PHYSICAL REVIEW D **93,** 043522 (2016)

Are cosmological data sets consistent with each other within the Λ cold dark matter model?

Marco Raveri[1,2,3]

[1]*SISSA–International School for Advanced Studies, Via Bonomea 265, 34136 Trieste, Italy*
[2]*INFN, Sezione di Trieste, Via Valerio 2, I-34127 Trieste, Italy*
[3]*INAF-Osservatorio Astronomico di Trieste, Via G.B. Tiepolo 11, I-34131 Trieste, Italy*

(Received 12 October 2015; published 16 February 2016)

We use a complete and rigorous statistical indicator to measure the level of concordance between cosmological data sets, without relying on the inspection of the marginal posterior distribution of some selected parameters. We apply this test to state of the art cosmological data sets, to assess their agreement within the Λ cold dark matter model. We find that there is a good level of concordance between all the experiments with one noticeable exception. There is substantial evidence of tension between the cosmic microwave background temperature and polarization measurements of the Planck satellite and the data from the CFHTLenS weak lensing survey even when applying ultraconservative cuts. These results robustly point toward the possibility of having unaccounted systematic effects in the data, an incomplete modeling of the cosmological predictions or hints toward new physical phenomena.

DOI: 10.1103/PhysRevD.93.043522

Our present understanding of the Universe is based on the combination of several different cosmological observations that are joined in order to exploit their complementary sensitivities to distinct characteristics of our Universe. Several supernovae (SN) surveys are added together, in a single catalog, to measure the expansion history at late times. Different baryon acoustic oscillation (BAO) surveys provide independent measurements of a cosmological standard ruler at several times. Large scale structure and weak lensing surveys measure the correlation of galaxies and weak lensing shear in many different redshift bins. These are combined together to get tomographic information on the clustering of cosmic structures. At last, measurements of the cosmic microwave background (CMB) are reaching an extraordinary level of sensitivity, which allows one to measure the CMB temperature fluctuations along with CMB polarization and lensing. These data are then joined together to exploit CMB sensitivity to both early and late times cosmology.

In the future, cosmological studies are going further in this direction. Wide large scale structure surveys, like Euclid [1], will combine maps of galaxies at several different redshifts that will be joined with measurements of the CMB from the Planck satellite [2] and suborbital experiments.

The observational efforts that are driving cosmology toward a phase of extremely accurate, large scale, measurements, will all be joined together to learn all possible information about the initial conditions and the evolution of our Universe. In this program, however, a problem arises.

How can we be sure that the data sets that we will be collecting, form a coherent picture, when interpreted within a model? How do we quantify the agreement between them, to be aware of the possible presence of unaccounted systematic effects or hints toward new physical phenomena?

Testing the agreement between data sets, in a rigorous way that goes beyond the comparison of the marginal distribution of some parameters, is critical in answering these questions. The posterior of the model parameters is, in fact, not guaranteed to show tensions due to the marginalization procedure, which can alter discrepancies that will be then misjudged. Assessing whether the posterior distribution of two different data sets occupies a substantially different volume in the parameter space of a model is instead crucial as it could provide a useful guidance for the future research. Answering these questions is also a useful sanity check for parameters estimation. The statistical inference on the parameters of a model should get stronger as we combine together different measurements and should not reflect the fact that we are joining low probability tails of the model posterior.

An estimate of the tensions between different data sets, based on the marginal posterior of cosmological parameters, has shown that indeed some discrepancies arise when combining several probes [3–7] that could point toward some extensions of the fiducial model [8–13].

In this paper we briefly review how Bayesian inference can be used to answer quantitatively these questions and we comment on the advantages and possible drawbacks of this approach. We apply, for the first time, this statistical test to state of the art cosmological measurements and the Λ cold dark matter (ΛCDM) model. We report and interpret the results commenting on their relevance for future studies.

The data concordance test (DCT).—Bayesian statistics provides a clean way of dealing with the problem of combining data sets by means of hypotheses testing and in particular with its application to the problem of classification and decision making. We have two data sets and we want to test whether we can describe them with the same set of parameters or not, within a given model. Based on the outcome of such an operation we shall take a decision about combining them [14–20].

