (on small scales)
- **Solution**:
  - Automatic estimation of bias parameters with MCMC
  - Replacing the gravity solver with an approximate N-body solver that yields a better 1-halo term clustering

New gravity solver: FastPM

- FastPM (Feng *et al.* 2016) : approximate particle mesh N-body solver
- Enforces large-scale linear growth
- Scales well with resolution, time step, force resolution, ...



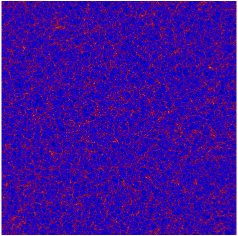
Strategy for generation of mocks

- Generation of a DM field with low resolution N-body
- Constraining the patchy bias parameters by fitting $P(k)$
- Generation of galaxy/halo mocks

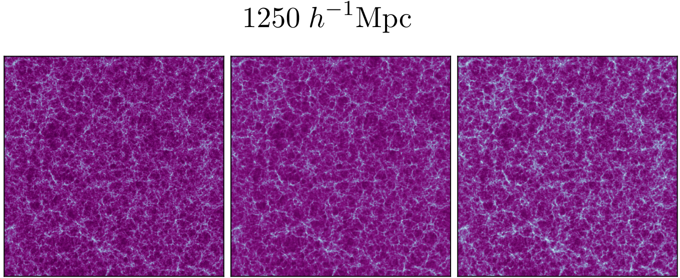
Method is currently being tested as part of the ***Euclid*** covariance project.

Comparison with the BigMultiDark simulation

Can we reproduce the population of *halos* (and *subhalos*) in the BigMultiDark N-body Simulation ($N_p^3 = 3840^3$) with a low-resolution FastPM-PATCHY ($N_p^3 = 960^3$)?



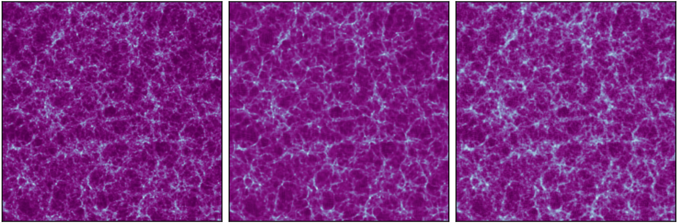
Dark matter density field



From left to right: BigMD, FastPM, ALPT.

Dark matter density ield

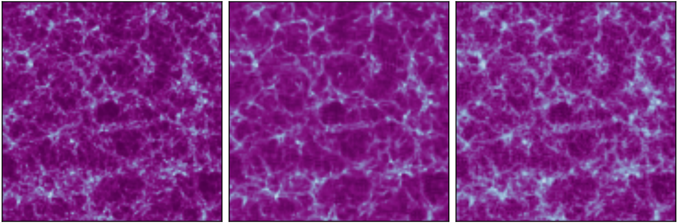
$625\ h^{-1}\mathrm{Mpc}$



From left to right: BigMD, FastPM, ALPT.

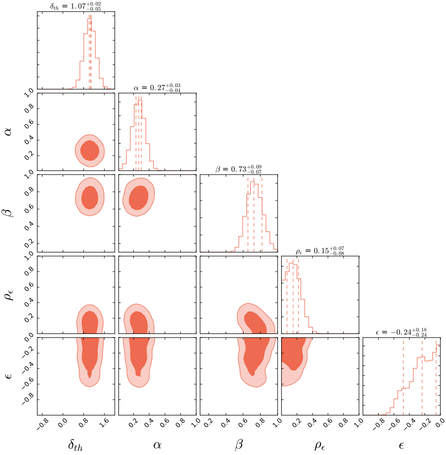
Dark matter density field

$312.5\ h^{-1}\mathrm{Mpc}$



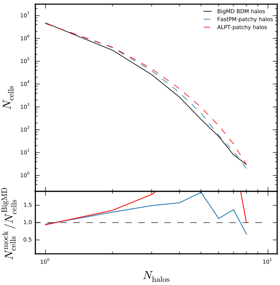
From left to right: BigMD, FastPM, ALPT.

