Case for support

Next generation cosmological analysis with nested sampling

Will Handley

1 Scientific case

Nested sampling [1] is a computational tool for extracting critical inference quantities from experimental and observational data. For example, given a dataset such as the cosmic microwave background, and our standard model of the universe ΛCDM [2], the critical concepts scientists are interested in are:

(a) What do the data tell us about the parameters of a model? (e.g. the size or age of a ΛCDM universe)

(b) How much do the data support a given model? (e.g. compared to a dynamic dark energy cosmology)

(c) Do different datasets make consistent predictions? (e.g. microwave background vs supernova data)

These three procedures are termed Parameter estimation, Model comparison and Tension quantification. The latter two of these have become increasingly relevant in recent years on account of discrepancies that have arisen in the context of the concordance model in the inferred value of the Hubble constant H_0 [3] by CMB and Supernovae data, clustering σ_8 [4] by CMB and weak lensing and curvature Ω_K [5, 6] by CMB and lensing/BAO, and between CMB datasets [7].

In the Bayesian paradigm adopted by many cosmologists [8], raw data via likelihood functions are processed on HPC machines by Markov-Chain Monte Carlo algorithms into data products, or sampling chains, which allow scientists to numerically infer the answers to the above questions. Parameter estimation can be accomplished by a variety of established techniques, whilst model comparison and tension quantification are computationally far harder tasks, requiring more specialist tools like nested sampling [9, 7, 10].

The Planck collaboration [11] enabled a great deal of community science by publishing the Planck Legacy Archive[1] (PLA). In this repository they provide the MCMC chains across a wide and deep grid of combinations of data and models. Since each individual element of this grid is theoretically and computationally expensive to produce, requiring HPC time and expert knowledge, the publication of the PLA grids (coupled with techniques and software for processing and reweighting them) has enabled scientists across the world to quickly perform research without the need to deploy HPC time themselves. However, the grid as it stands is primarily targeted at cosmological parameter estimation.

The primary outcome of this thematic project is to produce an analagous grid to the PLA, but using nested sampling chains to allow for model comparison and tension quantification in addition to parameter estimation, as well as updating to use to a broader variety of modern datasets. As nested sampling chains can also perform parameter estimation, this will act to both update and extend the PLA. The primary research goal is a systematic examination of which models best explain each dataset, and which datasets are inconsistent with each other. Patterns emerging from such a search will likely provide crucial insight into systematic errors in datasets, how to fix the cracks in our current concordance model of cosmology, or more excitingly what might replace it. The grid will be made publicly and permanently available on the open science platform Zenodo[2].

The secondary outcomes of this project will be to use the results of the partially completed grid to inform other analyses in collaboration with GAMBIT [12, 13] to combine particle physics data with cosmological observations in a coherent statistical framework in order to provide novel constraints on theories of dark matter and modified gravity. Such computational runs are more temporally inhomogeneous, requiring human cycles in between heavy CPU runs. The grid however once set up can be run continuously and started and paused at will, which can therefore be used to ensure more even usage over a long DiRAC allocation period.

The remainder of this document covers each of the assessment criteria detailed in the guidance notes[3].

[1] `https://pla.esac.esa.int/#cosmology`

[2] `https://about.zenodo.org/`

[3] `https://dirac.ac.uk/wp-content/uploads/2020/07/13th_Call_Guidance_Notes_Final.pdf`

1.1 Significance of the proposed research goals with reference to the STFC Roadmap

This thematic project covers several topics in the STFC programme[4]:

A:1 What are the laws of physics operating in the early Universe?

A:2 How did the initial structure in the universe form?

A:3 How is the universe evolving and what roles do dark matter and dark energy play?

With the number and variety of independent discrepancies that have been identified [3–7] in cosmological datasets, the community consensus is that whilst some or indeed many of these may be partially attributable to systematic experimental error, it is highly likely that our concordance model of the universe needs an upgrade or overhaul. As with any paradigm shift, this can either be accomplished by a modification of the contents of the theory (e.g. new types of dark matter or neutrinos), or by a modification of the forces in which they interact (e.g. adjustment of gravity or fundamental laws of physics at early times). The primary issue is that whilst we agree something is wrong, we have been unable to find a broadly convincing improvement to our concordance model.

