Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **466,** 369–377 (2017)
Advance Access publication 2016 December 1

doi:10.1093/mnras/stw3102

Constraining the dark energy equation of state using Bayes theorem and the Kullback–Leibler divergence

S. Hee,[1,2]★ J. A. Vázquez,[3] W. J. Handley,[1,2] M. P. Hobson[1] and A. N. Lasenby[1,2]

[1]*Astrophysics Group, Battcock Centre, Cavendish Laboratory, JJ Thomson Avenue, Cambridge CB3 0HE, UK*
[2]*Kavli Institute for Cosmology Cambridge, Madingley Road, Cambridge CB3 0HA, UK*
[3]*Brookhaven National Laboratory, 2 Center Road, Upton, NY 11973, USA*

Accepted 2016 November 28. Received 2016 October 24; in original form 2016 July 1

ABSTRACT

Data-driven model-independent reconstructions of the dark energy equation of state $w(z)$ are presented using *Planck* 2015 era cosmic microwave background, baryonic acoustic oscillations (BAO), Type Ia supernova (SNIa) and Lyman α (Lyα) data. These reconstructions identify the $w(z)$ behaviour supported by the data and show a bifurcation of the equation of state posterior in the range $1.5 < z < 3$. Although the concordance Λ cold dark matter (ΛCDM) model is consistent with the data at all redshifts in one of the bifurcated spaces, in the other, a supernegative equation of state (also known as 'phantom dark energy') is identified within the 1.5σ confidence intervals of the posterior distribution. To identify the power of different data sets in constraining the dark energy equation of state, we use a novel formulation of the Kullback–Leibler divergence. This formalism quantifies the information the data add when moving from priors to posteriors for each possible data set combination. The SNIa and BAO data sets are shown to provide much more constraining power in comparison to the Lyα data sets. Further, SNIa and BAO constrain most strongly around redshift range 0.1–0.5, whilst the Lyα data constrain weakly over a broader range. We do not attribute the supernegative favouring to any particular data set, and note that the ΛCDM model was favoured at more than 2 log-units in Bayes factors over all the models tested despite the weakly preferred $w(z)$ structure in the data.

Key words: equation of state – methods: data analysis – methods: statistical – cosmological parameters – dark energy.

1 INTRODUCTION

The nature of dark energy (DE) remains a significant outstanding problem in cosmology. The Λ cold dark matter (ΛCDM) model considers a constant equation of state (EoS) parameter $w = -1$ motivated by vacuum energy. The most frequent generalization of the ΛCDM DE EoS is to allow an alteration of the time-independent EoS parameter so that $w \neq -1$ (hereafter referred to as wCDM). Allowing w to vary in time $w = w(z)$ results in quintessence DE models. Many quintessence models (Ratra & Peebles 1988; Caldwell, Dave & Steinhardt 1998; Tsujikawa 2013), including phantom DE (Caldwell 2002; Sahni 2005), as well as modified GR theories (Sahni 2005), make predictions for the behaviour of $w(z)$ that may be tested against cosmological data sets (Planck Collaboration XIV 2016). Time-dependent behaviour can also be investigated by choosing equations that are simple or mathematically appealing, to test as a DE model. These phenomenological models include the CPL (Chevallier & Polarski 2001; Linder 2003), JPB (Jassal, Bagla & Padmanabhan 2004) and FNT (Felice, Nesseris & Tsujikawa 2012) models. Lastly, free-form approaches attempt to avoid any commitment to particular equations and instead aim to allow the observational data to define any time-dependent features in $w(z)$ (Huterer & Starkman 2003; Zunckel & Trotta 2007; Zhao, Huterer & Zhang 2008; Serra et al. 2009; Lazkoz, Salzano & Sendra 2012; Vázquez et al. 2012b). Other free-form reconstruction methods include Gaussian processes (Holsclaw et al. 2010a,b; Seikel, Clarkson & Smith 2012). We refer the reader to an older review by Sahni & Starobinsky (2006) that describes the general reconstruction process and new results by Planck Collaboration XIV (2016) for further reading on DE constraints.

In this paper, we use Bayes factors combined with a 'nodal' free-form method, which reconstructs a function using a spline between nodes whose amplitude and position can vary, as first proposed by Vázquez et al. (2012a), to investigate the constraints on $w(z)$. This approach has subsequently been used by Vázquez et al. (2012b), Aslanyan et al. (2014), Planck Collaboration XX (2016) and Hee et al. (2015) and has the benefit of remaining general and

★ E-mail: sh767@cam.ac.uk

focusing on the cosmological data sets rather than a specific model. The first aim of this paper is to investigate potential deviations from the ΛCDM constant DE EoS using Bayesian model selection. The second aim is to analyse the constraining power on $w(z)$ of the data sets using the Kullback–Leibler (KL) divergence ($\mathcal{D}_{\rm KL}$). Observational data are improving in quality with many upcoming missions promising to increase our ability to understand DE models. Assessing the data sets in the manner, this paper proposes provides a robust, quantitative measure of DE information that may easily be compared with past or future missions.