Bias parameters



Comparison with the BigMultiDark Simulation

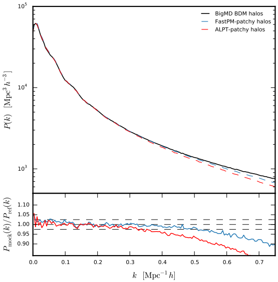
One-point PDF



Vakili *et al.* (2017)

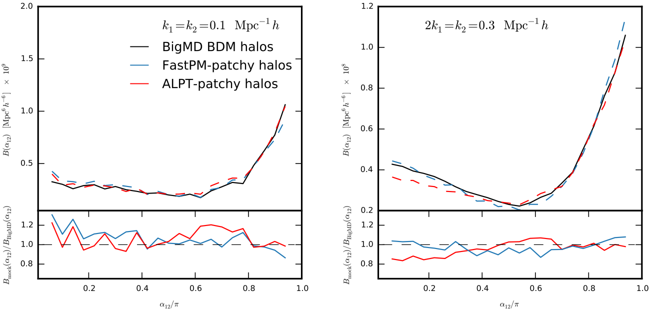
Comparison with BigMultiDark Simulation

Real Space P(k)



Vakili *et al.* (2017)

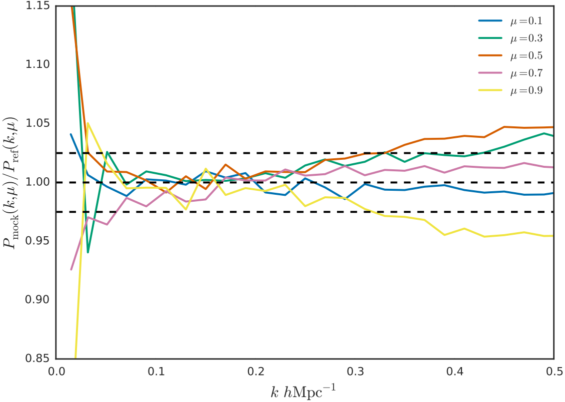
Bispectrum Comparison



Vakili *et al.* (2017)

Anisotropic RSD (Preliminary)

Work in progress!



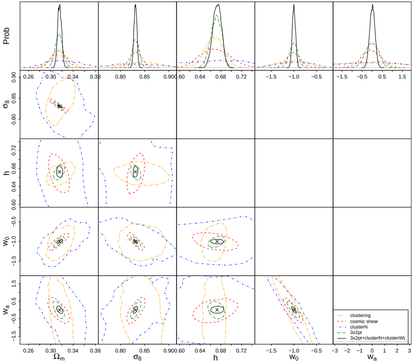
Summary

- We have presented a new version of the PATCHY code with MCMC estimation of bias parameters and FastPM gravity solver.
- By testing our method with the halos in the BigMultiDark simulation, we recover $P(k)$ at $\sim 2\%$ level to high k modes ($k \sim 0.4\ h\mathrm{Mpc}^{-1}$), and the bispectrum at a $\sim 15 - 20\%$ level!
- Redshift space clustering results are not ideal yet! But a different approach for treatment of RSD is currently being developed.

Tackling PSF and photometric redshift systematics in imaging surveys

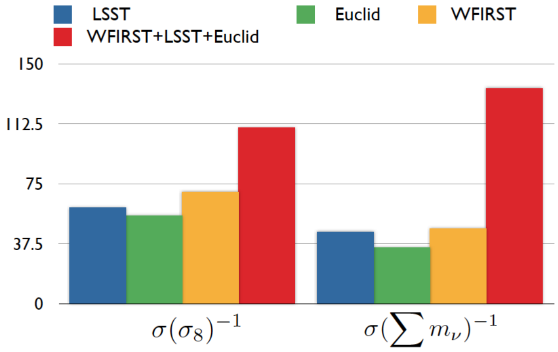
- Based on works in collaboration with:
  David Hogg (NYU, CCA), Alex Malz (NYU)