Cosmological parameter estimation, model comparison and tension quantification form the observational backdrop to items [A:1,2,3] on the STFC roadmap. The premise of the grid part of this project is that a systematic scan over many models and dataset combinations will be able to reveal any previously undetected tensions, and indicate broader patterns that are not visible on a case-by-case basis. This will aid the community in disentangling systematics from physics, and provide crucial insight into suggesting the next cosmological paradigm. The GAMBIT sub-projects will provide a more targeted attempt to explore specific resolutions, which may come from modifications to the content or forces of the early or late-time universe (or indeed combinations thereof).

C:4 What is the nature of dark matter and dark energy?

With direct detection experiments on the dark contents of our universe still unable to provide anything beyond upper bounds, for the foreseeable future our best observational handle on dark matter and energy is to use several independent probes from both particle physics and cosmology. The GAMBIT subprojects investigate concrete dark matter models motivated by beyond standard model particle physics. Portions of these models parameter space may be ruled out by particle physics experiments and cosmological observations from either ends of the energy scale, and model comparison allows us to select the models most preferred by data.

A:7 What is the True Nature of Gravity?

C:3 What is the nature of space-time?

Astrophysics can probe gravity on the largest scales with cosmology [14], but also on smaller scales using time delay likelihoods from compact object mergers (constraints on propagation speeds allows us to rule out or refine massive gravity models). More interestingly, many beyond standard model theories have consequences over a whole range of scales, with the corollary that this coupling allows large scale cosmology and astrophysics in theory to probe small scale particle physics models, such as super-renormalisable Higgs portal models. GAMBIT provides a unique framework for joining all of these otherwise disparate pieces of data to inform us about the true nature of gravity, and the interaction of space-time with particle physics.

C:7 Are there new phases of strongly interacting matter?

The early universe provides a particle accelerator beyond the realms of modern day terrestrial machines. It is anticipated that over the next few decades as gravitational wave observatories in the LISA epoch come online that these will provide a cosmological window onto the gravitational wave background and thus a direct probe onto the primordial universe. Theories with novel phase transitions such as symmetry breaking QCD will be able to be directly observed with such devices. The 'forecasting' element of the GAMBIT project will thus inform the next generation of scientists as to what will be possible with next-generation machines, motivating further theoretical work in this field.

[4] `https://stfc.ukri.org/research/science-challenges/`

Strategic value within the STFC programme

The grid of models and datasets, alongside website tutorials and Python processing tools [15] will provide the theoretical and observational communities across STFC's programme of research with an invaluable resource for testing out their models without need to deploy HPC time. In addition to saving research time, energy and money by reducing replication of effort, the completeness of the grid gives value beyond the sum of its parts, as it allows systematic checking of models and data. Additionally, as the PLA has done for parameter estimation, this grid will standardise model comparison, meaning that comparability and reproducibility is enhanced across the STFC cosmology programme.

The GAMBIT subprojects, aside from addressing the key issues in the STFC roadmap detailed above, will bring the GAMBIT community into the DiRAC and STFC remit. This is a community with a proven ability to convert large HPC allocations (such as PRACE grants) on international machines (such as Joliot-Curie) into cutting edge science at the frontiers of both cosmology and particle physics, and their involvement in DiRAC allocations should prove a welcome addition to the STFC community.

1.2 Appropriateness of the proposed methods/codes

The primary driving code `CosmoChord` [16] was written and has been maintained and operated for the past seven years by the PI. The code also has significant community up-take which has served debug the code across a wide range of cosmological models, meaning that running it across a wide and deep grid of models and datasets is a relatively low-risk affair.

`PolyChord` [17], the underlying nested sampling code behind `CosmoChord`, is the state-of-the-art in high-dimensional nested sampling, and the only implementation of nested sampling capable of sampling a hierarchy of fast-slow parameters which is essential for cosmological applications where in general foregrounds and calibration require the introduction of many nuisance parameters (this can rise to the region of ~ 40 when Planck and DES likelihoods together). It is written in FORTRAN90 and is highly optimised. As covered in the technical assessment, it have excellent HPC scaling properties, and are ideally suited to problems of this scale on CSD3 (a DiRAC-like system).