The paper is structured as follows. We first identify the data sets and computational techniques in Section 2. An analysis of $w(z)$ constraints from *Planck* satellite era cosmological data sets is presented in Section 3 and the analysis of these additional data sets using the $\mathcal{D}_{\rm KL}$ approach is presented in Section 4. We conclude in Section 5, considering the findings in relation to ΛCDM and constraints on $w(z)$, and comment on the efficacy of the techniques used for quantifying data set constraining power and information content.

2 DATA SETS AND COMPUTATION

We update the work of Vázquez et al. (2012b) and Hee et al. (2015), where time-dependent behaviour in $w(z)$ within a cold dark matter (CDM) universe is identified using a sequence of nodal reconstructions weighted by their Bayes factors. In addition, we use the KL divergence to analyse information content, expanding on a similar work by Trotta et al. (2008) and Bridges et al. (2009).

2.1 Bayes theorem and model selection

In order to reconstruct the $w(z)$ plane, we perform Bayesian parameter estimation and model comparison (Bayes & Price 1763; MacKay 2003; Sivia & Skilling 2006) on cosmological models to be defined shortly.

Bayesian parameter estimation is the process of determining the posterior probability distribution of a set of parameters θ for a given model $\mathcal{M}$ via Bayes theorem:

$$\Pr(\theta|\mathcal{D},\mathcal{M}) = \frac{\Pr(\mathcal{D}|\theta,\mathcal{M})\Pr(\theta|\mathcal{M})}{\Pr(\mathcal{D}|\mathcal{M})} \equiv \frac{\mathcal{L}\pi}{\mathcal{Z}}. \quad (1)$$

This requires a prior on the model parameters, $\Pr(\theta|\mathcal{M}) = \pi(\theta)$, and a means to calculate the likelihood, $\Pr(\mathcal{D}|\theta,\mathcal{M}) = \mathcal{L}(\theta)$. The evidence $\mathcal{Z}$ (or marginal likelihood) may be computed from the priors and likelihoods via

$$\mathcal{Z} \equiv \Pr(\mathcal{D}|\mathcal{M}) = \int \Pr(\mathcal{D}|\theta,\mathcal{M})\Pr(\theta|\mathcal{M})\, d\theta \equiv \int \mathcal{L}\pi. \quad (2)$$

Our priors are defined in Table 1, whilst the likelihood codes are defined in the references in the data set section below.

Bayesian model comparison uses Bayes theorem to make inferences about how probable a model is in light of the data:

$$P(\mathcal{M}|\mathcal{D}) = \frac{\Pr(\mathcal{D}|\mathcal{M})\Pr(\mathcal{M})}{\Pr(\mathcal{D})}. \quad (3)$$

Taking log-ratios of this equation for two different models yields the posterior odds ratio:

$$\exp(\mathcal{P}_{ij}) \equiv \frac{P(\mathcal{M}_j|\mathcal{D})}{P(\mathcal{M}_i|\mathcal{D})} = \frac{P(\mathcal{D}|\mathcal{M}_j)}{P(\mathcal{D}|\mathcal{M}_i)}\frac{P(\mathcal{M}_j)}{P(\mathcal{M}_i)}. \quad (4)$$

Thus, the critical data-dependent quantity is the Bayes factor:

$$\mathcal{B}_{ij} = \ln(\mathcal{Z}_j/\mathcal{Z}_i), \quad (5)$$

Table 1. The 31 priors that define the parameter space. The top set of parameters are the CDM parameters, the middle ones show the nuisance parameters associated with the data sets and the bottom set are the parameters introduced by the free-form DE model extensions. Planck Collaboration XI (2016) has more details about the CDM and nuisance parameters, whilst the DE extension parameters are defined in the text.