LSST and the next generation of imaging surveys



Kraus & Eifler 2016

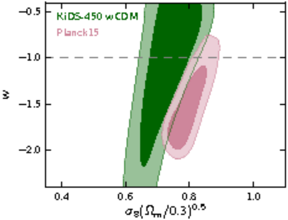
LSST and the next generation of imaging surveys



Jain *et al.* 2015

Weak lensing measurements

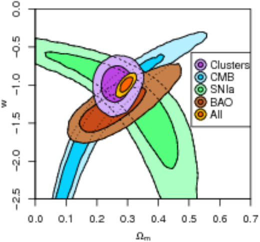
- Weak lensing measurements are the basis of many powerful probes:
  - Cosmic Shear
  - Galaxy Cluster Cosmology
  - Cross-correlation with CMB and galaxies



Hildebrandt *et al.* 2016

Weak lensing measurements

- Weak lensing measurements are the basis of many powerful probes:
  - Cosmic Shear
  - Galaxy Cluster Cosmology
  - Cross-correlation with CMB and galaxies

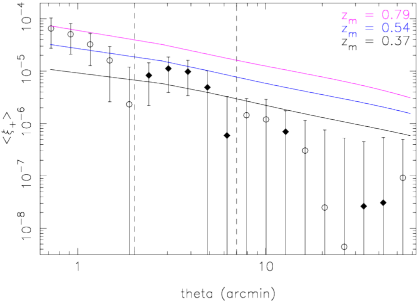


Mantz *et al.* 2014

Weak lensing is limited by systematics

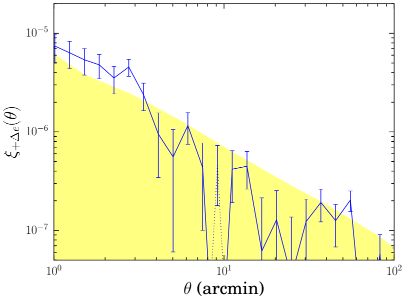
- The problem of inferring the cosmic shear signals from observations is far from idealized. Cosmic shear signal is dominated by:
  - the PSF
  - shape noise
  - Intrinsic alignments
  - and many more: Blending, noise bias, ...

Impact of the PSF (CFHTLenS)



Heyman *et al.* 2011

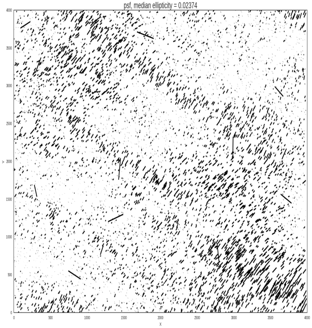
Impact of the PSF (DES)



Jarvis *et al.* 2016

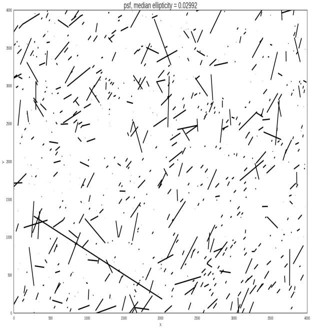
A closer look at the atmospheric PSF

Variation of LSST atmospheric PSF ellipticities across the FoV Simulations run by LSST Photon Simulator (Peterson 2011)



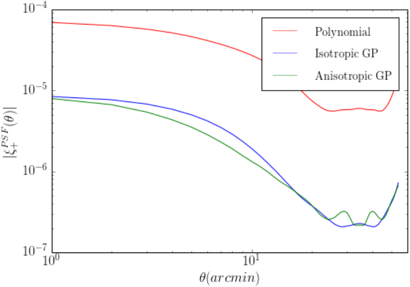
A closer look at the atmospheric PSF

In practice, we can only empirically estimate the PSF at the positions of stars and predict its value elsewhere



LSST Atmospheric turbulence

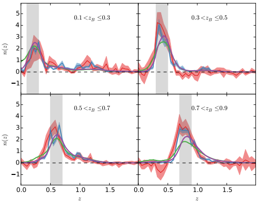
How can we optimally interpolate the PSF?

