Request for Information Related to High Energy Physics and Space-Based Astrophysics

Combining information from multiple cosmological surveys: inference and modeling challenges

Focus Areas:

☐ Focus Area 1: Lunar farside
☐ Focus Area 2: International Space Station
■ Focus Area 3: Cosmology from Optical Surveys

Contact Information:
Erminia Calabrese (Cardiff University) [calabresee@cardiff.ac.uk]
Tim Eifler (University of Arizona) [timeifler@arizona.edu]

Authors (alphabetical):
David Alonso (Oxford University), Erminia Calabrese (Cardiff University), Tim Eifler (University of Arizona), Giulio Fabbian (Cardiff University/CCA), Simone Ferraro (Lawrence Berkeley National Laboratory), Eric Gawiser (Rutgers), J. Colin Hill (Columbia University/Flatiron Institute), Elisabeth Krause (University of Arizona), Mathew Madhavacheril (Perimeter Institute), Anže Slosar (Brookhaven National Laboratory), David N. Spergel (CCA)

Endorsers (Group): Roman Space Telescope Science Investigation Team “Cosmology with the High Latitude Survey”

Endorsers (Individuals): Maximilian H. Abitbol (University of Oxford), Marcelo Alvarez (Lawrence Berkeley National Lab), Jonathan Aumont (IRAP), Susanna Azzoni (University of Oxford / Kavli IPMU), Carlo Baccigalupi (SISSA), Mario Ballardini (University of Bologna), Nicola Bartolo (University of Padova), Rito Ban Basu Thakur (California Institute of Technology), Nicholas Battaglia (Cornell University), Rachel Bean (Cornell University), Karim Benabed (Institut d’astrophysique de Paris - Sorbonne Université/CNRS), Federico Bianchini (Stanford/SLAC), Julian Borrill (LBNL & UC Berkeley), Thejs Brinckmann (Stony Brook University), Anthony Challinor (University of Cambridge), Susa E. Clark (Institute for Advanced Study), Anantha Cooray (UC Irvine), Gabriele Coppi (Department of Physics, University of Milano-Bicocca), Jacques Delabrouille (CNRS, laboratoire APC), Eleonora Di Valentino (Durham University), Bradley Dober (NIST / CU Boulder), Olivier Doré (JPL/Caltech), Cyrille Doux (University of Pennsylvania), Adriaan Duivenvoorden (Princeton University, Department of Physics), Jo Dunkley (Princeton University), Cora Dvorkin (Harvard University), Josquin Errard (APC / CNRS), Katherine Freese (University of Texas, Austin), Nicholas Galitzki (University of California San Diego), Kenneth Ganga (APC/CNRS/France), Ricardo Génova-Santos (Instituto de Astrofísica de Canarias), Martina Gerbino (INFN Ferrara), Vera Gluscevic (University of Southern California), Jon E. Gudmundsson (Stockholm University), Salman Habib (Argonne National Laboratory), Ian Harrison (The University of Manchester), Katrin Heitmann (Argonne National Laboratory), Brandon Hensley (Princeton University), Christopher M. Hirata (The Ohio State University), Renée Hlozek (University of Toronto), Eric Huff (JPL/Caltech), John P. Hughes (Rutgers University), Andrew Jaffe (Imperial College London), Alejandro Jiménez Muñoz (LPSC, CNRS, Grenoble, France), Arthur Kosowsky (University of Pittsburgh, Massimiliano
Abstract:
The tightest and most robust cosmological results of the next decade will be achieved by bringing together multiple surveys of the Universe. This endeavor has to happen across multiple layers of the data processing and analysis, e.g., enhancements are expected from combining Euclid, Rubin, and Roman (as well as other surveys) not only at the level of joint processing and catalog combination, but also during the post-catalog parts of the analysis such as the cosmological inference process. While every experiment builds their own analysis and inference framework and creates their own set of simulations, cross-survey work that homogenizes these efforts, exchanges information from numerical simulations, and coordinates details in the modeling of astrophysical and observational systematics of the corresponding datasets is crucial.
A decade of optical and millimeter-wave surveys of the Universe to be combined