Parameter	Prior range	Prior type
$\Omega_{\rm b}h^2$	[0.019, 0.025]	Uniform
$\Omega_{\rm c}h^2$	[0.095, 0.145]	Uniform
$100\theta_{\rm MC}$	[1.03, 1.05]	Uniform
τ	[0.01, 0.4]	Uniform
$n_{\rm s}$	[0.885, 1.04]	Uniform
$\ln(10^{10}A_{\rm s})$	[2.5, 3.7]	Uniform
$y_{\rm cal}$	[0.9, 1.1]	Uniform
$\alpha_{\rm JLA}$	[0.01, 2.00]	Uniform
$\beta_{\rm JLA}$	[0.9, 4.6]	Uniform
$A_{217}^{\rm CIB}$	[0, 200]	Uniform
$\xi^{\rm tSZ-CIB}$	[0, 1]	Uniform
$A_{143}^{\rm tSZ}$	[0, 10]	Uniform
$A_{100}^{\rm PS}$	[0, 400]	Uniform
$A_{143}^{\rm PS}$	[0, 400]	Uniform
$A_{143\times217}^{\rm PS}$	[0, 400]	Uniform
$A_{217}^{\rm PS}$	[0, 400]	Uniform
$A^{\rm kSZ}$	[0, 10]	Uniform
$A_{100}^{\rm dustTT}$	[0, 50]	Uniform
$A_{143}^{\rm dustTT}$	[0, 50]	Uniform
$A_{143\times217}^{\rm dustTT}$	[0, 100]	Uniform
$A_{217}^{\rm dustTT}$	[0, 400]	Uniform
c_{100}	[0, 30]	Uniform
c_{217}	[0, 30]	Uniform
$w(z_i)\vert_{i=1\ldots5}$	$[-2, -0.01]$	Uniform
$z_i\vert_{i=2\ldots4}$	[0.01, 3.0]	Sorted-uniform

Table 2. Jeffreys guideline for interpreting posterior odds ratios. As $\mathcal{P}_{ji} = -\mathcal{P}_{ij}$, negative values imply model favouring is reversed.

Posterior odds ratio	Favouring of $\mathcal{M}_j$ over $\mathcal{M}_i$
$0.0 \le \mathcal{P}_{ij} \le 1.0$	None
$1.0 \le \mathcal{P}_{ij} \le 2.5$	Slight
$2.5 \le \mathcal{P}_{ij} \le 5.0$	Significant
$5.0 \le \mathcal{P}_{ij}$	Decisive

where $\Pr(\mathcal{D}|\mathcal{M}_i) = \mathcal{Z}_i$ is the evidence of model i. We quantify a model's favouring using the Jeffreys (1961) guideline defined in Table 2. We use Bayes factors and posterior odds ratios interchangeably, as we assume uniform model priors $P(\mathcal{M}_j) = P(\mathcal{M}_i)$ throughout.

2.2 Data sets

In order to investigate possible time-dependent behaviour in the DE EoS, we use likelihood codes from *Planck* cosmic microwave background (CMB) measurements, baryonic acoustic oscillations (BAO), Type Ia supernovae (SNIa) and Lyman α (Lyα) BAO data. Specifically, for the CMB data we use the low-l TEB and high-l TT likelihoods from the *Planck* satellite 2015 data release (Planck

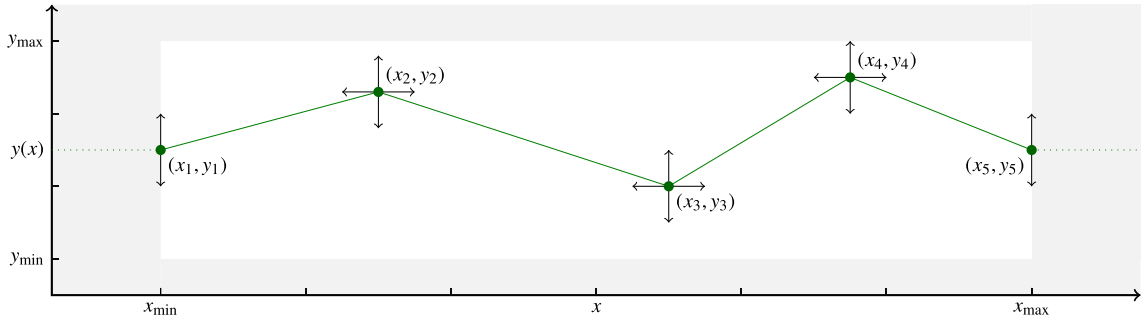

Figure 1. Piecewise linear interpolation function. We place n internal nodes (x_i, y_i) in the rectangle bounded by (x_{min}, y_{min}) and (x_{max}, y_{max}), where the positions x_i and amplitudes y_i are model parameters to be varied. At x_{min} and x_{max} fixed-position nodes are placed with varying amplitude only, such that for the model defined by n internal nodes, there are $2 + 2n$ parameters. Linear interpolation between the nodes (x_i, y_i) is used to construct y at all points, with $y(x)$ set constant outside the range $[x_{min}, x_{max}]$.