Let us now consider two data sets D_1 and D_2 and a model $\mathcal{M}$. The two competing hypotheses that we want to compare are

(i) $\mathcal{I}_0$: the two data sets can be characterized, within model $\mathcal{M}$, by the same (unknown) parameters;

(ii) $\mathcal{I}_1$: the two data sets can be described, within model $\mathcal{M}$, with different (unknown) parameters.

Then, we compare the evidences for these two statements to obtain their odds ratio. In particular by assigning noncommittal priors for the two hypotheses we immediately have

$$\mathcal{C}(D_1, D_2, \mathcal{M}) = \frac{P(D_1 \cup D_2 | \mathcal{I}_0, \mathcal{M})}{P(D_1 \cup D_2 | \mathcal{I}_1, \mathcal{M})} = \frac{P(D_1 \cup D_2 | \mathcal{M})}{P(D_1 | \mathcal{M}) P(D_2 | \mathcal{M})}, \quad (1)$$

where $P(D_1 \cup D_2 | \mathcal{M})$ is the evidence of the joint data sets and $P(D_1 | \mathcal{M})$ and $P(D_2 | \mathcal{M})$ are the evidences of the single data sets. The last equality follows from the definition of the two hypotheses and the fact that under the hypothesis $\mathcal{I}_1$, the two data sets, D_1 and D_2, pertain to distinctly different classes and knowledge of one of them tells us nothing about the other.

We can interpret the odds resulting from this calculation with a classification scheme, like the Jeffreys' scale or others, depending on the decision that we have to perform afterwards. In particular, when we have to decide if it is appropriate to combine two data sets, we can establish a threshold for the positive answer, based on the risk that we are willing to take, and act accordingly. A common choice [14] with this respect is to decide to follow the $\mathcal{I}_0$ hypothesis, combining the data, if $\log \mathcal{C} > 0$ and $\mathcal{I}_1$ otherwise.

The DCT is relatively easy to compute, once we have at our disposal the tools to perform efficient evidence computations, and has some other advantages. First of all the DCT is a quantitative and statistically rigorous prescription. It measures the odds, within model $\mathcal{M}$, of obtaining one data set given the other one. Tensions between them are quantified in terms of odds of agreement or disagreement and are not based on the marginal distribution of the parameters. While the latter approach might point in the right direction if the likelihood is Gaussian, both in the data and the parameters, it might fail as soon as the posterior is slightly non-Gaussian. Indeed, with the above assumptions, it can be shown [14,15] that the DCT reduces to the usual prescription for the marginal posterior of uncorrelated parameters. When these requirements break down, however, the inspection of the parameters posterior becomes unreliable, in assessing tensions, as it tends to be biased. In addition, as common sense suggests, the DCT naturally favors the combination of data sets, as long as there is no strong evidence that should not be done [14,17]. The way in which this is automatically encoded in the computation of $\mathcal{C}$ is by weighting the prior volume with the likelihood volume, in a manner that resembles the Occam razor common to Bayesian model selection. If there is no clear indication on how to set the prior ranges, i.e. the previous knowledge of the model is vague, and the prior are consequently wide, the DCT favors the combination of data sets, as this might help in gaining knowledge of the model. Conversely if the priors are stronger than the data the DCT will disfavor the combination, as we already included in the prior choice the information that is coming from the combination of the data sets.

The DCT has also some disadvantages. It does not give any indication whether the model is good by itself in fitting the data. Being a comparative test, we can use it to judge if the agreement, within a given model, improves or not when combining two data sets but it is possible to have a model that fits very badly the data while the DCT might still favor their combination. Another problem that is particularly relevant when the DCT is used more than once on some data sets, is that it is not robust against over fitting. As immediately follows from the previous points, enlarging the parameter space with the introduction of an additional parameter will not decrease $\mathcal{C}$. As a consequence, it is always possible to relax a tension between different measurements by introducing a new parameter, being it just a nuisance parameter, describing some systematic effects, or a parameter related to a different underlying physical modeling. For this reason it is critical to use other statistical tools to assess whether the introduction of the additional parameter is really justified. It is worth noticing that as a by-product of the computation of $\mathcal{C}$, for the two different models, one has the relevant information to perform evidence based model comparison. The last source of biases in the DCT is due to unaccounted correlations between the data sets. If two data sets are assumed to have independent errors and this is not the case, $\mathcal{C}$ will be biased toward positive values if the covariance between the errors of the two experiments is positive and toward negative values in the opposite case [17].