Our current best understanding of the origin, composition, and evolution of the Universe is drawn from observations of the sky at different wavelengths collected by different kinds of experiments. For the last few billion years of cosmic time, as the large-scale structure (LSS) of the Universe becomes more established, information encoded in the spatial distribution, clustering, and shape distortion of galaxies is measured with optical imaging and spectroscopic redshift surveys. The very early phases are instead anchored with high-precision, high-resolution surveys of the Cosmic Microwave Background (CMB). As the CMB photons travel through the emerging and evolving LSS, multiple physical processes such as gravitational lensing, thermal and kinetic Sunyaev-Zel’dovich effect correlate the two. All these observations probe the same underlying matter and energy distribution of the Universe and combining them provides several advantages: (i) multiple, complementary routes to elucidate the cosmological model, e.g., to establish if the standard \( \Lambda \)CDM model is indeed correct or whether there are other subtle physical effects from additional unknown physics, (ii) increased statistical power, (iii) help in isolating physical effects that are otherwise degenerate, and (iv) minimization of the impact of survey-specific instrumental and astrophysical systematics.

For past and current surveys this combination has been a trivial exercise, with simple addition of probes at the likelihood level when exploring models of the Universe and constraining cosmological parameters – a good enough approximation given the level of noise and overlap in the observed sky. The increased sensitivity and the wider scientific reach of future surveys will make this work much more challenging. Many new surveys are on the horizon for this decade. Each of them will work to deliver their specific science program, but combined analyses will be the pillar for the tightest and most robust cosmological results, and therefore for potential new discoveries. In practice, this will require developing new infrastructure for combining multiple surveys and consistency across simulations.

Multi-Survey Cosmological Analyses

Future DOE experiments and NASA/ESA space missions such as the Vera C. Rubin Observatory’s Legacy Survey of Space and Time (LSST)\(^1\), the Euclid satellite mission\(^2\), and the Nancy Grace Roman Space Telescope\(^3\) aim to unveil the physical mechanism causing the accelerated expansion of the Universe (“The Cosmic Frontier”). Possible theoretical explanations require fundamentally new physics, elevating dark energy research into one of the most exciting problems in science today. Analyzing these LSS datasets and fully exploiting their synergies requires a concerted effort across many areas of science, most prominently cosmology, astrophysics, computer science, statistics, and data analysis techniques. Expected enhancement and synergies from this combination have been highlighted in several reports [1–6].

In addition to the three exciting datasets from Rubin, Roman, and Euclid, the community will have access to data from a variety of next-generation CMB experiments, in particular the ground-based Simons Observatory (SO)\(^4\) and the DOE-led CMB-S4\(^5\) endeavors which will provide new maps of the microwave sky with significant overlap in survey footprint with Rubin, Roman and Euclid. These CMB datasets are highly complementary to the LSS information, both in terms of increasing the constraining power on cosmology and in terms of mitigating critical systematics of LSS analyses [7–14]. LSS will also provide tracers to delens potential future CMB missions with NASA contribution such as LiteBIRD [15].

\(^1\)https://www.lsst.org/
\(^2\)https://sci.esa.int/web/euclid/
\(^3\)https://roman.gsfc.nasa.gov/
\(^4\)https://simonsobservatory.org/
\(^5\)https://cmb-s4.org/
While synergies between Euclid, Rubin, and Roman at the imaging and catalog levels are an important focus area that requires funding support, this document focuses on the post-catalog parts of the analysis, specifically the details of the cosmological inference process. Currently, every experiment (Euclid, Rubin, Roman, SO, CMB-S4) builds their own analysis and inference framework and creates their own set of simulations. Each experiment also has a specific working group or team exploring CMB and LSS correlations. We propose a cross-survey work effort that homogenizes these efforts, exchanges information from numerical simulations, and coordinates details in the modeling of astrophysical and observational systematics of the corresponding datasets. Ultimately, this effort will help the community build and use a joint analysis framework that can robustly combine LSS and CMB information.

