# Snowmass2021 - Letter of Interest

# *CosmoSIS for the Next Decade*

**Thematic Areas:** (check all that apply □/■)
■ (CompF1): Experimental Algorithm Parallelization
■ (CompF5): End user analysis
■ (CompF7): Reinterpretation and long-term preservation of data and code
■ (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

**Contact Information:**
Submitter Name/Institution: Joe Zuntz / University of Edinburgh
Collaboration (optional): CosmoSIS Development Team
Contact Email: joe.zuntz@ed.ac.uk

**Abstract:** CosmoSIS is a computational framework to organize and implement forward-modeling cosmological analysis that compares theoretical predictions to observational data. With its extreme modularity and compatibility with different sampling methods and programming languages, CosmoSIS can play a critical role in the coming decade of ongoing and emerging cosmological analyses. These features can especially facilitate collaborations in the cosmology community and enable cross-analyses of data from different cosmic experiments. We further itemize improvement plans that can help addressing future computing challenges in cosmology in the coming decade.

**Authors:** (names and institutions in alphabetical order)
Camille Avestruz (University of Michigan)
Scott Dodelson (Carnegie Mellon University)
Marc Paterno (Fermilab)
Yuanyuan Zhang (Fermilab)
Joe Zuntz (University of Edinburgh)

## Introduction to CosmoSIS

The analysis of cosmological data is generally performed using a Bayesian approach: forward-modelling probabilistic theoretical predictions and obtaining likelihoods of observed data with respect to them, and then exploring parameter spaces with Monte Carlo methods.

CosmoSIS[1] is a computational framework to organize and implement this analysis. It defines a modular structure, in which computational libraries and tools (in C, C++, Fortran, Python, or Julia) are used to perform individual pieces of a wider predictive pipeline, and connects the pieces together to contain all results in a single *data block*. The CosmoSIS *standard cosmology library* collects the most important cosmological libraries, and puts them into this format. The framework then connects these pipelines to parameter space exploration tools (e.g. optimizers, MCMC samplers, and machine learning tools) to compute posterior constraints, best fits, or other statistics.

This approach solves the major problems in the building of cosmology analysis pipelines. It offers flexibility, allowing users to draw on a broad range of software; users can mix, match, and replace pieces easily. It offers uniformity, putting diverse libraries in the same network and all calculated results for a pipeline in the same data structure. It improves debugging, by isolating pieces of an analysis for independent verification. Finally, the multi-language design allows slow pieces of code to be written in a compiled language and fast ones to be written in an interpreted one.

The CosmoSIS framework is widely used in Cosmology, particularly in weak gravitational lensing, where it has powered the analysis of two of the three large stage III (the current generation) of photometric surveys, the *Dark Energy Survey*[2] and *Kilo-Degree Survey*[3].

## Unique Features of CosmoSIS

The critical unique feature of CosmoSIS, when compared to alternative MCMC parameter estimation frameworks, is its extreme *modularlity*. Each calculation module in CosmoSIS can be replaced by an alternative one, simply by changing a configuration file option. The common data interface makes this change seamless. For example, a user can choose to switch to a modified gravity model, just by changing a single path and adding any parameters the new model requires. Modularity is the key to allowing CosmoSIS to leverage the advances in understanding of correlated systematics, of astrophysical models, and of variant cosmologies made by a diverse group of scientists. This modularity makes CosmoSIS especially powerful in the challenging development phase of cosmological modelling, when building, testing, and comparing models.

CosmoSIS also provides an interface to a large suite of different algorithms for exploring parameter spaces, *samplers*. It is straightforward to add a new sampler to the framework, and this has enabled easy tests of new algorithms and approaches.

The system is also highly scriptable: any piece of the code can be imported by a user and used largely independently, for example to run a single calculation with varying inputs and compare the results, or to run the same MCMC process multiple times to check for sampling variance.

CosmoSIS is also compatible with a variety of programming languages – different modules in CosmoSIS can be programmed in different languages including Fortran, C, C++, Julia, and Python, allowing scientists of different scientific backgrounds, or with different programming habits, to work together.

## Future Improvements

We believe that CosmoSIS can play a significant role in the coming decade with ongoing and emerging cosmological analysis. With the modularity and compatibility, CosmoSIS is a platform that allows code re-running from the broad cosmology community, and for facilitating cross analysis of data from multiple scientific collaborations.