Collaboration I 2016; Planck Collaboration XI 2016; Planck Collaboration XIII 2016), which we will refer to as *Planck*. For the BAO data, we use the Baryon Oscillation Spectroscopic Survey (BOSS) Data Release 11 likelihoods (Anderson et al. 2014), or *BAO*. For the SNIa data, we use the JLA supernovae catalogue likelihoods (Betoule et al. 2014), *JLA* for short. For the Lyα data, we use the likelihood codes described by Font-Ribera et al. (2014) ($A_{Ly\alpha}$; BOSS autocorrelation) and Delubac et al. (2015) ($B_{Ly\alpha}$; BOSS cross-correlation). For a good summary of the BAO data, see Aubourg et al. (2015). Using the above notation, the whole data set combination can be referred to as $Planck + BAO + JLA + A_{Ly\alpha} + B_{Ly\alpha}$.

2.3 Computational tools

To carry out Bayesian inference, we use CosmoMC (Lewis & Bridle 2002) containing the Boltzmann CAMB code (Lewis, Challinor & Lasenby 2000; Howlett et al. 2012). We substitute the default Metropolis–Hastings sampler with the PolyChord nested sampling plug-in (Handley, Hobson & Lasenby 2015a,b), an effective nested sampling implementation (Skilling 2004, 2006; Sivia & Skilling 2006) for evidence calculations and parameter estimation with proven efficacy using *Planck* era data (Planck Collaboration XX 2016). Aside from the Lyα data sets, all data sets used are default CosmoMC options. To facilitate deviations from the standard ΛCDM EoS parameter $w = -1$, we implement the parametrized post-Friedmann framework (PPF) modification to CAMB (Fang, Hu & Lewis 2008), which has sound speed equal to c and no scalar anisotropic stress. The free-form reconstruction we use is the nodal reconstruction as proposed by Vázquez et al. (2012b) and successfully used in several cosmological applications to date (Vázquez et al. 2012a,b; Aslanyan et al. 2014; Hee et al. 2015; Planck Collaboration XX 2016).

2.4 Nodal reconstruction

We model a one-dimensional function $y(x)$ using a piecewise linear interpolation between a set of n nodes (Fig. 1), where the positions of the nodes are model parameters to be varied. Alternative interpolation schemes may be used, for example, the cubic spline studied by Vázquez et al. (2012a), although we do not consider these here since the continuity requirements of the interpolation functions and its derivatives limit its ability to model sharply changing functions $y(x)$.

A model is defined by how many nodes are used in reconstructing $y(x)$. We use Bayes factors to compare models with increasing numbers of nodes, which quantify how many nodes are needed to fit the data.

Table 3. The six models we consider. Priors on each w parameter are uniform on the range $[-2, 0]$, and were chosen to be conservative (Vázquez et al. 2012b). Priors on each z parameter are uniform on $[0, 3]$ and sorted, such that for more than one internal node, we have $z_i < z_{i+1}$ (i.e. sorted uniform priors).

Model name	Description
ΛCDM	$w = -1$
wCDM	w constant in z, but allowed to vary
tCDM	tilted spectrum: two fixed-position nodes at $z = 0, 3$
1CDM	One internal node
2CDM	Two internal nodes
3CDM	Three internal nodes

Further, as each posterior sample defines a function in $y(x)$, we can calculate the posterior probability of y in normalized slices of constant x, $\Pr(y|x, \mathcal{D}, \mathcal{M})$, to obtain the plane reconstruction of a model. We plot these as a function of σ confidence intervals to show the statistical significance of deviations from the maximal y at each x. One can plot $\Pr(y|x, n_\star)$, where $n_\star$ denotes the number of nodes in the most favoured model. In order to identify the nature of constraints from various models, one should also plot $\Pr(y|x)$ averaged over all models weighted by their posterior odds ratios (Parkinson & Liddle 2013; Hee et al. 2015; Planck Collaboration XX 2016).

A key strength of this reconstruction procedure is its free-form nature, which can capture any shape of function in the $y(x)$ plane by adding arbitrarily large numbers of nodes. The Bayes factor penalizes overcomplex models, identifying how much complexity the data are able to support. Model selection techniques can thus be used to solve questions on the constraining power of the data in cosmological applications (Vázquez et al. 2012a,b; Aslanyan et al. 2014; Hee et al. 2015; Planck Collaboration XX 2016).