Data sets and model.—We use several available cosmological data sets to perform a DCT over all the possible independent data couples, within the ΛCDM model.

The first data set that we consider consists of the "Joint Light-curve Analysis" (JLA) Supernovae sample, as introduced in [21], which is constructed by the combination of the SNLS, SDSS and HST SNe data, together with several low redshift SNe.

We use the WiggleZ Dark Energy Survey (WZ) [22] measurements of the galaxy power spectrum as inferred from 170,352 blue emission line galaxies over a volume of

1 Gpc^3 [23,24] up to $k_{\max} = 0.2\ h/\mathrm{Mpc}$. We marginalize over a scale independent linear galaxy bias for each of the four redshift bins, as in [24].

The third data set that we examine consists of the measurements of the galaxy weak lensing shear correlation function as provided by the Canada-France-Hawaii Telescope Lensing Survey (CFHTLenS) [3]. This is a 154 square degree multicolor survey, optimized for weak lensing analyses, which spans redshifts ranging from $z \sim 0.2$ to $z \sim 1.3$. Here we consider the data subdivided into six redshift bins and we applied ultraconservative cuts, as in [11], that exclude ξ_- completely and cut the ξ_+ measurements at scales smaller than $\theta = 17'$ for all the tomographic redshift bins. As discussed in [11], these cuts make the CFHTLenS data insensitive to the modeling of the nonlinear evolution of the power spectrum. For the Planck best-fit ΛCDM cosmology [5] these cuts correspond to: $k_{\max} = 0.18\ \mathrm{Mpc}^{-1}$ for the CFHTLenS bin with mean redshift $\bar{z} = 0.36$; $k_{\max} = 0.15\ \mathrm{Mpc}^{-1}$ for $\bar{z} = 0.50$; $k_{\max} = 0.13\ \mathrm{Mpc}^{-1}$ for $\bar{z} = 0.68$; $k_{\max} = 0.12\ \mathrm{Mpc}^{-1}$ for $\bar{z} = 0.87$; $k_{\max} = 0.11\ \mathrm{Mpc}^{-1}$ for $\bar{z} = 1.00$ and $\bar{z} = 1.16$.

We include in this study the measurements of the CMB fluctuations in both temperature and polarization as released by the Planck satellite [5,25]. At large angular scales the Planck release implements a joint pixel-based likelihood including both temperature and E-B mode polarization for the multipoles range of $\ell \leq 29$, as described in [25]. At smaller angular scales we use the Plik likelihood [25] for CMB measurements of the TT, TE and EE power spectra, as extracted from the 100, 143, and 217 GHz high frequency instrument channels. We refer to the combination of the low-ℓ TEB measurements and the high-ℓ TT TE EE data as the CMB compilation.

We also include in the analysis the Planck 2015 full-sky lensing potential power spectrum [26] in the multipoles range $40 \leq \ell \leq 400$ as obtained with the SMICA code, hereafter called CMBL.

We also employ BAO measurements of: the SDSS main galaxy sample at $z_{\mathrm{eff}} = 0.15$ [27]; the BOSS DR11 “LOWZ” sample at $z_{\mathrm{eff}} = 0.32$ [28]; the BOSS DR11 CMASS at $z_{\mathrm{eff}} = 0.57$ of [28]; and the 6dFGS survey at $z_{\mathrm{eff}} = 0.106$ [29], all joined together in the data set that we dub BAO.

In addition we consider the redshift space distortion (RSD) measurements of BOSS CMASS-DR11 as analysed in [30] and [31]. When these data are used we exclude the BOSS-CMASS results of [28] from the BAO likelihood to avoid double counting. We refer to the data set obtained by combining BOSS CMASS-DR11 RSD measurements and the above BAO measurements, excluding BOSS-CMASS points, as the RSD one.