Such a framework for combined cosmological analyses should include: (i) Plans for joint simulations: building sky models that use the same underlying assumptions; (ii) Software infrastructure to bridge and/or merge likelihood codes from different experiments: creating a platform that standardizes code comparison between different analysis codes and identifying the nodes where connection between codes is needed; (iii) Software to compute consistent theoretical predictions for all observational probes: identifying key approximations, fiducial values and priors; (iv) Prescriptions for building joint terms, e.g., covariances between probes.

This RFI was specific to Rubin, Euclid, and Roman, but the range of datasets that are important to consider extends beyond these three. We specifically mention CMB experiments in this document, however data from the Dark Energy Spectroscopic Instrument, the NASA SPHEREx explorer mission, radio surveys, and other experiments will exist as well on the relevant timescale. All these missions face similar challenges as described here (analysis frameworks and numerical simulations). The advent of so many, diverse and complementary datasets in the near future poses a fantastic opportunity for the community, however in order to fully explore the joint constraining power of these datasets a coordinated effort is necessary.

Response to Specific Questions

Science Enhancements

a. What are the key dark energy science areas that will be enhanced by these activities?

Dark Energy as a term describes our lack of understanding of the physical concepts that underlie cosmic acceleration. As such it encompasses a wide variety of fundamental physics topics including modified gravity, neutrino physics, dark matter-dark energy coupling, early dark energy, and many more. A joint analysis is required to control the systematics budget and to increase the constraining power such that the community can discriminate between the different physical concepts that explain cosmic acceleration. Below we list the main focus areas that need to be addressed in order to fully extract the cosmological information from joint survey analyses:

- Observational uncertainties: For example, photo-z errors, shear calibration, depth variations need to be parameterized consistently across the different surveys if the datasets are combined.

- Astrophysical uncertainties: For example, nonlinear modeling of the density field, baryonic physics, intrinsic alignment, galaxy bias and Halo Occupation Distribution models are key astrophysical uncertainties. Better coordination on how to model these effects and how to mitigate their uncertainties is required.

- Statistical uncertainties: For example, the functional form of the likelihood and, if a multivariate Gaussian is assumed, the computation of data covariances are key uncertainties in a joint CMB-LSS analysis.
Simulated analyses - quantifying the science return as a function of analysis choices. Analysis frameworks can inform simulation campaigns, survey strategy, and ultimately the analyses about the error budget as a function of the analysis choices (scales, redshifts, galaxy samples, summary statistics).

Numerical simulations: Nonlinear modeling of the density field and exploring the statistical uncertainties mentioned above requires numerical simulations. The initial conditions of these simulations should be coordinated across all survey collaborations to enable a better comparison.

Hydrodynamic simulations: Baryonic physics, intrinsic alignment, galaxy bias and Halo Occupation Distribution models require a hydrodynamic simulation campaign that is computationally expensive. In order to utilize the available computing resources most effectively this simulation campaign must be informed by the composition of the error budget of a joint analysis. In other words, the requirements for a simulation campaign will be different when analyzing data from a single survey as opposed to data from multiple surveys. Simulated cosmological likelihood analyses of multi-survey data can identify the main contributors to the overall error budget and can inform a corresponding simulation campaign. A close connection between the simulated analyses and the simulation effort is required.

See also the report from the simulation task force presented in Ref. [16].

What level of scientific enhancement is expected by carrying them out after the datasets are public?

Coordinating the combination of multiple surveys before the datasets become public will guarantee that the release of the data will be accompanied with the tightest and most robust cosmological results. Starting this work after the datasets become public will not only slow down the path to this outcome, but will also encounter additional complications due to potential conflicts between individual analysis software tools and inconsistency between simulations. All information collected through this cross-survey effort should be made public immediately and code development for the analysis frameworks should be open source and in the public domain. The code and information can be used by anyone to analyze the joint catalogs.

What additional enhancements are expected if plans are put in place in the near term to enable joint data processing and analysis of public data sets?