We apply this reconstruction to $w(z)$. The models we consider, along with their priors are detailed in Table 3. The previous work using *Wilkinson Microwave Anisotropy Probe* (*WMAP*) satellite era data by Vázquez et al. (2012b) found that ΛCDM was favoured, whilst 2CDM had the second largest evidence, pointing to structure in $w(z)$ that could not be captured by a constant EoS parameter wCDM, or even the 1 internal node model. Subsequent work with *Planck* 2013 era data by Hee et al. (2015) showed that ΛCDM was again favoured, and that each model of increasing complexity was more disfavoured than the last. We now investigate this more fully

with *Planck* 2015 era data sets, the addition of Lyα data and further data set analysis tools.

2.5 Kullback–Leibler divergence and data set analysis

We expand on the model selection complexity analysis through the use of the KL divergence. The KL divergence of P from Q is defined as

$$\mathcal{D}_{\rm KL}(P||Q) \equiv \int_{-\infty}^{\infty} p(x) \ln\left[\frac{p(x)}{q(x)}\right] {\rm d}x = \int \ln\left[\frac{{\rm d}P}{{\rm d}Q}\right] {\rm d}P, \qquad (6)$$

where $p(x)$ and $q(x)$ are the probability density functions of probability distributions P and Q. Evaluating the KL divergence (6) of a posterior distribution from its prior provides a measure of information gained from the data (Kullback & Leibler 1951; Trotta et al. 2008; Bridges et al. 2009; Seehars et al. 2014, 2016; Grandis et al. 2016).

We wish to restrict our analysis to the constraining power of the data sets on $w(z)$, and not the other cosmological and nuisance parameters as a whole. First, we can calculate the KL divergence of the marginalized posterior $\Pr(w|z)$ from the marginalized prior $\pi(w|z)$ for w at each z to obtain a function:

$$\mathcal{D}_{\rm KL}(z) = \int \Pr(w|z) \ln\left[\frac{\Pr(w|z)}{\pi(w|z)}\right] {\rm d}w \qquad (7)$$

which defines the gain in information on w at each z. Second, we calculate the $\mathcal{D}_{\rm KL}$ for the whole plane by using the function $\Pr(w, z)$ and its prior, which can be written as

$$\mathcal{D}_{\rm KL} = \int \mathcal{D}_{\rm KL}(z) \Pr(z) {\rm d}z, \qquad (8)$$

where $\Pr(z)$ is flat (as z is not constrained by the analysis given that every posterior sample for a nodal reconstruction passes through every point in z). Note that it is also possible to integrate over da or dlog(a) to compress higher redshifts, however dz is more natural here given how we have defined our reconstruction. Together, the two values allow us to analyse the gain in information due to different data sets using $\mathcal{D}_{\rm KL}$ as well as to understand where each data set provides the greatest gains in information using $\mathcal{D}_{\rm KL}(z)$. We obtain the posterior plane reconstructions from PolyChord and the prior distributions based on $\pi(z_i)$ and $\pi(w_i)$ together with the physical restrictions imposed by CosmoMC.

Typically, a gain in information can occur for two reasons: either due to an increase in parameter constraints, or due to a shift in the position of the peak from prior to posterior (Trotta et al. 2008; Seehars et al. 2014, 2016; Grandis et al. 2016). It is not yet possible to differentiate between the two cases for non-Gaussian distributions. In order to identify the constraining power of the data, we supplement our analysis by calculating the $\mathcal{D}_{\rm KL}$ and $\mathcal{D}_{\rm KL}(z)$ when moving from a completely flat prior on $w(z)$ to the posterior. As there is no peak to shift from for a flat posterior, this measure only identifies how tightly constrained the posterior is, due both to the priors and data. In cases where the CosmoMC prior divergences are larger than those from the flat prior, we can deduce that a significant shift is present.