By means of the DCT we perform a test of the data concordance within the six parameter ΛCDM model. To compute nonlinear corrections to the matter power spectrum and the lensed CMB power spectra, we use the halofit approach [32] with the updates of [33]. We notice that the halofit approach gives a reliable fitting formula over a limited volume in parameter space [33] and this might result in biases in the evidence computation. We use the CAMB code [34,35] to compute the predictions for all cosmological observables of interest and we use the likelihoods of the previously described data sets, as implemented in CosmoMC [36]. We compute the evidence by means of the nested sampling algorithm and its implementation in the PolyChord code [37,38]. The PolyChord code also outputs error estimates, intrinsic to the nested sampling evidence calculation, which are reported as well. In order to assess the agreement between the above data sets, in the setup commonly used for parameter estimation, we use the standard CosmoMC prior on the ΛCDM model parameters as summarized in Table I.

TABLE I. Prior on the six cosmological parameters of the ΛCDM model.

	Prior on cosmological parameters
Parameter	Prior range
$\Omega_b^0 h^2$	[0.005, 0.1]
$\Omega_c^0 h^2$	[0.001, 0.99]
$100\theta_{MC}$	[0.5, 10]
τ	[0.01, 0.8]
n_s	[2, 4]
$\ln(10^{10} A_s)$	[0.8, 1.2]

Results.—The results of the DCT of all the independent couples of the data sets described above are shown in Fig. 1. We can see that the combination of CMB and WL data shows evidence of substantial disagreement. It is worth noticing that the marginal distribution of the parameters is not displaying strong discrepancies. In Fig. 2(a), we show the joint marginalized posterior of the parameters σ_8 and Ω_m that is commonly used [5,6,39,40] to discuss tensions between these kinds of data sets. As we can see, the constraints coming from the two data sets seem consistent at 68% C.L. as in [5,11]. In Fig. 2(b) we show the joint marginalized posterior of two combinations of cosmological parameters, $\sigma_8 \Omega_m^{0.5}$ and $\sigma_8 / h^{0.5}$, that can be used to discuss tensions between data sets [6]. As we can clearly see, the tension between CMB and WL seems enhanced, as they seem not to be consistent at 68%. This is a clear example where marginalizing over a high dimensional non-Gaussian likelihood to get the posterior of some parameters biases the conclusions on the possible tensions between data sets. It is clear, from this study, that the DCT helps in assessing whether discrepancies, over the whole parameter space of a model, are statistically relevant and require further investigation.

The $\mathcal{C}$ values involving CMB lensing (CMBL) data are all weakly pointing toward agreement. CMBL + BAO and CMBL + RSD, in particular, are borderline between