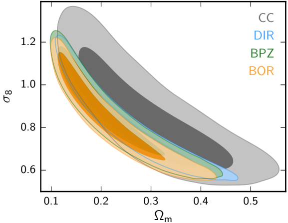


Vakili *et al.* in preparation: Gaussian Process interpolation method beats a more traditional polynomial interpolation. Atmosphere still causes confusion in sub-arcminute scales!

Weak lensing is limited by systematics : the impact of Photo-z's

- Accurate redshift probabilities are needed for tomographic two-point function calculations, determination of redshift distributions, inference of cluster masses.

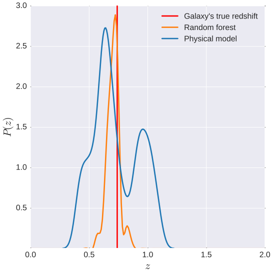




Hildenbrandt *et al.* 2016

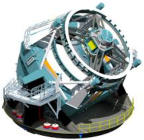
Common photo-z estimation methods

- Template fitting
- Machine Learning
- Cross-correlation with spectroscopic sample

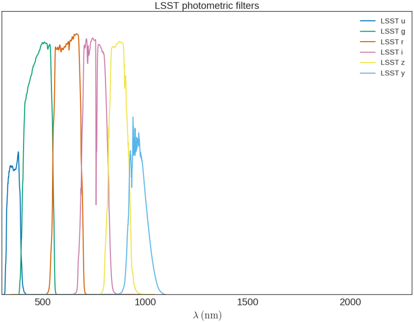


Combining different datasets : WFIRST and LSST

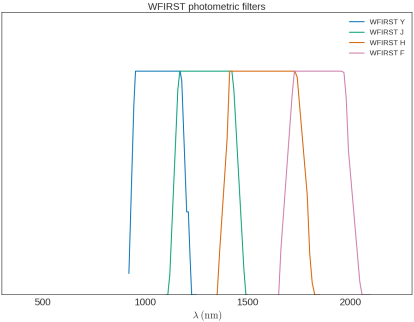
- Accuracy and precision of $P(z)$ for individual galaxies can be enhanced by combining the data from overlapping surveys:



LSST filters



WFIRST filters



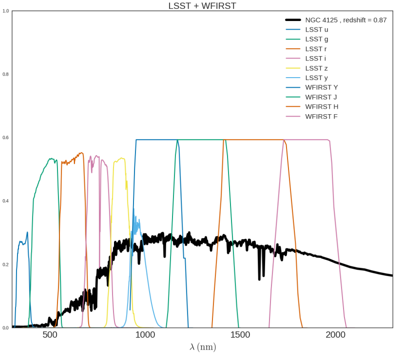
LSST and WFIRST

$$P(z|\hat{\mathbf{F}}, \{\mathrm{SED}_k\}) = \int \prod_k dt_k P(z, t_k|\hat{\mathbf{F}}, \{\mathrm{SED}_k\})$$

$$\hat{\mathbf{F}} = \{\hat{F}_{\mathrm{LSST}}, \hat{F}_{\mathrm{WFIRST}}\}$$

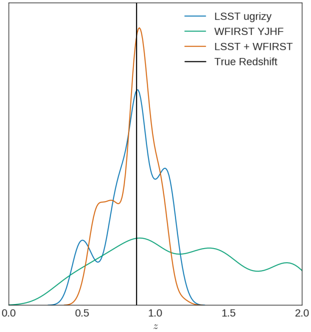
Template library $\{\mathrm{SED}_k\}$ from Brown *et al.* (2014) used in LSST DC1.

$P(z|\hat{\mathbf{F}})$ with single exposure LSST and WFIRST?

