Snowmass2021 - Letter of Interest

Unified inference of the cosmological parameters from noise-dominated galaxy clustering and shear

Thematic Areas: (check all that apply ☐/■)

☐ (CF1) Dark Matter: Particle Like
☐ (CF2) Dark Matter: Wavelike
☐ (CF3) Dark Matter: Cosmic Probes
■ (CF4) Dark Energy and Cosmic Acceleration: The Modern Universe
☐ (CF5) Dark Energy and Cosmic Acceleration: Cosmic Dawn and Before
☐ (CF6) Dark Energy and Cosmic Acceleration: Complementarity of Probes and New Facilities
☐ (CF7) Cosmic Probes of Fundamental Physics
☐ (Other) *[Please specify frontier/topical group]*

Contact Information:
Submitter Name/Institution: Robert Armstrong (armstrong46@llnl.gov)

Authors: Robert Armstrong (LLNL), Alex I. Malz (German Centre for Cosmological Lensing), Michael Schneider (LLNL)

Abstract: Large optical surveys have traditionally constrained dark energy by comparing two-point statistics of galaxy catalogs calculated from independent measurements of positions, shears, and photometric redshifts to those computed analytically from potential values of cosmological parameters. While this approach has been successful for current surveys, it is lacking in a number of areas. We highlight some of the limitations of this approach. First, the measurement methods employed are chosen without sound theoretical motivation, ignoring implicit knowledge of inherent uncertainties and making it difficult to properly characterize galaxies at low signal-to-noise. Second, combining information at the catalog level can lead to subtle selection biases where some measurement algorithms will fail on a subset of galaxies while others succeed. Third, the current approach is not able to propagate uncertainties from pixels through to cosmological constraints. To maximize the scientific gain from the flood of new data from upcoming surveys, it is critical that we develop algorithms that tightly control systematic errors. We recommend support for infrastructure enabling a unified Bayesian analysis at the pixel level that overcomes the limitations of standard methodologies. Such an approach will circumvent the need to match independent catalogs by instead propagating varied sources of uncertainty through the entire analysis pipeline. By properly accounting for correlations between shear, flux, and position, this paradigm shift will increase sensitivity to cosmology by leveraging, rather than neglecting, the large fraction of low signal-to-noise galaxies traditionally omitted from cosmological analyses.

Background: Two decades after the discovery of the accelerating expansion of the universe, the HEP community still seeks a better understanding of how dark energy fits into the standard cosmological model. Many experiments point towards dark energy being a cosmological constant[1–4], but current data is not able to distinguish between alternative dark energy models or modified gravity[5;6].

The HEP community is currently preparing for the imminent onset of several large scale surveys (Rubin[7], Euclid[8], Roman[9]) aiming to provide insight into the nature of dark energy. These data sets will cover significant fractions of the sky measuring billions of galaxies to an unprecedented depth. The vastly expanded volumes of data, particularly data with a low signal-to-noise ratio (S/N), in upcoming surveys, combined with increased precision requirements, merit a new approach to survey analysis that has the potential to be ground-breaking. In anticipation of the deluge of new data, we must examine our analyses ensuring they have sufficient precision for our needs and allow a comprehensive understanding of systematic errors.

One of the most powerful techniques to constrain dark energy is to combine measurements of weak lensing with galaxy clustering. These probes are sensitive to both the expansion history of the universe and the growth of cosmological structure. Several ongoing experiments have combined these measurements to constrain cosmological parameters (DES[10], HSC[11], KiDS[12]). The typical approach in these analyses is to compress information from catalogs of galaxy position, shear, and redshift into a small number of two-point functions. As two-point functions can be readily computed analytically, they provide a fast method to constrain cosmology. Also, the correlation of different two-point functions employs inherent redundancy to asses and mitigate systematic errors.

Extent of applicability of current methods: Though the tried-and-true analysis approach has enabled the breakthroughs of previous data sets, many aspects of the standard approach merit re-examination prior to application to anticipated data sets, which will be more strongly affected by systematic errors.

- **Next-generation cosmological inference must be robust to violations of the assumption of supportive prior information assumed for calibration of derived quantities, such as photometric redshifts.** For example, model uncertainty in photometric redshift estimation may take the form of a nonrepresentative spectroscopic training set for a machine learning algorithm or an unrealistic template library for a model fitting technique, and such assumptions are not currently taken into account in ongoing analyses.
- **Advancement of the constraints on the cosmological parameters necessitates the effective utilization of noise-dominated data as a replacement for traditional cutting procedures.** Applying the cuts, e.g. S/N $\lesssim$20[13], of modern surveys, enforced to circumvent the difficulties of shear calibration and photometric redshift estimation, to the deep catalogs of the future, populated predominantly by low S/N galaxies, would negate the statistical advantages of the larger data sets and introduce systematic biases.
- **Valid constraints on the cosmological parameters will require precise quantification of unavoidable methodological systematics that will dominate over systematic errors.** While current surveys have mechanisms that attempt to alleviate selection biases resulting from inconsistencies between which galaxies are subject to measurement failures under diverse algorithms, it remains a significant problem that could be avoided by a unified approach that does not reject galaxies on the basis of intermediate measurements.
- **Characterization of the uncertainty landscape mandates a self-consistent error propagation procedure that does not wash out error properties closer to the data at each step.** To meet external requirements in the absence of mutual coverage with calibration data, we must marginalize over arbitrary uncertainties, such as the novel systematics virtually guaranteed by any new instrument, which may only be accomplished by propagating pixel-level uncertainties self-consistently throughout an analysis that leverages the natural covariances of the data.