GAMBIT [12, 13] has a similarly wide community user base and has been tested and optimised on many HPC machines. The framework is written in C++ and designed as a flexible backending build and install system to seamlessly incorporate new data and likelihoods from both astronomy and particle physics communities. At the time of writing it is the only code capable of combining cosmological and particle physics likelihoods in a unified framework, and unlike traditional cosmology build systems (such as `cobaya`, `CosmoMC`, `cosmosis` or `MontePython`) it provides frequentist techniques alongside Bayesian samplers. `PolyChord` is integrated into GAMBIT's build system.

1.3 Justification of the requested resources

CPU time

In general a nested sampling run on ΛCDM takes roughly four hours per live point. This value depends on several factors, such as the number of live points, prior widths, degree of posterior constraint and number of nuisance parameters associated with a dataset. More importantly, this runtime will increase by a factor of up to twenty for more complicated models such as those involving multiple neutrino species and/or curvature. Science-grade runs generally have 1000 live points, so to run this over the roughly 1000 elements of the grid, this comes out as 4 million hours of ΛCDM runs, or 80 million for more complicated models. This project takes an average over the grid as a whole and asks for 30M CPU hours, with any additional time absorbed by sub-projects or widening th egrid. A more precise estimate will not be available until the grid is begun on account of interactions between combined likelihood codes which affect degrees of constraint, but this will be reviewed every month so that the plan of the grid can be expanded or prioritised accordingly. The GAMBIT community has been historically awarded PRACE grants of order 50M core hours to satisfy their working group needs, which justifies the GAMBIT sub projects being included as a fractional part of this 30M allocation. See project management structure for more detail.

Storage

Each dataset under consideration is at most a few GB, so for the ten datasets, PolyChord and GAMBIT codes, only tens of GB are necessary for fast and backed up storage. At runtime, codes can generate large output files for resuming a run (up to the size of the memory limit of the machine which is 192G), so if several runs are in progress it is prudent to have a few terabytes of `/work` storage, and an order of magnitude more `/data` for temporary storage of data for periodic review of slices of the grid. For permanent storage and distribution resume files are removed and these runs are compressed to 100s of MB which will be uploaded to Zenodo, so no archive space is needed. See Data Management plan for more detail.

The project therefore requests 30GB of `/home`, 3TB of `/work` and 30TB of `/data`

1.4 Justification for any research software engineering support requested

Not applicable

1.5 Suitability of the investigator(s) for the proposed research

Will Handley is a newly appointed Royal Society University Research Fellow at the University of Cambridge, with a 5+3 year research programme in the field of cosmological parameter estimation, model comparison and tension quantification, entitled "Bayesian machine learning and tensions in cosmology". The DiRAC allocation requested in this proposal enhances his existing highly competitively awarded fellowship, giving him access to an order of magnitude more computing power which enables a more systematic and publicly usable grid of results to be computed using the same tools and techniques he deploys for his ongoing research programme.

He has a strong track record of publication in the fields of astrophysics and cosmology, with 73 articles[5] published or under review since he began his PhD 8 years ago. He has participated in several collaborations of a range of sizes (Planck, CORE, GAMBIT, REACH), and shown evidence of leadership even at an early career stage in graduate student and postdoctoral supervision, writing subsequently awarded grants and leading working groups of GAMBIT and REACH. More detail can be found on his CV[6].

He has 7 years and 5 million hours of HPC experience both for his own research and as part of the Planck [18], CORE [19] and GAMBIT [12, 13] collaborations. As his first DiRAC application, an awarded allocation will allow him to further scale up his HPC research. Over the course of the 5+3 year URF research programme he plans to bring in more PhD and postdocs in both a funded and collaborative capacity, and hopes that DiRAC allocations will form an integral part of his and his group's research programme going forward.