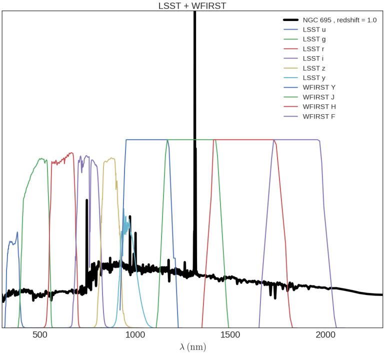


$P(z|\hat{\mathbf{F}})$ with single exposure LSST and WFIRST?

WFIRST photo-z is limited by distinguishing galaxy SED's at WFIRST wavelengths

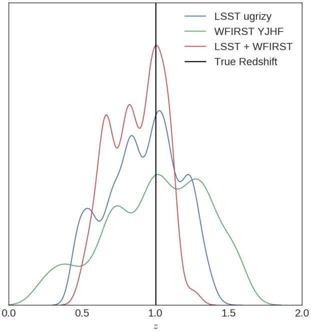


$P(z|\hat{\mathbf{F}})$ with single exposure LSST and WFIRST?



$P(z|\hat{\mathbf{F}})$ with single exposure LSST and WFIRST?

WFIRST photo-z is limited by distinguishing galaxy SED's at WFIRST wavelengths



$n(z)$ with single exposure LSST and WFIRST?

How well can we recover the redshift distributions?

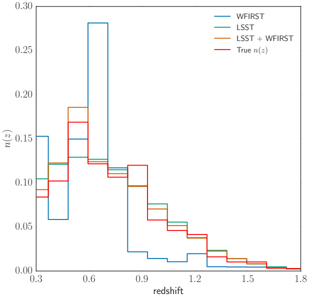
$p(\mathcal{N}|\{d_k\}) \propto$

$p(\mathcal{N}) \exp[-\int \mathcal{N}(z)dz] \times \prod_k \int \frac{p(z_k|d_k)}{p(z_k)} dz_k$

$n(z) = \frac{d\mathcal{N}}{dz}$

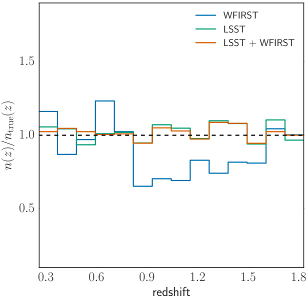
$n(z)$ with single exposure LSST and WFIRST?

How well can we recover the redshift distributions?



$n(z)$ with single exposure LSST and WFIRST?

How well can we recover the redshift distributions?



How do we optimally combine different datasets

- Treat different datasets independently
- Simultaneously constrain photometry and shapes with both datasets:

$$P(\hat{\mathbf{F}}, e|\mathbf{d}_{\mathrm{pixel}})$$

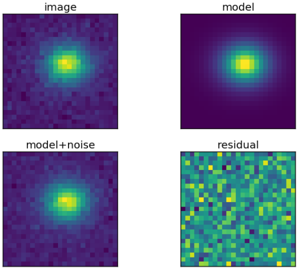
where

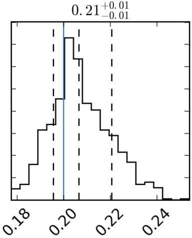
$$\mathbf{d}_{\mathrm{pixel}}$$

is the pixel-level data from all band-passes

How do we optimally combine different datasets

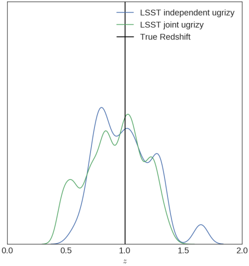
- Real World scenario:





Joint vs Independent modeling of bandpasses

- Joint modeling of all band-passes at the pixel level could mitigate the biases in flux estimates and hence the redshifts



Summary

- The impact of PSF residual systamtics can be controlled if we use a more flexible Gaussian Process model for PSF interpolation.
- We have presented results showing that **accuracy** and **precision** of photometric redshift probabilities can be enhanced by combining datasets.
- Joint modeling of all bandpasses at the pixel level leads to more robust photometric redshift estimation.