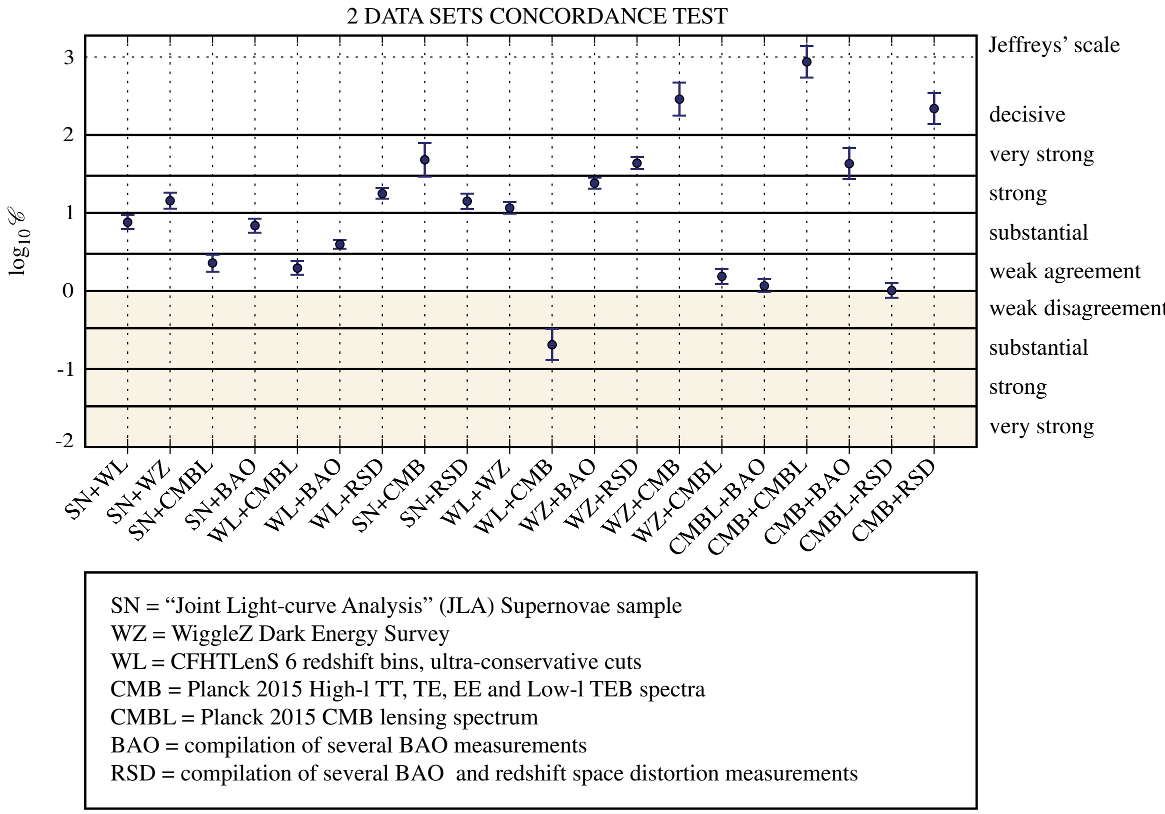

FIG. 1. The data concordance test (DCT) performed on all the independent couples of the cosmological data sets described in the text. The shaded region highlights the values of $\mathcal{C}$ that point toward disagreement between data sets. The error bars represent the uncertainty in the nested sampling computation of the evidence.

agreement and disagreement. For both data sets, this comes from some discrepancies in the determination of the background parameters. RSD data, in addition, are also penalized by some discrepancy in the determination of the amplitude of scalar perturbations. The results of the DCT involving CMB and SN, BAO, RSD and WZ are on the high end of the comparison scale, with values that range from very strong to decisive. This is largely expected as we are combining a probe that is extremely sensitive to all the cosmological parameters (CMB) with other data that probe only a subset of them.

Surprisingly enough, when combining CMB and CMBL the DCT reaches a very high value, the maximum achieved in this comparison. This seems suspicious for two reasons. The first one is the known lensing amplitude tension discussed in [5,12] and that is not found here. The second reason is the fact that CMBL was found in weak agreement with all other data sets while CMB was displaying a good

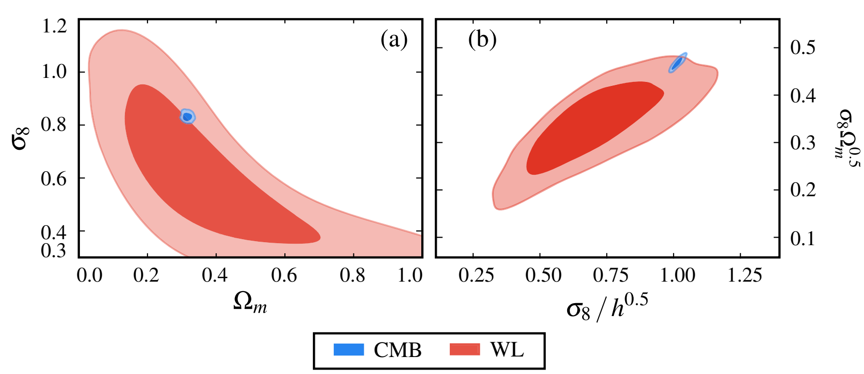

FIG. 2. Panel (a): The marginalized joint posterior for Ω_m and the amplitude of the linear power spectrum on the scale of $8h^{-1}$ Mpc for different data sets, as shown in legend. Panel (b): The marginalized joint posterior for $\sigma_8 \Omega_m^{0.5}$ and $\sigma_8 / h^{0.5}$ for different data sets, as shown in legend. In both panels the darker and lighter shades correspond respectively to the 68% C.L. and the 95% C.L. regions.