GAMBIT[7] is an international community of ~ 60 particle physicists, cosmologists and statisticians who maintain the community code and participate in working groups which aim to use both particle collider experimental data and astrophysical observations in a coherent statistical framework (either Bayesian or Frequentist). Will Handley was appointed convener of the GAMBIT cosmology working group in February of this year at the last face-to-face meeting in Melbourne, with the intention of taking this leadership role for the long term future of the group after the first stable version release [12, 13]. There is therefore a pool of HPC and physics expertise from this community that the PI can draw upon over the three years of the allocation to aid in the GAMBIT thematic sub-projects.

1.6 Justification of any periods of machine use in exclusive mode

Not applicable

[5] https://arxiv.org/a/handley_w_1.html
[6] https://github.com/williamjameshandley/CV/raw/master/CV.pdf
[7] https://gambit.hepforge.org/

1.7 A prioritised list of the projects within the proposal

1.7.1 Main grid

The final grid will be across a cube with three axes

- 10 datasets: Planck temperature and polarisation [20], CMB lensing [21], ACT [22], SPT [23], DES [24], BAO [25], KiDS [26], Pantheon [27], JLA [28], SH0ES [29].
- Pairwise combinations of the above
- 18 models from the PLA (base, Alens, Aphiphi, alpha1, mnu, nnu, nnu+meffsterile, nnu+mnu, nnu+nrun, nnu+yhe, nrun, nrun+nrunrun, nrun+r, omegak, r, w, w+wa, yhe)

although technically because $A+B=B+A$ for the purposes of data combination it is in reality a triangular prism rather than a cube. Whilst these are the preliminary sets of data, there is capacity to expand or contract the grid as necessary, so if we have N_{data} datasets and N_{model} models, the total time for the grid T_{total} is

$$T_{\mathrm{total}} = T_{\mathrm{run}} \times \frac{1}{2}N_{\mathrm{data}}(N_{\mathrm{data}}+1) \times N_{\mathrm{model}}, \qquad 4\mathrm{h} \lesssim T_{\mathrm{run}}/\mathrm{nlive} \lesssim 80\mathrm{h}, \qquad \mathrm{nlive} \sim 1000 \tag{1}$$

where T_{run} is an estimate for the time taken for an individual run, which is model, dataset and prior dependent, but in practice is anywhere between $4\times n_{\mathrm{live}}$ and $80\times n_{\mathrm{live}}$ cpu hours, where n_{live} is a sampling resolution parameter, and $n_{\mathrm{live}}\sim 1000$ is considered science grade for cosmology. Runs are parallelised up to the number of live points, so the time frame of 4–80 hours per run is a good proxy for walltime. Expansion of the grid on either models or data axes may occur either if runs proceed faster than anticipated, or if more time is granted (for example in follow-up allocations).

1.7.2 GAMBIT: dark matter constraints

This sub-project will focus around constraining particle physics inspired candidates for dark matter. For example, axion-like particles are a favoured by theoreticians, and decaying subcomponents may be constrained using Big Bang nucleosynthesis and spectral distortions (from COBE-FIRAS). Using late time cosmological datasets such as the Lyman-alpha forest one can constrain dark matter theories which contain interactions. With the recent claimed detection of the 21cm global signal by EDGES [30], there was a flurry of theoretical papers designing exotic dark matter models to explain why the absorption trough was deeper than expected. In order to sift through these potential theories, one requires additional constraints from other areas of particle physics and cosmology, which GAMBIT is ideally suited to bringing in. Will Handley is involved in the Cambridge-based 21cm global signal detection experiment REACH and will exploit his position on both collaborations to bring global 21cm cosmology into GAMBIT.

1.7.3 GAMBIT: particle physics and modified gravity

This sub-project focuses on beyond standard model extensions which also implicitly include modified gravity. Models including Higgs portal mechanisms can have consequences on a whole range of scales, and therefore can be informed by a wide variety of probes from solar neutrino observations through modified gravity likelihoods and X-ray and stellar astronomy to CMB and BBN observables. In this instance recent gravitational wave data can feed into these analyses using time-delay likelihoods from compact object mergers with electromagnetic counterparts, where constraints on propagation speeds can inform massive gravity models.

