

Big Data? Big LSS!

Eric Gawiser

(Rutgers, LSST-DESC Deputy Spokesperson)

(A Highly Biased Sample of)

Large Galaxy Surveys

LSST Dark Energy Science Collaboration (DESC)

- Optical survey; photo-z only
- Will discover >10 billion galaxies from 2022-2032
- Distance measurements are highly imprecise
 - Milky Way dust strongly impacts the apparent LSS
 - Galaxies will be sorted into ~10 bins => contamination.

Hobby-Eberly Telescope Dark Energy Experiment (HETDEX)

- Will detect 1 million Emission Line galaxies 2018-2022
- Relative probability that a given EL galaxy high- or low-z => contamination

Simultaneous estimation of LSS and MW dust extinction from galaxy surveys

Matías Bravo (ICRAR, U. Western Australia)

Eric Gawiser (Rutgers)

Nelson Padilla (P.U. Católica-Chile)

Milky Way dust maps

Measurements of dust from IR emission:

- Schlegel, Finkbeiner & Davis (1998)
- Planck Collaboration et al. (2014)

Measurements from dust extinction of stars:

- Schlafly et al. (2014)

What about using galaxy statistics?

- Can use counts, magnitudes and colors to measure the extinction.
- Li et al. (2017) used counts and colors from SCUSS + SDSS to test the SFD map.
- Current galaxy catalogues with wide area coverage (like SDSS) are not deep enough to produce useful maps at even $\sim 1^\circ$ resolution.

What about using galaxy statistics?

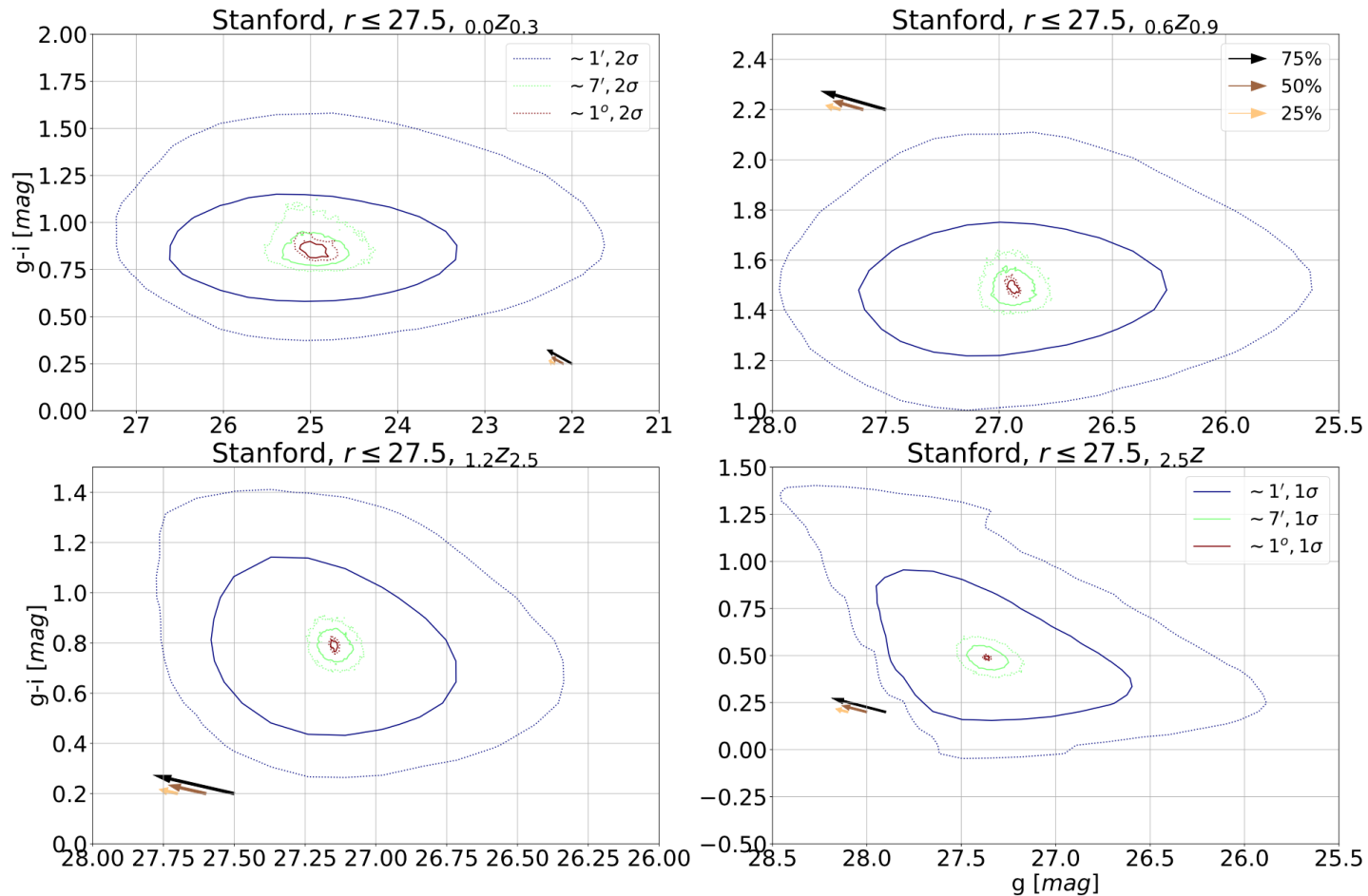
- Can use counts, magnitudes and colors to measure the extinction.
- Li et al. (2017) used counts and colors from SCUSS + SDSS to test the SFD map.
- Current galaxy catalogues with wide area coverage (like SDSS) are not deep enough to produce useful maps at even $\sim 1^\circ$ resolution.