3 RESULTS: DARK ENERGY EQUATION OF STATE RECONSTRUCTION

The columns in Fig. 2 show from left to right the prior, posterior and marginalized 1D and 2D posteriors for the $w(z)$ plane reconstructions alongside the Bayes factors for the five model extensions compared to ΛCDM. ΛCDM is the favoured model in the Bayesian model selection analysis. wCDM is disfavoured by more than 2 log-units, a slight disfavouring on the Jeffreys scale, whilst all other models are significantly disfavoured at beyond 2.5 log-units. We conclude that the additional flexibility in capturing $w(z)$ features provided by additional parameters does not produce favourable Bayes factors. This is consistent with previous results obtained with *Planck* (Hee et al. 2015). The systematic dropping off in Bayes factors for models with increasing numbers of parameters used for defining $w(z)$ suggests that there is not sufficient time dependence in the true EoS function to overcome the Occam factor associated with the additional parameters (MacKay 2003). Specifically, one can estimate the evidence integral using a Laplace approximation to obtain an Occam factor given by $\sigma_{\theta|\mathcal{D}}/\sigma_\theta$ (MacKay 2003, p. 349; Hee et al. 2015), where $\sigma_{\theta|\mathcal{D}}$ is the width around the peak of the posterior and σ_θ is the same for the prior, and use this to determine the size of the Occam factor between models. When moving from 2CDM to 3CDM we obtain an Occam factor of approximately 0.72, where we assume the posterior on the additional nodal position parameter is equal to the prior, as there is little additional structural information, and have taken the average full width half-max value of the five 3CDM amplitude parameters to estimate the effect of adding the additional node (the prior is flat so σ_θ is the width, 2). This shows that the observed Bayes factor drop of 0.54 (with errors on order 0.29) is comparable to the Occam factor and therefore the information gained from the additional node, which should compensate the effect, is small.

The plane reconstructions show clear constraining power compared to the priors. In all cases that allow for time-dependence, there is the suggestion that a supernegative EoS fits the data best at higher redshifts. Specifically, the tCDM model deviates from ΛCDM by 1σ already before $z = 1$, whilst the models with internal nodes, which are able to identify more flexibly where deviations occur, suggest a 1σ deviation around $z = 1.5$. No model deviates at 2σ, however. It should also be noted that the tightest constraints on the EoS are around redshift $z = 0.1$–0.5, and all models tend to ΛCDM in this region. This suggests that conclusions are still data-limited but that time-dependent behaviour of a supernegative EoS is hinted at by the combinations of $Planck + BAO + JLA + {\rm Ly}\,\alpha$.

We can look at the marginalized posteriors of nodes and amplitude parameters to gather further insights. Interestingly, the 1D marginalized posteriors on the $w(z = 0)$ parameters of the models seem to favour $w > -1$, whilst the wCDM model does not specifically as the single amplitude parameter has simultaneously to model the late time behaviour. This suggests that using wCDM simplifies the DE problem in a way that can obscure underlying DE physics. Given that the difference in Bayes factors between wCDM and any of the more flexible models is indistinguishable on the Jeffreys scale, using a more flexible model is statistically valid and therefore advisable if wishing to analyse $w(z)$.

Looking at the 2D marginalized node positions in the $w(z)$ plane, it is clear that in all cases, the lowest redshift node is well defined as agreeing with ΛCDM. In 1CDM, where there is only one internal node, the plane reconstruction takes a very similar form to tCDM as a result. For the 2CDM and 3CDM models, the additional nodes then have considerable freedom and the plane reconstruction shape at higher redshifts reflects this via a more constant value of w from about redshift 2 onwards. The last node for both the 2CDM and 3CDM models is largely consistent with ΛCDM as the amplitude is poorly constrained beyond $z = 2$, whilst in the range $1.5 < z < 2.0$, it deviates by 1σ, as consistent with the plane reconstructions. Generally, we conclude that all the amplitude parameters are in good agreement with ΛCDM, which is why the additional parameters do