- **Using derived data products can complicate the analysis procedure and largely ignores correlations between products** A catalog-based approach can introduce extra steps in the pipeline. For example, the metacalibration shear[14] method requires that any measurement used in galaxy selection, including object detection, be run multiple times on sheared versions of the image. The omitted correlations of galaxy properties can lead to an improved parameter inference[15;16].

To neglect these hazards would go beyond mere nullification of upcoming data sets according to a sub-optimal balance of the bias-variance trade-off; the inaccurate characterization and propagation of systematic errors that would result from application of traditional methodologies to next-generation data would render the derived constraints on the cosmological parameters meaningless.

Recommendation for Snowmass 2021: To address the shortcomings of traditional methodologies and maximize the constraining power of the data sets of the immediate future, we recommend adoption of a pixel-level framework that unifies traditionally separate measurements into a single inference pipeline. The pixel-level framework would operate directly on survey images and hierarchically infer interim measurements such as shear and redshift. The requisite analysis approach employs a Bayesian forward model to utilize interim galaxy properties to determine probabilstically, the lens potential, galaxy density and intrinsic galaxy distributions. One can then produce robust constraints by simply feeding these derived quantities into a standard cosmology pipeline, but such a Bayesian hierarchical model can also facilitate the extension of the inference to encompass the cosmological parameters, enabling a fully coherent marginalization over the systematic uncertainties that would otherwise preclude the illumination of dark energy with next-generation photometric data sets.

In addition to putting cosmological inference on a statistically rigorous foundation, we provide the following motivations for the proposed paradigm shift:

- A unified framework eliminates the need for separate measurement methods reliant on independent and possibly contradictory assumptions, thereby avoiding the subtle selection biases that arise from matching the outputs of independent analysis stages.
- The hierarchical method maximizes the statistical constraining power by using all the data gathered, rendering cuts entirely unnecessary and preventing their otherwise unavoidable information loss.
- Model uncertainty, for example, in the point spread function, that cannot be accounted for in traditional approaches is instead coherently propagated from pixels to the cosmological parameters via a forward model, not only reducing biases in, say, shear measurement, but also ensuring an accurate assessment of the errors at all stages.[17].
- The fully probabilistic approach exploits inherent per-object and population-level correlations between galaxy size, shape, and redshift to model diverse sources of uncertainty.
- A pixel-level analysis ensures a statistically sound combination of distinct data sets and instruments in a single inference, a known need identified in the context of Rubin, Euclid and WFIRST[18–20].

Though early efforts toward this goal[21–25] have been restricted to a subset of measurements, recent advances in computing now bring within reach an end-to-end pipeline. For these emerging technologies to realize a principled inference that deepens our understanding of dark energy, however, we must devote resources to developing, implementing, and validating such methods.

Inference of the cosmological parameters in the low-signal-to-noise regime necessitates a robust, self-consistent analysis framework unlike those that have long sufficed for the smaller, higher fidelity data sets of the past. A unified Bayesian approach will open the door to significant improvements over current methods that constrain cosmology, delivering optimal measurements on more galaxies while minimizing the impact of systematic errors, a prerequisite for distinguishing between dark energy mechanisms and modified gravity.

References

[1] eBOSS Collaboration, S. Alam, M. Aubert, S. Avila, C. Balland, J.E. Bautista et al., *The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Cosmological Implications from two Decades of Spectroscopic Surveys at the Apache Point observatory*, *arXiv e-prints* (2020) arXiv:2007.08991 [`2007.08991`].

[2] T.M.C. Abbott, F.B. Abdalla, A. Alarcon, S. Allam, J. Annis, S. Avila et al., *Dark Energy Survey year 1 results: Joint analysis of galaxy clustering, galaxy lensing, and CMB lensing two-point functions*, **100** (2019) 023541 [`1810.02322`].

[3] M. Betoule, R. Kessler, J. Guy, J. Mosher, D. Hardin, R. Biswas et al., *Improved cosmological constraints from a joint analysis of the SDSS-II and SNLS supernova samples*, **568** (2014) A22 [`1401.4064`].

[4] T.M.C. Abbott, M. Aguena, A. Alarcon, S. Allam, S. Allen, J. Annis et al., *Dark Energy Survey Year 1 Results: Cosmological constraints from cluster abundances and weak lensing*, **102** (2020) 023509.

[5] A. Spurio Mancini, F. Köhlinger, B. Joachimi, V. Pettorino, B.M. Schäfer, R. Reischke et al., *KiDS + GAMA: constraints on horndeski gravity from combined large-scale structure probes*, **490** (2019) 2155 [`1901.03686`].