The post-catalog efforts described here will use the joint catalog that could be created from Rubin, Euclid, and Roman data. Additionally, thanks to its open source nature this framework can be immediately applied to precursor data from current LSS and CMB experiments. On the simulation front, this infrastructure could provide value-added products for the cosmological community such as (i) ray-tracing for lensing observables, (ii) painting of CMB observables like the thermal and kinetic Sunyaev Zeldovich effect, (iii) Halo Occupation Distribution models for the cosmic infrared background and radio galaxies at millimeter wavelengths, and optical and infrared galaxies that mimic the populations observed by various surveys, and (iv) LSS catalogs (e.g., red sequence galaxy and galaxy cluster catalogs).

b. What is the scope of work required, as well as the opportunities and costs?

We suggest three stages:

1. **Stage 1**: Forming an exploratory team from all interested surveys that includes funded FTE and postdocs. The main task of this team is to build an analysis framework interfacing analysis codes of different surveys and quantify the precision needed in suitable simulation campaigns.

2. **Stage 2**: Informing and engaging the leading simulation experts in the effort defined in Stage 1. The goal here is to synergize existing efforts in the individual collaborations and to help design a simulation campaign that addresses the main contributors to the error budget of a joint analysis. The cost for such a simulation campaign will be defined in Stage 1.
3. **Stage 3:** Include the information obtained through the simulation campaign in the analysis frameworks and help the community to analyze public datasets. We of course acknowledge that any joint analysis of proprietary datasets requires MOUs between the collaborations/projects.

c. *What are key obstacles, impediments, or bottlenecks to advancing development of these plans?*

- Person power and time commitment.
- Computing time for the simulated analyses and the simulation campaigns. It is important to point out that coordinating the individual simulation efforts within each collaboration will utilize synergies, which will make the individual computing efforts more efficient. However, a joint analysis likely has more stringent requirements (size, resolution, parameter space, systematics modeling) on a simulation campaign, which will add to the computing time that has been allocated for the individual analysis efforts.
- Infrastructure to build and develop complex software and to collaborate across the different survey efforts. Infrastructure to host simulations or post-processed simulation results and make them accessible to the community.

d. *Are there other science topics besides dark energy that drive the requirements for joint data processing or analysis?*

Combining multiple surveys will benefit exploration of all physical processes that affect the distribution of matter and energy in the Universe, such as modified theories of gravity and the presence of massive neutrinos. For this latter case, a significant detection of the neutrino mass is only possible in a combined survey analysis (see, e.g., [17, 18]).

**Collaboration and Partnerships**

k. *What cooperation or partnerships between DOE and NASA could further the scientific and technology advances?*

Collaboration between the DOE experiment collaborations and the science teams of the NASA missions is critically important as well as the connection to the numerical simulation experts. We also stress the importance of international members of the science collaborations, who should be included in such partnerships wherever possible.

l. *What mix of institutions or collaboration models could best carry out the envisioned research and/or development?*

The survey collaborations need to identify and interface the relevant experts for Stage 1. Stage 2 requires High Performance Computing resources, probably at DOE leadership computing facilities and NASA Supercomputers. It also requires highly qualified personnel that work at the interface of cosmological data analysis and HPC software development. Stage 3 again requires close interaction of the numerical simulation experts and the analysis framework experts. At all stages a close contact to the upstream catalog generation efforts is required such that the error budget is continuously updated.

m. *What resources, capabilities and infrastructure at DOE National Laboratories or the NASA Centers (including the Jet Propulsion Laboratory (JPL)) would be beneficial for and could accelerate or facilitate research in this topic?*

Access to high-performance computing resources, e.g., at DOE facilities (NERSC, ACLF, OCLF) or NASA facilities (Pleiades), including trained personnel that works at the interface of cosmological data analysis and numerical simulations on HPC systems, is the key. Training of said personnel will require time and and funding.
Are there other factors, not addressed by the questions above, which should be considered in planning HEP and APD activities in this subject area?