1.7.4 GAMBIT: forecasting

This sub-project looks more long-term into the capability of GAMBIT to constrain fundamental particle physics using astronomical observations. With gravitational wave observatories in the LISA era, we will have access to a cosmological gravitational wave window probing the universe at extremely high energies, where beyond standard model predictions such as phase transitions and symmetry breaking QCD effects become relevant. Here we would build simulated likelihoods which will have similar constraining power, and examine the synergies that come from other observables, and the kind of particle physics questions which an observatory of this scale will be able to answer.

1.8 Absolute minimum time required for the proposed work

As quantified in equation (1), this project is highly scalable in both directions. A minimum viable product could be viewed as a few models and datasets $N_{\mathrm{models}} \sim N_{\mathrm{data}} \sim 5$, which taking a value of $T_{\mathrm{run}} \sim 30$kh yields $T_{\mathrm{total}} \sim 2.250$Mh. Anything less ambitious than this could arguably be accomplished comfortably on the free CSD3 tier.

The converse of this observation is that it can scale to fit the allocation it is given by either adding more models or more datasets. In general therefore the scientific value scales approximately linearly with the number of models considered, and by square root of the number of datasets. Were more time to be allocated or become available, additional models would be prioritised over additional datasets in proportion to this relation.

This same scalability can be used to de-risk the project in the event that the run-time estimates are either too large or too small, as detailed in Section 2.7.

The last CosmoBit paper [12] consumed approximately 1M hours for its neutrino likelihoods, and we anticipate this as a lower bound for each of the three GAMBIT sub projects.

We therefore estimate a minimum of 2.25 + 3M hours.

1.9 Code usage

See section 2.9 of the scientific proposal form.

2 Project & data management

2.1 The project management structure

The thematic project is structured with one large grid, which is flexible and straightforwardly structured (and able to be run in the background, be scaled up, down or paused using scontrol commands as necessary), and three GAMBIT sub-projects with more temporally inhomogeneous loads (with a series of large jobs, followed by a reasonable quantity of parallelised human cycles of analysis which inform the design of the next batch of cpu cycles). The strength of this thematic project as a whole from a DiRAC allocation perspective is that without much effort the grid-based set of runs can be used to 'smooth out' the temporal inhomogeneity of the other three, ensuring even usage over a long period of time.

2.2 Data management plan

2.2.1 The types of data generated

Nested sampling runs generate several types of output files: Posterior samples, runtime statistics files, parameter names, resume files, dead points files, live points files, prior details. These are of varying size, the largest of which by some margin are the resume files giving the precise implementation details needed to resume a run, and can run to tens to hundreds of GB in size.

2.2.2 Selection and preservation of data

A near lossless compression of a run is achieved by simply saving the 'dead points file' containing the birth contours, which are 10-100MB in side. With the correct software [15] the rest of the files may be reconstructed. These data files will be preserved in compressed form on the free and permanent open science server Zenodo, with a corresponding DOI. Once a run is completed, it is wasteful to preserve resume files, so these will be deleted after the run's validity has been confirmed.

2.2.3 Software and metadata

Alongside the data, the precise software versions will be archived (which are all also available under version control available on GitHub). All input files required to run the software will also be clearly provided alongside the compressed output files.

An example of this in action can be seen in these data products [31], and an example of their reuse in this paper [32].

2.2.4 How long the data will be preserved

Zenodo represents the state-of-the-art in indefinite open storage offering a strong guarantee far beyond the lifetime of this project.

2.2.5 Which data will be shared

All preserved data detailed above will be shared in a fully open manner on Zenodo.

2.2.6 Proprietary considerations

Whilst there are no intellectual property considerations, the PI and collaborators will potentially withold any data products for at most six months after they are produced and processed. This embargo period is to allow any more controversial or unexpected results to be fully explored before data release.

2.3 Availability of sufficient researcher effort to carry out the proposed research

As this thematic project acts to enhance the science outlined in the PI's University Research Fellowship proposal, the PI effectively has 100% time commitment to using these resources.