[6] M. Ishak, T. Baker, P. Bull, E.M. Pedersen, J. Blazek, P.G. Ferreira et al., *Modified Gravity and Dark Energy models Beyond $w(z)$CDM Testable by LSST*, *arXiv e-prints* (2019) arXiv:1905.09687 [`1905.09687`].

[7] LSST Science Collaboration, P.A. Abell, J. Allison, S.F. Anderson, J.R. Andrew, J.R.P. Angel et al., *LSST Science Book, Version 2.0*, *ArXiv e-prints* (2009) [`0912.0201`].

[8] R. Laureijs, J. Amiaux, S. Arduini, J.. Auguères, J. Brinchmann, R. Cole et al., *Euclid Definition Study Report*, *ArXiv e-prints* (2011) [`1110.3193`].

[9] D. Spergel, N. Gehrels, C. Baltay, D. Bennett, J. Breckinridge, M. Donahue et al., *Wide-Field InfrarRed Survey Telescope-Astrophysics Focused Telescope Assets WFIRST-AFTA 2015 Report*, *ArXiv e-prints* (2015) [`1503.03757`].

[10] T.M.C. Abbott, F.B. Abdalla, A. Alarcon, J. Aleksić, S. Allam, S. Allen et al., *Dark Energy Survey year 1 results: Cosmological constraints from galaxy clustering and weak lensing*, **98** (2018) 043526 [`1708.01530`].

[11] C. Hikage, M. Oguri, T. Hamana, S. More, R. Mandelbaum, M. Takada et al., *Cosmology from cosmic shear power spectra with Subaru Hyper Suprime-Cam first-year data*, **71** (2019) 43 [`1809.09148`].

[12] C. Heymans, T. Tröster, M. Asgari, C. Blake, H. Hildebrandt, B. Joachimi et al., *KiDS-1000 Cosmology: Multi-probe weak gravitational lensing and spectroscopic galaxy clustering constraints*, *arXiv e-prints* (2020) arXiv:2007.15632 [`2007.15632`].

[13] R. Mandelbaum, H. Miyatake, T. Hamana, M. Oguri, M. Simet, R. Armstrong et al., *The first-year shear catalog of the Subaru Hyper Suprime-Cam Subaru Strategic Program Survey*, **70** (2018) S25 [`1705.06745`].

[14] E.S. Sheldon, M.R. Becker, N. MacCrann and M. Jarvis, *Mitigating Shear-dependent Object Detection Biases with Metacalibration*, *arXiv e-prints* (2019) arXiv:1911.02505 [`1911.02505`].

[15] E.M. Huff and G.J. Graves, *Magnificent Magnification: Exploiting the Other Half of the Lensing Signal*, *arXiv e-prints* (2011) arXiv:1111.1070 [1111.1070].

[16] S.-M. Niemi, T.D. Kitching and M. Cropper, *On weak lensing shape noise*, **454** (2015) 1221 [1509.05058].

[17] M.D. Schneider and J.E. Meyers, *Probabilistic point spread function modeling*, Tech. Rep. LSST Dark Energy Science Collaboration: Report from the PSF Task Force.

[18] W. Dawson and M. Schneider, *Complementarity of lsst and wfirst: Regarding object blending*, Tech. Rep. Lawrence Livermore National Lab.(LLNL), Livermore, CA (United States) (2014).

[19] B. Jain, D. Spergel, R. Bean, A. Connolly, I. Dell'antonio, J. Frieman et al., *The Whole is Greater than the Sum of the Parts: Optimizing the Joint Science Return from LSST, Euclid and WFIRST*, *arXiv e-prints* (2015) arXiv:1501.07897 [1501.07897].

[20] R. Chary, G. Helou, G. Brammer, P. Capak, A. Faisst, D. Flynn et al., *Joint Survey Processing of Euclid, Rubin and Roman: Final Report*, *arXiv e-prints* (2020) arXiv:2008.10663 [2008.10663].

[21] M.D. Schneider, D.W. Hogg, P.J. Marshall, W.A. Dawson, J. Meyers, D.J. Bard et al., *Hierarchical probabilistic inference of cosmic shear*, *The Astrophysical Journal* **807** (2015) 87.

[22] G.M. Bernstein and R. Armstrong, *Bayesian lensing shear measurement*, **438** (2014) 1880 [1304.1843].

[23] J. Alsing, A. Heavens, A.H. Jaffe, A. Kiessling, B. Wandelt and T. Hoffmann, *Hierarchical cosmic shear power spectrum inference*, **455** (2016) 4452 [1505.07840].

[24] M.D. Schneider, K.Y. Ng, W.A. Dawson, P.J. Marshall, J.E. Meyers and D.J. Bard, *Probabilistic Cosmological Mass Mapping from Weak Lensing Shear*, **839** (2017) 25 [1610.06673].

[25] C. Sánchez and G.M. Bernstein, *Redshift inference from the combination of galaxy colours and clustering in a hierarchical Bayesian model*, **483** (2019) 2801 [1807.11873].