We anticipate major improvements in CosmoSIS that can greatly enhance its versatility and compatibility. Those include:

- *Likelihood-free inference:* LFI[4–7] enables implementation of machine learning methods in cosmological analysis. With increasing interests in the usage of machine learning methods, we expect those methods to raise the needs for different set-ups or parallelization methods, such as those using GPUs.
- *Quality Assurance:* Unit Tests, physics validation and documentation. Currently the CosmoSIS framework does not require unit tests in the modules contributed by the community. In the future, quality assurance in those modules can be achieved through setting up requirements of unit tests, code reviews, physics validation code, and documentation. These quality assurance measures can be enforced in the version of software that is centrally installed and maintained by scientific experiments. A framework for automated testing of modules, comparing a known input data block to a known output, would also be extremely valuable with continuous integration of developed code.
- *User interface:* While very useful for experts, CosmoSIS can be daunting for new users. A simple browser-based user interface, allowing users to choose and configure existing modules, would greatly ease adoption.
- *Automatic gradients:* Some of the most important algorithms in parameter estimation, notably Hamiltonian Monte Carlo[8], require the derivatives of likelihoods or theory predictions. While this can currently be done using finite differences, the results are generally unreliable without fine-tuning. The modular structure of CosmoSIS lends itself to efficient automatic differentiation, applied to each module in turn.
- *GPU interfaces:* The most important advances in computational technology are taking place on GPUs rather than CPUs, including at HPC scale. To make full use of GPU facilities, for example via the Jax-Cosmo project, CosmoSIS could be upgraded to use GPU calculations for numerically intensive modules.
- *Distribution:* Currently CosmoSIS has been centrally installed and maintained by each of the scientific experiments (e.g., DES) that employs the software. We expect the central installation and maintenance of CosmoSIS in a computing hub that is accessible to multiple scientific experiments will greatly enhance the cross-analyses of data between them.

## References

[1] J. Zuntz, M. Paterno, E. Jennings, D. Rudd, A. Manzotti, S. Dodelson, S. Bridle, S. Sehrish, and J. Kowalkowski. CosmoSIS: Modular cosmological parameter estimation. *Astronomy and Computing*, 12:45–59, September 2015.

[2] T. M. C. Abbott, F. B. Abdalla, A. Alarcon, J. Aleksić, S. Allam, S. Allen, A. Amara, J. Annis, J. Asorey, S. Avila, D. Bacon, E. Balbinot, M. Banerji, N. Banik, W. Barkhouse, M. Baumer, E. Baxter, K. Bechtol, M. R. Becker, A. Benoit-Lévy, B. A. Benson, G. M. Bernstein, E. Bertin, J. Blazek, S. L. Bridle, D. Brooks, D. Brout, E. Buckley-Geer, D. L. Burke, M. T. Busha, A. Campos, D. Capozzi, A. Carnero Rosell, M. Carrasco Kind, J. Carretero, F. J. Castander, R. Cawthon, C. Chang, N. Chen, M. Childress, A. Choi, C. Conselice, R. Crittenden, M. Crocce, C. E. Cunha, C. B. D'Andrea, L. N. da Costa, R. Das, T. M. Davis, C. Davis, J. De Vicente, D. L. DePoy, J. DeRose, S. Desai, H. T. Diehl, J. P. Dietrich,