For the grid, after a few days of setting up the initial grid and cross-checking scripts, it is anticipated that this will be a relatively low cost in human-cycles (although high cost in CPU cycles), requiring roughly on average 5% time to check on the grid, process any results and inspect and deal with any bugs which may or may not arise.

For the other sub-projects, this again will form part of the primary research of the PI and any other researchers making use of the less temporally homogeneous aspects of the proposal.

2.4 A description of the internal allocation process for the allocation of time to sub-projects

The grid will fill up the grid "from the corner" gradually adding models and datasets in proportion to their fractional contribution to the grid. This will ensure that at each stage the grid is maximally complete enabling science results to be extracted as the grid fills up.

Once the grid is up and running, it will be able to be paused and resumed using `scontrol` statements to make space for the larger bursts of activity from the other thematic subprojects. Such pausing and resuming will only occur once a proposed "interjection run" has been thoroughly tested on login and interactive nodes, and any pausing will be approved, performed and synchronised by the PI.

2.5 Work plan & Milestones

WP-1.1: Grid Preliminary sweep

2021 Q2

Initial sweep through LCDM + all datasets and Planck + all models. This will provide a high-level view on in practice how long each run will take and reveal any bugs which have not yet been rooted out by the community. $N_{\text{data}} + N_{\text{model}}$ runs complete. Write scripts for automating job submission and checking. Set up GitHub repository with frontpage website and tutorial of how to process runs, and Zenodo uploads of initial sweep.

WP-1.2: Model selection grid

2021 Q3 – 2022 Q2

Cover all models and all individual datasets. $N_{\text{data}} \times N_{\text{model}}$ runs complete. This sweep should be first by dataset, then by model. Release first paper indicating the layout of the grid, any science arising or expected to arise, and a forward reference to the full grid with scanning strategy.

WP-1.3: Tension quantification grid

2022 Q3 – 2024 Q1

Sweep through all models and all pairwise combinations of datasets. $\frac{1}{2}N_{\text{data}}(N_{\text{data}}+1) \times N_{\text{model}}$ runs completed. This sweep will be set up such that it is evenly balanced and maximally complete at any given stage. Produce second paper highlighting global patterns that have emerged from the grid, and suggestions of future directions in cosmology.

WP-2→4: GAMBIT subprojects

These work packages are structured similarly, with each expecting to take a year, and will be approximately sequentially executed in the order detailed in Section 1.7, (WP-2 in 2021, WP-3 in 2021, WP-4 in 2023). In general the GAMBIT cosmology working group begins with a preliminary run, followed by a detailed analysis of the results to work out if any of the particle physics or cosmological modelling need improvement. This is then followed up by large scale HPC runs for production quality. The expense of the particle physics likelihoods and complexity of the parameter spaces which require navigating mean that there are usually several cycles of science-grade runs before the community are confident in releasing the results.

2.6 Publication plans

As detailed above, the grid itself will have a master publication laying out the key strategy and design choices which will form a citation point that is clearly advertised in both the download repository and a tutorial site. This will be released after WP-1.2 is completed. There will then be a second paper at the end of WP-1.3 highlighting any global patterns that have emerged and suggestions of future directions in cosmology, and the future of the grid if further funding is to be gained in expanding it after this initial run. It will be made clear that any publications using grid results should cite DiRAC and STFC resources.

Any particularly interesting tensions or science arising from this systematic examination of the current state-of-the-art datasets will be published as separate publications, which will also serve as milestones of what point the grid has reached. Each of the GAMBIT work packages will have at least one associated paper.

2.7 Project risk and mitigation analysis

As both the sampling code and the likelihoods are all relatively well-tested there is little risk in the grid analysis not being able to be completed. In general, nested sampling has a tendency to catch bugs in likelihoods and Boltzmann codes that only occur in the deep tails of the distribution (and hence have been missed by a more traditional MCMC chain), but these can be quickly fixed by existing collaborations the PI has with maintainers of likelihood codes. If the bugs are insurmountable, this risk can be mitigated by narrowing the prior widths and correcting evidences later, although this manual intervention should be viewed as a last resort. The interchangeability of grid elements means that whilst bugs are being fixed, other more stable models or datasets can continue unimpeded.