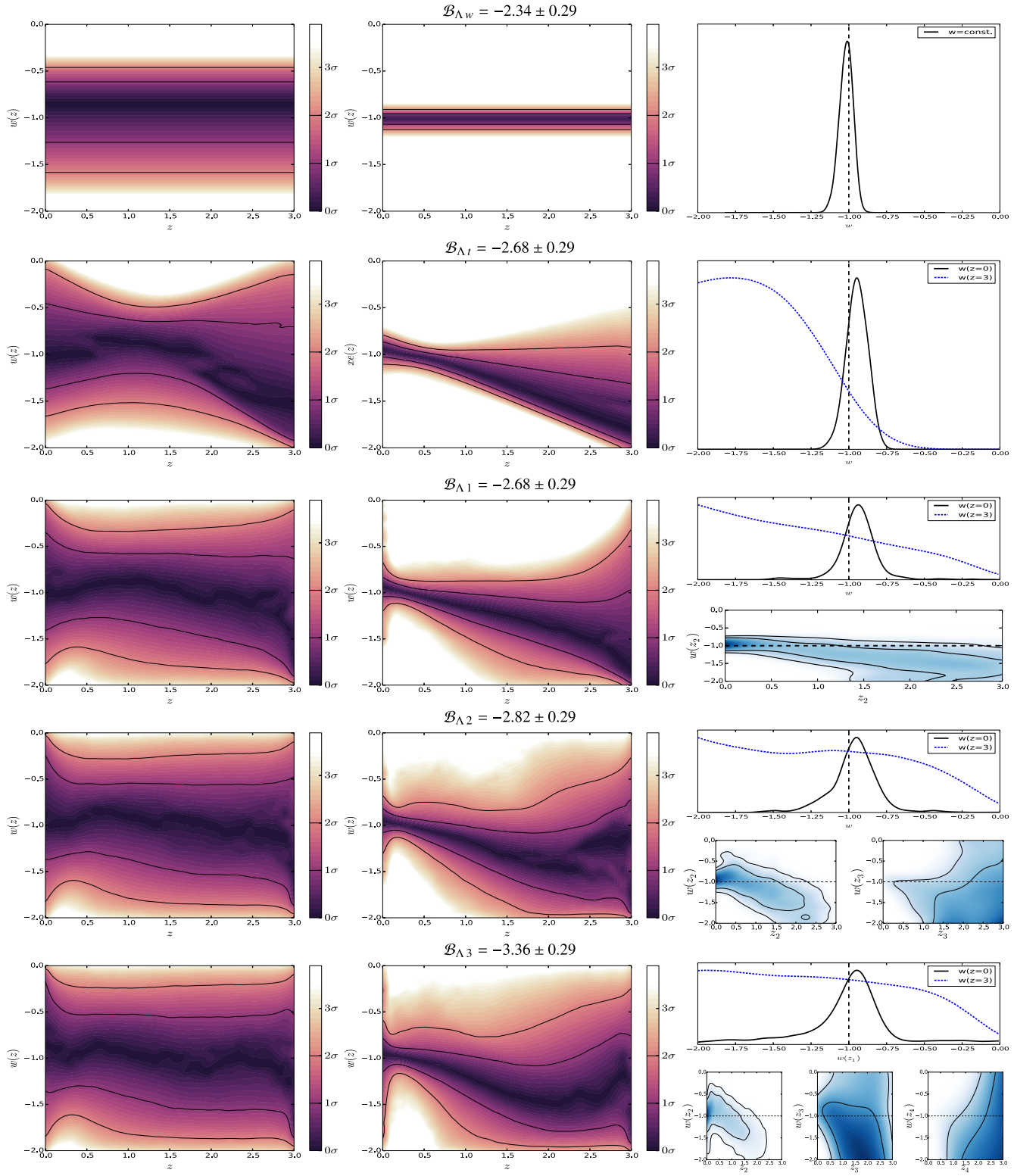

Figure 2. The $w(z)$ priors, $w(z)$ reconstructions and parameter constraints for each of the five model extensions beyond ΛCDM. The leftmost plots are the prior space on the function $w(z)$ as a result of our uniform nodal reconstruction parameters and CosmoMC's sampling, and the central plots show the constraints on $w(z)$ as a result of the data. These plots show the posterior probability $\Pr(w|z)$: the probability of w as normalized in each slice of constant z, with colour scale in confidence interval values. The 1σ and 2σ confidence intervals are plotted as black lines. Note that the σ-deviations are plotted assuming a central value such that a flat prior would not have a uniform colour, thus interpreting regions of the posterior space that are highly unconstrained is more difficult, such as when interpreting the lower bounds of w at high redshifts. Reviewing priors, we see a slight favouring in $w(z)$ of the central values, as expected when calculating priors analytically and given that CosmoMC restricts the permissible parameter space. The posteriors show that the data constrain $w(z)$ strongly compared to our priors. Rightmost are the 1D and 2D marginalized posteriors of the additional model parameters. Marginalized plots were produced using GetDist and $w(z)$ reconstructions were produced in python with the cubehelix colour scheme by Green (2011) for linearity in grey-scale.