agreement. It is beyond the scope of this paper to investigate the possible causes of this behavior. We can, however, have some ideas on its origin from the properties of the DCT discussed before. It seems unlikely that this discrepancy arises because of the unaccounted cross correlation between CMB temperature fluctuations and the CMB lensing spectrum. This was indeed found to result in negligible corrections for Planck-like observations [41]. This leaves the factors involving the likelihood volume and the best fit χ^2, that improve significantly with the introduction of CMB lensing, and, as shown here, are surely worth a deeper investigation. SN data show good agreement with BAO and RSD measurements, from substantial to strong on a Jeffreys' scale, as they agree on the determination of the parameters describing background evolution. The agreement between SN and RSD is slightly higher than BAO as this data set is also sensitive to some perturbations parameters. Agreement between WL and SN, BAO and RSD is also good as the DCT is rewarding the additional leverage on perturbation parameters that comes from WL measurements. For the same reason, a good agreement is found also for WZ and SN, BAO and RSD. Noticeably the values of $\mathcal{C}$ are slightly higher than the previous ones. This reflects the fact that, due to the presence of nonlinear scales in WZ data, the constraints on perturbation parameters are stronger than the previous ones. Testing the combination of WL and WZ data then results in strong agreement. The two data roughly agree on the background parameters and the additional constraining power of WZ on perturbation parameters favors the combination of these two data sets.

In conclusion, we have used Bayesian hypothesis testing to assess quantitatively whether there is concordance, within the ΛCDM model, between several different cosmological experiments. This test, which we dubbed DCT, allows one to compute the odds that two data sets can be described by the same choice of cosmological parameters and thus gives a way of measuring the statistical significance of tensions between different measurements. We have commented on some of the properties that make this test a reliable tool that extends, with statistical rigor, other commonly used approaches. We applied this test to the combinations of some of the most relevant cosmological data sets to date and found, overall, a good agreement between geometrical probes and other perturbations measurements. We showed, however, that the lensing of the CMB is only weakly in agreement with all other cosmological data sets but CMB itself. The odds of this agreement are suspiciously high, given the other results, and require further analysis. At last, we found substantial evidence for a disagreement between WL data of CFHTLenS and CMB measurements of Planck. We showed that a similar conclusion would not be drawn by inspecting the marginal posterior of some parameters. This tension might be a sign of new physics, pointing toward mechanisms that suppress the growth of structures in the late time Universe [11,42,43]. It might also be a signal of the presence of unknown systematic effects [44,45], such as the presence of B-mode signal in weak lensing observations [46]. At last it might point toward the inadequacy of present cosmological predictions in fitting the data. A failure in modeling the evolution of perturbations on nonlinear scales might bias our conclusions resulting in spurious tensions between data sets.

The investigation of the tensions found in this paper, and how they are relieved in extended models, is a primary goal as these could point toward the presence of unaccounted systematic effects, an incomplete modeling of the cosmological predictions or the presence of new physical phenomena.

ACKNOWLEDGMENTS

I am grateful to Carlo Baccigalupi, Nicola Bartolo, Stefano Camera, Noemi Frusciante, Alan Heavens, Bin Hu, Michele Liguori, Matteo Martinelli, Sabino Matarrese, Levon Pogosian, Alessandra Silvestri, Licia Verde and Matteo Viel for useful discussions and comments on the manuscript. I acknowledge partial support from the INFN-INDARK initiative. I thank the Instituut Lorentz (Leiden University) for the computing resources made available for conducting the research reported in this paper.

[1] I. Tereno *et al.* (Euclid Collaboration), Euclid space mission: Building the sky survey, IAU Symp. **306**, 379 (2015).
[2] R. Adam *et al.* (Planck Collaboration), Planck 2015 results. I. Overview of products and scientific results, arXiv: 1502.01582.
[3] C. Heymans *et al.*, CFHTLenS tomographic weak lensing cosmological parameter constraints: Mitigating the impact of intrinsic galaxy alignments, Mon. Not. R. Astron. Soc. **432**, 2433 (2013).
[4] É. Aubourg *et al.*, Cosmological implications of baryon acoustic oscillation (BAO) measurements, Phys. Rev. D **92**, 123516 (2015).
[5] P. A. R. Ade *et al.* (Planck Collaboration), Planck 2015 results. XIII. Cosmological parameters, arXiv: 1502.01589.