S. Dodelson, P. Doel, A. Drlica-Wagner, T. F. Eifler, A. E. Elliott, F. Elsner, J. Elvin-Poole, J. Estrada, A. E. Evrard, Y. Fang, E. Fernandez, A. Ferté, D. A. Finley, B. Flaugher, P. Fosalba, O. Friedrich, J. Frieman, J. García-Bellido, M. Garcia-Fernandez, M. Gatti, E. Gaztanaga, D. W. Gerdes, T. Giannantonio, M. S. S. Gill, K. Glazebrook, D. A. Goldstein, D. Gruen, R. A. Gruendl, J. Gschwend, G. Gutierrez, S. Hamilton, W. G. Hartley, S. R. Hinton, K. Honscheid, B. Hoyle, D. Huterer, B. Jain, D. J. James, M. Jarvis, T. Jeltema, M. D. Johnson, M. W. G. Johnson, T. Kacprzak, S. Kent, A. G. Kim, A. King, D. Kirk, N. Kokron, A. Kovacs, E. Krause, C. Krawiec, A. Kremin, K. Kuehn, S. Kuhlmann, N. Kuropatkin, F. Lacasa, O. Lahav, T. S. Li, A. R. Liddle, C. Lidman, M. Lima, H. Lin, N. MacCrann, M. A. G. Maia, M. Makler, M. Manera, M. March, J. L. Marshall, P. Martini, R. G. McMahon, P. Melchior, F. Menanteau, R. Miquel, V. Miranda, D. Mudd, J. Muir, A. Möller, E. Neilsen, R. C. Nichol, B. Nord, P. Nugent, R. L. C. Ogando, A. Palmese, J. Peacock, H. V. Peiris, J. Peoples, W. J. Percival, D. Petravick, A. A. Plazas, A. Porredon, J. Prat, A. Pujol, M. M. Rau, A. Refregier, P. M. Ricker, N. Roe, R. P. Rollins, A. K. Romer, A. Roodman, R. Rosenfeld, A. J. Ross, E. Rozo, E. S. Rykoff, M. Sako, A. I. Salvador, S. Samuroff, C. Sánchez, E. Sanchez, B. Santiago, V. Scarpine, R. Schindler, D. Scolnic, L. F. Secco, S. Serrano, I. Sevilla-Noarbe, E. Sheldon, R. C. Smith, M. Smith, J. Smith, M. Soares-Santos, F. Sobreira, E. Suchyta, G. Tarle, D. Thomas, M. A. Troxel, D. L. Tucker, B. E. Tucker, S. A. Uddin, T. N. Varga, P. Vielzeuf, V. Vikram, A. K. Vivas, A. R. Walker, M. Wang, R. H. Wechsler, J. Weller, W. Wester, R. C. Wolf, B. Yanny, F. Yuan, A. Zenteno, B. Zhang, Y. Zhang, J. Zuntz, and Dark Energy Survey Collaboration. Dark Energy Survey year 1 results: Cosmological constraints from galaxy clustering and weak lensing. *PRD*, 98(4):043526, August 2018.

[3] B. Joachimi, C. A. Lin, M. Asgari, T. Tröster, C. Heymans, H. Hildebrandt, F. Köhlinger, A. G. Sánchez, A. H. Wright, M. Bilicki, C. Blake, J. L. van den Busch, M. Crocce, A. Dvornik, T. Erben, F. Getman, B. Giblin, H. Hoekstra, A. Kannawadi, K. Kuijken, N. R. Napolitano, P. Schneider, R. Scoccimarro, E. Sellentin, H. Y. Shan, M. von Wietersheim-Kramsta, and J. Zuntz. KiDS-1000 Methodology: Modelling and inference for joint weak gravitational lensing and spectroscopic galaxy clustering analysis. *arXiv e-prints*, page arXiv:2007.01844, July 2020.

[4] Elise Jennings, Rachel Wolf, and Masao Sako. A new approach for obtaining cosmological constraints from Type Ia Supernovae using Approximate Bayesian Computation. *arXiv e-prints*, page arXiv:1611.03087, November 2016.

[5] Justin Alsing, Tom Charnock, Stephen Feeney, and Benjamin Wand elt. Fast likelihood-free cosmology with neural density estimators and active learning. *MNRAS*, 488(3):4440–4458, September 2019.

[6] Yu-Chen Wang, Yuan-Bo Xie, Tong-Jie Zhang, Hui-Chao Huang, Tingting Zhang, and Kun Liu. Likelihood-free Cosmological Constraints with Artificial Neural Networks: An Application on Hubble Parameters and SN Ia. *arXiv e-prints*, page arXiv:2005.10628, May 2020.

[7] Stephen R. Green and Jonathan Gair. Complete parameter inference for GW150914 using deep learning. *arXiv e-prints*, page arXiv:2008.03312, August 2020.

[8] Simon Duane, A. D. Kennedy, Brian J. Pendleton, and Duncan Roweth. Hybrid Monte Carlo. *Physics Letters B*, 195(2):216–222, September 1987.