The RFI was specific to Rubin, Euclid, and Roman, but the range of datasets that are important to consider is much larger. We specifically mention CMB experiments in this document, however data from the Dark Energy Spectroscopic Instrument, the NASA SPHEREx explorer mission, radio surveys, and other experiments will exist as well on the relevant timescale. All these missions will face similar problems as described here (analysis frameworks and numerical simulations) and how to best coordinate and synergize the individual efforts should be discussed as part of future HEP and APD activities.

References

[1] B. Jain, D. Spergel, R. Bean, A. Connolly, I. Dell’antonio, J. Frieman, E. Gawiser, N. Gehrels, L. Gladney, K. Heitmann, G. Helou, C. Hirata, S. Ho, Ž. Ivezic, M. Jarvis, S. Kahn, J. Kalirai, A. Kim, R. Lupton, R. Mandelbaum, P. Marshall, J. A. Newman, S. Perlmuter, M. Postman, J. Rhodes, M. A. Strauss, J. A. Tyson, L. Walkowicz, and W. M. Wood-Vasey, “The Whole is Greater than the Sum of the Parts: Optimizing the Joint Science Return from LSST, Euclid and WFIRST,” arXiv e-prints (Jan., 2015) arXiv:1501.07897, arXiv:1501.07897 [astro-ph.IM].

[2] J. Rhodes, R. C. Nichol, É. Aubourg, R. Bean, D. Boutigny, M. N. Bremer, P. Capak, V. Cardone, B. Carry, C. J. Conselice, A. J. Connolly, J.-C. Cuillandre, N. A. Hatch, G. Helou, S. Hemmati, H. Hildebrandt, R. Hložek, L. Jones, S. Kahn, A. Kiessling, T. Kitching, R. Lupton, R. Mandelbaum, K. Markovic, P. Marshall, R. Massey, B. J. Maughan, P. Melchior, Y. Mellier, J. A. Newman, B. Robertson, M. Sauvage, T. Schrabback, G. P. Smith, M. A. Strauss, A. Taylor, and A. Von Der Linden, “Scientific Synergy between LSST and Euclid,” ApJS 233 no. 2, (Dec., 2017) 21, arXiv:1710.08489 [astro-ph.IM].

[3] K. Dawson, J. Frieman, K. Heitmann, B. Jain, S. Kahn, R. Mandelbaum, S. Perlmutter, and A. Slosar, “Cosmic Visions Dark Energy: Small Projects Portfolio,” arXiv e-prints (Feb., 2018) arXiv:1802.07216, arXiv:1802.07216 [astro-ph.CO].

[4] P. Capak, J.-C. Cuillandre, F. Bernardeau, F. Castander, R. Bowler, C. Chang, C. Grillmair, P. Gris, T. Eifler, C. Hirata, I. Hook, B. Jain, K. Kuijken, M. Lochner, P. Oesch, S. Palanit, J. Rhodes, B. Robertson, D. Rubin, R. Scaramella, C. Scarlata, D. Scolnic, J. Silverman, S. Wachter, Y. Wang, and The Tri-Agency Working Group, “Enhancing LSST Science with Euclid Synergy,” arXiv e-prints (Apr., 2019) arXiv:1904.04349, arXiv:1904.04349 [astro-ph.IM].

[5] T. Eifler, M. Simet, E. Krause, C. Hirata, H.-J. Huang, X. Fang, V. Miranda, R. Mandelbaum, C. Doux, C. Heinrich, E. Huff, H. Miyatake, S. Hemmati, J. Xu, P. Rogozenski, P. Capak, A. Choi, O. Dore, B. Jain, M. Jarvis, N. MacCrann, D. Masters, E. Rozo, D. N. Spergel, M. Troxel, A. von der Linden, Y. Wang, D. H. Weinberg, L. Wenzl, and H.-Y. Wu, “Cosmology with the Wide-Field Infrared Survey Telescope – Synergies with the Rubin Observatory Legacy Survey of Space and Time,” arXiv e-prints (Apr., 2020) arXiv:2004.04702, arXiv:2004.04702 [astro-ph.CO].