The primary risk is associated with the fact that there are substantial error bars on the estimates as to how long the grid will take. However, the flexibility and expandability of the grid, alongside the scanning strategy mitigate this risk from both directions. If estimates are conservative, the grid can be expanded by adding more models of interest, or increasing the number of live points for greater accuracy. If the runtime has been underestimated, the scanning strategy of the grid ensures that the subset that has been run is "maximally complete"

As this project is not storage intensive (and the project plans to upload all data to zenodo), there is no risk of storage running out, but there would be a small risk of running out of disk space temporarily. The additional large `/data` allocation, alongside regular checks and cron jobs for python scripts to examine quotas and notify users via email if the disc is close to full will mitigate this risk.

References

[1] John Skilling. Nested sampling for general bayesian computation. *Bayesian Anal.*, 1(4):833–859, 12 2006. doi: 10.1214/06-BA127. URL `https://doi.org/10.1214/06-BA127`.

[2] Douglas Scott. The Standard Model of Cosmology: A Skeptic's Guide. *arXiv e-prints*, art. arXiv:1804.01318, April 2018.

[3] Licia Verde, Tommaso Treu, and Adam G. Riess. Tensions between the early and late Universe. *Nature Astronomy*, 3:891–895, September 2019. doi: 10.1038/s41550-019-0902-0.

[4] Shahab Joudaki, Alexander Mead, Chris Blake, Ami Choi, Jelte de Jong, Thomas Erben, Ian Fenech Conti, Ricardo Herbonnet, Catherine Heymans, Hendrik Hildebrandt, Henk Hoekstra, Benjamin Joachimi, Dominik Klaes, Fabian Köhlinger, Konrad Kuijken, John McFarland, Lance Miller, Peter Schneider, and Massimo Viola. KiDS-450: testing extensions to the standard cosmological model. MNRAS, 471(2):1259–1279, October 2017. doi: 10.1093/mnras/stx998.

[5] Will Handley. Curvature tension: evidence for a closed universe. *arXiv e-prints*, art. arXiv:1908.09139, August 2019.

[6] Eleonora Di Valentino, Alessandro Melchiorri, and Joseph Silk. Planck evidence for a closed Universe and a possible crisis for cosmology. *Nature Astronomy*, 4:196–203, February 2020. doi: 10.1038/s41550-019-0906-9.

[7] Will Handley and Pablo Lemos. Quantifying the global parameter tensions between ACT, SPT and Planck. *arXiv e-prints*, art. arXiv:2007.08496, July 2020.

[8] Roberto Trotta. Bayes in the sky: Bayesian inference and model selection in cosmology. *Contemporary Physics*, 49(2):71–104, March 2008. doi: 10.1080/00107510802066753.

[9] Will Handley and Pablo Lemos. Quantifying tensions in cosmological parameters: Interpreting the DES evidence ratio. Phys. Rev. D, 100(4):043504, August 2019. doi: 10.1103/PhysRevD.100.043504.

[10] Pablo Lemos, Fabian Köhlinger, Will Handley, Benjamin Joachimi, Lorne Whiteway, and Ofer Lahav. Quantifying Suspiciousness within correlated data sets. MNRAS, 496(4):4647–4653, June 2020. doi: 10.1093/mnras/staa1836.

[11] Planck Collaboration. Planck 2018 results. VI. Cosmological parameters. A&A, 641:A6, September 2020. doi: 10.1051/0004-6361/201833910.

[12] The GAMBIT Cosmology Workgroup. Strengthening the bound on the mass of the lightest neutrino with terrestrial and cosmological experiments. *arXiv e-prints*, art. arXiv:2009.03287, September 2020.

[13] The GAMBIT Cosmology Workgroup. CosmoBit: A GAMBIT module for computing cosmological observables and likelihoods. *arXiv e-prints*, art. arXiv:2009.03286, September 2020.