Solution: LSST

Proposed method for map construction

- Divide the sky into HEALPixels and redshift bins.
- In each pixel, measure overdensity, median galaxy magnitude, and median galaxy color.
- Use displacement of pixels in that 3-D space versus its centroid to infer the MW dust reddening and LSS.
- Determine uncertainties via the dispersion of *simulated* pixels (without dust) along the vector from the MW dust extinction law.

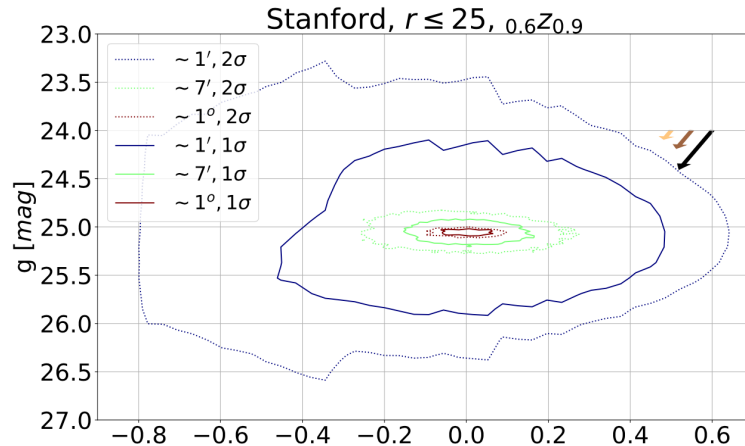
Color-magnitude diagrams!



1.2 < z < 2.5 bin most constraining

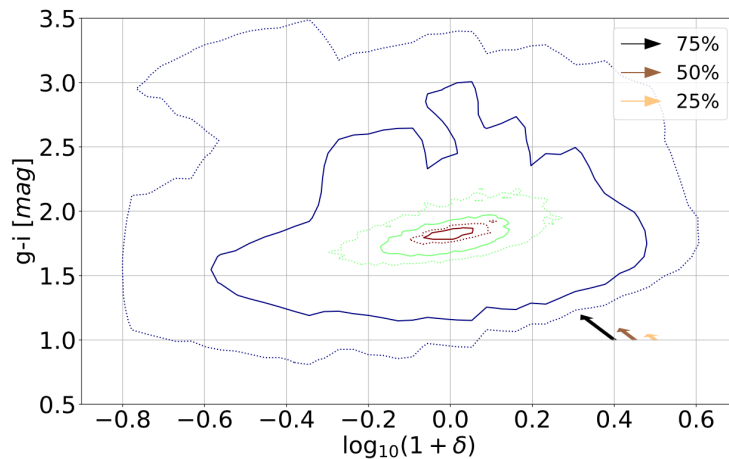
Use inverse-variance weighting to combine bins

Overdensity vs. color & magnitude



Overdensity is the LSS we want to measure.