[6] T. Abbott *et al.* (DES Collaboration), Cosmology from cosmic shear with DES science verification data, arXiv:1507.05552.

[7] R. A. Battye, T. Charnock, and A. Moss, Tension between the power spectrum of density perturbations measured on large and small scales, Phys. Rev. D **91**, 103508 (2015).

[8] M. Wyman, D. H. Rudd, R. A. Vanderveld, and W. Hu, Neutrinos Help Reconcile Planck Measurements with the Local Universe, Phys. Rev. Lett. **112**, 051302 (2014).

[9] J. Hamann and J. Hasenkamp, A new life for sterile neutrinos: Resolving inconsistencies using hot dark matter, J. Cosmol. Astropart. Phys. 10 (2013) 044.

[10] M. Costanzi, B. Sartoris, M. Viel, and S. Borgani, Neutrino constraints: What large-scale structure and CMB data are telling us, J. Cosmol. Astropart. Phys. 10 (2014) 081.

[11] P. A. R. Ade *et al.* (Planck Collaboration), Planck 2015 results. XIV. Dark energy and modified gravity, arXiv:1502.01590.

[12] B. Hu and M. Raveri, Can modified gravity models reconcile the tension between the CMB anisotropy and lensing maps in Planck-like observations?, Phys. Rev. D **91**, 123515 (2015).

[13] E. Di Valentino, A. Melchiorri, and J. Silk, Beyond six parameters: Extending ΛCDM, Phys. Rev. D **92**, 121302 (2015).

[14] P. Marshall, N. Rajguru, and A. Slosar, Bayesian evidence as a tool for comparing data sets, Phys. Rev. D **73**, 067302 (2006).

[15] F. Feroz, B. C. Allanach, M. Hobson, S. S. AbdusSalam, R. Trotta, and A. M. Weber, Bayesian selection of signμ within mSUGRA in global fits including WMAP5 results, J. High Energy Phys. 10 (2008) 064.

[16] M. C. March, R. Trotta, L. Amendola, and D. Huterer, Robustness to systematics for future dark energy probes, Mon. Not. R. Astron. Soc. **415**, 143 (2011).

[17] L. Amendola, V. Marra, and M. Quartin, Internal robustness: Systematic search for systematic bias in SN Ia data, Mon. Not. R. Astron. Soc. **430**, 1867 (2013).

[18] L. Verde, P. Protopapas, and R. Jimenez, Planck and the local Universe: Quantifying the tension, Phys. Dark Univ. **2**, 166 (2013).

[19] J. Martin, C. Ringeval, R. Trotta, and V. Vennin, Compatibility of Planck and BICEP2 in the light of inflation, Phys. Rev. D **90**, 063501 (2014).

[20] N. V. Karpenka, F. Feroz, and M. P. Hobson, Testing the mutual consistency of different supernovae surveys, Mon. Not. R. Astron. Soc. **449**, 2405 (2015).

[21] M. Betoule *et al.* (SDSS Collaboration), Improved cosmological constraints from a joint analysis of the SDSS-II and SNLS supernova samples, Astron. Astrophys. **568**, A22 (2014).

[22] http://smp.uq.edu.au/wigglez-data.

[23] M. J. Drinkwater *et al.*, The WiggleZ Dark Energy Survey: Survey design and first data release, Mon. Not. R. Astron. Soc. **401**, 1429 (2010).

[24] D. Parkinson *et al.*, The WiggleZ Dark Energy Survey: Final data release and cosmological results, Phys. Rev. D **86**, 103518 (2012).

[25] N. Aghanim *et al.* (Planck Collaboration), Planck 2015 results. XI. CMB power spectra, likelihoods, and robustness of parameters, arXiv:1507.02704.

[26] P. A. R. Ade *et al.* (Planck Collaboration), Planck 2015 results. XV. Gravitational lensing, arXiv:1502.01591.

[27] A. J. Ross, L. Samushia, C. Howlett, W. J. Percival, A. Burden, and M. Manera, The clustering of the SDSS DR7 main galaxy sample I. A 4 percent distance measure at $z = 0.15$, Mon. Not. R. Astron. Soc. **449**, 835 (2015).