[14] W. E. V. Barker, A. N. Lasenby, M. P. Hobson, and W. J. Hand ley. Addressing H_0 tension with emergent dark radiation in unitary gravity. *arXiv e-prints*, art. arXiv:2003.02690, March 2020.

[15] Will Handley. anesthetic: nested sampling visualisation. *The Journal of Open Source Software*, 4:1414, May 2019. doi: 10.21105/joss.01414.

[16] W. J. Handley, M. P. Hobson, and A. N. Lasenby. polychord: nested sampling for cosmology. MNRAS, 450: L61–L65, June 2015. doi: 10.1093/mnrasl/slv047.

[17] W. J. Handley, M. P. Hobson, and A. N. Lasenby. POLYCHORD: next-generation nested sampling. MNRAS, 453(4):4384–4398, November 2015. doi: 10.1093/mnras/stv1911.

[18] Planck Collaboration. Planck 2018 results. X. Constraints on inflation. A&A, 641:A10, September 2020. doi: 10.1051/0004-6361/201833887.

[19] CORE Collaboration. Exploring cosmic origins with CORE: Inflation. J. Cosmology Astropart. Phys., 2018(4): 016, April 2018. doi: 10.1088/1475-7516/2018/04/016.

[20] Planck Collaboration. Planck 2018 results. V. CMB power spectra and likelihoods. A&A, 641:A5, September 2020. doi: 10.1051/0004-6361/201936386.

[21] Planck Collaboration. Planck 2018 results. VIII. Gravitational lensing. A&A, 641:A8, September 2020. doi: 10.1051/0004-6361/201833886.

[22] ACT collaboration. The Atacama Cosmology Telescope: DR4 Maps and Cosmological Parameters. *arXiv e-prints*, art. arXiv:2007.07288, July 2020.

[23] SPT Collaboration. Maps of the Southern Millimeter-wave Sky from Combined 2500 deg^2 SPT-SZ and Planck Temperature Data. ApJS, 239(1):10, November 2018. doi: 10.3847/1538-4365/aae694.

[24] DES Collaboration. Dark Energy Survey year 1 results: Cosmological constraints from galaxy clustering and weak lensing. Phys. Rev. D, 98(4):043526, August 2018. doi: 10.1103/PhysRevD.98.043526.

[25] eBOSS Collaboration. 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*, art. arXiv:2007.08991, July 2020.

[26] KiDS Collaboration. KiDS-1000 Cosmology: Cosmic shear constraints and comparison between two point statistics. *arXiv e-prints*, art. arXiv:2007.15633, July 2020.

[27] Pantheon. The Complete Light-curve Sample of Spectroscopically Confirmed SNe Ia from Pan-STARRS1 and Cosmological Constraints from the Combined Pantheon Sample. ApJ, 859(2):101, June 2018. doi: 10.3847/1538-4357/aab9bb.

[28] JLA. Improved cosmological constraints from a joint analysis of the SDSS-II and SNLS supernova samples. A&A, 568:A22, August 2014. doi: 10.1051/0004-6361/201423413.

[29] Adam G. Riess, Stefano Casertano, Wenlong Yuan, Lucas M. Macri, and Dan Scolnic. Large Magellanic Cloud Cepheid Standards Provide a 1% Foundation for the Determination of the Hubble Constant and Stronger Evidence for Physics beyond ΛCDM. ApJ, 876(1):85, May 2019. doi: 10.3847/1538-4357/ab1422.

[30] Judd D. Bowman, Alan E. E. Rogers, Raul A. Monsalve, Thomas J. Mozdzen, and Nivedita Mahesh. An absorption profile centred at 78 megahertz in the sky-averaged spectrum. Nature, 555(7694):67–70, March 2018. doi: 10.1038/nature25792.

[31] Will Handley. Curvature tension: evidence for a closed universe (supplementary inference products), August 2019. URL `https://doi.org/10.5281/zenodo.3371152`.

[32] Andrew Fowlie, Will Handley, and Liangliang Su. Nested sampling cross-checks using order statistics. MNRAS, 497(4):5256–5263, August 2020. doi: 10.1093/mnras/staa2345.