[6] R. Chary, G. Helou, G. Brammer, P. Capak, A. Faisst, D. Flynn, S. Groom, H. C. Ferguson, C. Grillmair, S. Hemmati, A. Koekemoer, B. Lee, S. Malhotra, H. Miyatake, P. Melchior, I. Momcheva, J. Newman, J. Masiero, R. Paladini, A. Prakash, B. Rusholme, N. R. Stickley, A. Smith, W. M. Wood-Vasey, and H. I. Teplitz, “Joint Survey Processing of Euclid, Rubin and Roman: Final Report,” arXiv e-prints (Aug., 2020) arXiv:2008.10663, arXiv:2008.10663 [astro-ph.IM].
[7] A. Vallinotto, “Using Cosmic Microwave Background Lensing to Constrain the Multiplicative Bias of Cosmic Shear,” ApJ 759 no. 1, (Nov., 2012) 32, arXiv:1110.5339 [astro-ph.CO].

[8] S. Das, J. Errard, and D. Spergel, “Can CMB Lensing Help Cosmic Shear Surveys?,” arXiv e-prints (Nov., 2013) arXiv:1311.2338, arXiv:1311.2338 [astro-ph.CO].

[9] E. Schaan, E. Krause, T. Eifler, O. Doré, H. Miyatake, J. Rhodes, and D. N. Spergel, “Looking through the same lens: Shear calibration for LSST, Euclid, and WFIRST with stage 4 CMB lensing,” Phys. Rev. D 95 no. 12, (June, 2017) 123512, arXiv:1607.01761 [astro-ph.CO].

[10] R. Cawthon, “Effects of redshift uncertainty on cross-correlations of CMB lensing and galaxy surveys,” Phys. Rev. D 101 no. 6, (Mar., 2020) 063509, arXiv:1809.09251 [astro-ph.CO].

[11] E. Schaan, S. Ferraro, and U. Seljak, “Photo-z outlier self-calibration in weak lensing surveys,” J. Cosmology Astropart. Phys. 2020 no. 12, (Dec., 2020) 001, arXiv:2007.12795 [astro-ph.CO].

[12] R. Pearson and O. Zahn, “Cosmology from cross correlation of CMB lensing and galaxy surveys,” Phys. Rev. D 89 no. 4, (Feb., 2014) 043516, arXiv:1311.0905 [astro-ph.CO].

[13] P. M. Merkel and B. M. Schäfer, “Parameter constraints from weak-lensing tomography of galaxy shapes and cosmic microwave background fluctuations,” MNRAS 469 no. 3, (Aug., 2017) 2760–2770, arXiv:1707.08153 [astro-ph.CO].

[14] S.-F. Chen, H. Lee, and C. Dvorkin, “Precise and Accurate Cosmology with CMBxLSS Power Spectra and Bispectra,” arXiv e-prints (Mar., 2021) arXiv:2103.01229, arXiv:2103.01229 [astro-ph.CO].

[15] A. Manzotti, “Future cosmic microwave background delensing with galaxy surveys,” Phys. Rev. D 97 no. 4, (Feb., 2018) 043527, arXiv:1710.11038 [astro-ph.CO].

[16] N. Battaglia, A. Benson, T. Eifler, A. Hearin, K. Heitmann, S. Ho, A. Kiessling, Z. Lukic, M. Schneider, E. Sellentin, and J. Stadel, “Report from the Tri-Agency Cosmological Simulation Task Force,” arXiv e-prints (May, 2020) arXiv:2005.07281, arXiv:2005.07281 [astro-ph.CO].

[17] CMB-S4 Collaboration, “CMB-S4 Science Book, First Edition,” arXiv e-prints (Oct., 2016) arXiv:1610.02743, arXiv:1610.02743 [astro-ph.CO].

[18] Simons Observatory Collaboration, “The Simons Observatory: science goals and forecasts,” J. Cosmology Astropart. Phys. 2019 no. 2, (Feb., 2019) 056, arXiv:1808.07445 [astro-ph.CO].