[28] L. Anderson *et al.* (BOSS Collaboration), The clustering of galaxies in the SDSS-III baryon oscillation spectroscopic survey: Baryon acoustic oscillations in the data releases 10 and 11 galaxy samples, Mon. Not. R. Astron. Soc. **441**, 24 (2014).

[29] F. Beutler, C. Blake, M. Colless, D. Heath Jones, L. Staveley-Smith, L. Campbell, Q. Parker, W. Saunders, and F. Watson, The 6dF galaxy survey: Baryon acoustic oscillations and the local Hubble constant, Mon. Not. R. Astron. Soc. **416**, 3017 (2011).

[30] F. Beutler *et al.* (BOSS Collaboration), The clustering of galaxies in the SDSS-III baryon oscillation spectroscopic survey: Testing gravity with redshift-space distortions using the power spectrum multipoles, Mon. Not. R. Astron. Soc. **443**, 1065 (2014).

[31] L. Samushia *et al.*, The clustering of galaxies in the SDSS-III baryon oscillation spectroscopic survey: Measuring growth rate and geometry with anisotropic clustering, Mon. Not. R. Astron. Soc. **439**, 3504 (2014).

[32] R. E. Smith, J. A. Peacock, A. Jenkins, S. D. M. White, C. S. Frenk, F. R. Pearce, P. A. Thomas, G. Efstathiou, and H. M. P. Couchman (VIRGO Consortium Collaboration), Stable clustering, the halo model and nonlinear cosmological power spectra, Mon. Not. R. Astron. Soc. **341**, 1311 (2003).

[33] R. Takahashi, M. Sato, T. Nishimichi, A. Taruya, and M. Oguri, Revising the halofit model for the nonlinear matter power spectrum, Astrophys. J. **761**, 152 (2012).

[34] http://camb.info.

[35] A. Lewis, A. Challinor, and A. Lasenby, Efficient computation of CMB anisotropies in closed FRW models, Astrophys. J. **538**, 473 (2000).

[36] A. Lewis and S. Bridle, Cosmological parameters from CMB and other data: A Monte Carlo approach, Phys. Rev. D **66**, 103511 (2002).

[37] W. J. Handley, M. P. Hobson, and A. N. Lasenby, PolyChord: Nested sampling for cosmology, Mon. Not. R. Astron. Soc. **450**, L61 (2015).

[38] W. J. Handley, M. P. Hobson, and A. N. Lasenby, PolyChord: Next-generation nested sampling, Mon. Not. R. Astron. Soc. **450**, L61 (2015).

[39] N. MacCrann, J. Zuntz, S. Bridle, B. Jain, and M. R. Becker, Cosmic discordance: Are Planck CMB and CFHTLenS weak lensing measurements out of tune?, Mon. Not. R. Astron. Soc. **451**, 2877 (2015).

[40] J. Liu and J. C. Hill, Cross correlation of Planck CMB lensing and CFHTLenS galaxy weak lensing maps, Phys. Rev. D **92**, 063517 (2015).

[41] M. M. Schmittfull, A. Challinor, D. Hanson, and A. Lewis, Joint analysis of CMB temperature and lensing-reconstruction power spectra, Phys. Rev. D **88**, 063012 (2013).

[42] R. A. Battye and A. Moss, Evidence for Massive Neutrinos from Cosmic Microwave Background and Lensing Observations, Phys. Rev. Lett. **112**, 051303 (2014).

[43] J. N. Dossett, M. Ishak, D. Parkinson, and T. Davis, Constraints and tensions in testing general relativity from Planck and CFHTLenS data including intrinsic alignment systematics, Phys. Rev. D **92**, 023003 (2015).
[44] J. Liu, A. Ortiz-Vazquez, and J. C. Hill, Constraining multiplicative bias in CFHTLenS weak lensing shear data, arXiv:1601.05720.
[45] S. Joudaki *et al.*, CFHTLenS revisited: Assessing concordance with Planck including astrophysical systematics, arXiv:1601.05786.
[46] M. Asgari, C. Heymans, C. Blake, J. Harnois-Deraps, P. Schneider, and L. Van Waerbeke, Revisiting CFHTLenS cosmic shear: Optimal E/B mode decomposition using COSEBIs and compressed COSEBIs, arXiv:1601.00115.