It does not correlate with median brightness but correlates mildly with color.



As expected, MW dust makes pixels dimmer, redder, and less overdense.

$0.6 < z < 0.9$

Bayesian analysis

Calculate the probability that the excursions in galaxy properties, \vec{d} , in a given HEALPixel, are due to a particular combination of Galactic dust extinction, A_V , and large-scale structure in several redshift bins, $\vec{\delta}$.

$$P(\vec{\delta}, A_V | \vec{d}) = L(\vec{d} | \vec{\delta}, A_V) p(\vec{\delta}) p(A_V)$$

Angular Resolution for MW Dust

- $\sim 1^\circ$ is lower resolution of SFD dust map; LSST data enable an independent dust map at this resolution.
- $\sim 6'$ is the highest resolution of SFD; in Bayesian analysis, LSST adds information at this resolution.
- $\sim 1'$ was chosen to test if we could improve the resolution of current maps; this appears difficult.

Partial Galaxy Clustering

Humna Awan (Rutgers)

Eric Gawiser (Rutgers)

Partial Galaxies?

- Traditional LSS assumes that galaxies lie at known positions in redshift space (RA, dec, z)
- Even with spectroscopy, a galaxy's radial position is uncertain
- With photo-z, some galaxies have significant probability of lying in multiple redshift bins

Partial Galaxies?

- Traditional LSS assumes that galaxies lie at known positions in redshift space (RA, dec, z)
- Even with spectroscopy, a galaxy's radial position is uncertain
- With photo-z, some galaxies have significant probability of lying in multiple redshift bins
- **Possible Solution: Estimators that incorporate uncertainty in galaxy radial positions**

Galaxy Clustering: Traditional Estimator

(2D) 2pt galaxy autocorrelation function $w(\theta)$

- Landy-Szalay estimator:

$$w_{auto}(\theta) = \frac{(D - R)(D - R)(\theta)}{RR(\theta)} = \frac{DD(\theta) - 2DR(\theta) + RR(\theta)}{RR(\theta)}$$

DD, DR, RR are histograms.

Explicitly, e.g. ,

$$DD(\theta_k) = \frac{\sum_i^N \sum_{j>i}^N \Theta(\theta_{ij} - \theta_{\min,k}) [1 - \Theta(\theta_{ij} - \theta_{\max,k})]}{\sum_i^N \sum_{j>i}^N}$$

where $\Theta(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases}$ is the Heaviside step function.

Probability-Weighted Estimator

Marked correlations: extract features in correlations.

Weigh each galaxy by e.g., its classification probability

⇒ Consider **all** galaxies, without divisions into subsamples.

⇒ Probability-weighted estimator

$$\tilde{w}_{AB}^{obs}(\theta_k) = \frac{(\widetilde{DD})_{AB}(\theta_k) - (\widetilde{DR})_A(\theta_k) - (\widetilde{DR})_B(\theta_k) + RR(\theta_k)}{RR(\theta_k)};$$

where

$$(\widetilde{DD})_{AB}(\theta_k) = \frac{\sum_i^{N_{tot}} \sum_{j \neq i}^{N_{tot}} w_i^A w_j^B \Theta(\theta_{ij} - \theta_{\min,k}) [1 - \Theta(\theta_{ij} - \theta_{\max,k})]}{\sum_i^{N_{tot}} \sum_{j \neq i}^{N_{tot}} w_i^A w_j^B},$$

$$(\widetilde{DR})_A(\theta_k) = \frac{\sum_i^{N_{tot}} \sum_j^{N_R} w_i^A \Theta(\theta_{ij} - \theta_{\min,k}) [1 - \Theta(\theta_{ij} - \theta_{\max,k})]}{\sum_i^{N_{tot}} \sum_j^{N_{R,tot}} w_i^A}$$

Conclusions

- “Big LSS” comes from innovative analysis techniques, including revisiting assumptions
- For Milky Way dust, use dust map as informative prior for simultaneous measurement of LSS and dust. LSST adds constraints on dust at $\sim 1^\circ$ scales.
(Matías Bravo et al., under DESC WG review)
- For galaxies with uncertain radial positions, partial galaxy clustering offers unbiased LSS measurements with reduced variance
(Humna Awan & EG, in prep.)