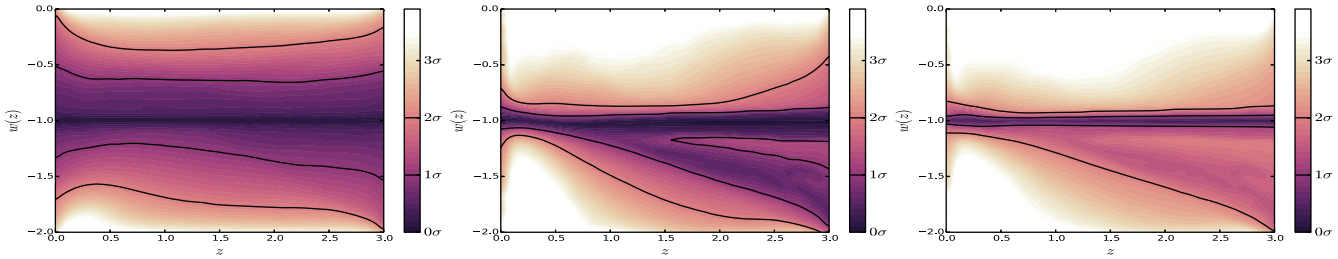

Figure 3. Summarizing the DE model extension results for the constraints on the $w(z)$ plane. All models are weighted by their evidences to give a model averaged plane reconstruction (Parkinson & Liddle 2013; Hee et al. 2015; Planck Collaboration XX 2016), and plotted as in Fig. 2. The three plots show the prior space (left) contracting down to the posterior odds ratio averaged $w(z)$ plane reconstruction for all of the model extensions beyond ΛCDM (middle) and for all of the models including ΛCDM (right). For the model extension averaged reconstruction, it is clear that there is one solution around $w = -1$ and another favouring a supernegative EoS. When including ΛCDM, the significance of the supernegative solution wanes due to the associated large Bayes factor for the $w = -1$ equation.

not add sufficient constraining power to generate Bayes factors that favour the models over ΛCDM.

Reviewing the model averaged plane reconstructions shown in Fig. 3, we observe the conclusions noted above quite clearly in the bifurcation of probabilities on $w(z)$. In the central plot averaging over all models that allow for deviation from ΛCDM, a supernegative solution creates a second peak in the posterior of w for $z > 1.5$. As the reconstruction colour represents posteriors on w in constant slices of z measured by σ confidence intervals with respect to the maximum, the dual peak structure defined by the 1σ contour suggests that the data are sufficiently powerful to resolve a distinct supernegative solution. This supernegative structure is well within the 1σ confidence intervals of the posterior distribution, fitting the data well, whilst $w(z) = -1$ creates the peak probability that defines the 0σ confidence interval. When including ΛCDM in the model averaging, to produce the right-hand plot, again the statistical significance and consistent identification of deviations away from ΛCDM in the reconstructions identifies the alternative supernegative EoS structure. However, the significant Bayes factor favouring of ΛCDM ensures that the functional reconstruction heavily favours $w = -1$ for all redshifts. When including ΛCDM in the model averaging, we conclude that a supernegative EoS fits the observed data at best to within the 1.5σ confidence interval. It should be noted that the model averaging has been done over four models with very similar features identified, which no doubt adds to the strength of the bifurcation when averaging.

4 RESULTS: KULLBACK–LEIBLER DIVERGENCE AND DATA SET ANALYSIS

To understand how the various data sets constrain the $w(z)$ EoS, we analyse every combination of the data sets using the 2CDM model and the KL divergence ($\mathcal{D}_{\rm KL}$). We chose the 2CDM model for its flexibility to capture features whilst not being as computationally demanding as 3CDM. For each combination, we present the $w(z)$ plane reconstruction to identify features visually, the single value $\mathcal{D}_{\rm KL}$ to understand the total information gained and data set constraining power and the distribution $\mathcal{D}_{\rm KL}(z)$ to localize these effects as a function of redshift. As discussed in Section 2.5, the $\mathcal{D}_{\rm KL}$ values and $\mathcal{D}_{\rm KL}(z)$ functions are presented for each data set using both the CosmoMC priors to calculate the $\mathcal{D}_{\rm KL}$, which reflect the data set information content when updating our knowledge from prior to posterior, and also using a flat prior when calculating the $\mathcal{D}_{\rm KL}$ to quantify only the strength of the posterior distribution constraints.

Fig. 4 shows the plane reconstructions and plane $\mathcal{D}_{\rm KL}$ for each data set combination in a grid of Lyα versus non-Lyα data sets. The $\mathcal{D}_{\rm KL}$ values in brackets show the constraining power only, whilst the $\mathcal{D}_{\rm KL}$ values not in brackets show the information content of the data sets. Note that for the top row, containing only *Planck* and Lyα data set combinations, the information content is larger than the constraining power. As discussed in Section 2.5, this happens when the posterior peak shifts from the prior peak, and the $\mathcal{D}_{\rm KL}$ analysis therefore is consistent with the observed posterior reconstruction shift to a supernegative EoS for these four data set combinations. From reviewing the constraining power in plane reconstructions along each row (where the combinations vary in use of Lyα data sets), it appears that Lyα data sets do not strongly affect the constraining power despite their large information content. Comparing *P* with *Pab*, we observe an increase in constraining power of 0.14 nats only. When reviewing plots along the columns (where the combinations vary in use of *BAO* and *JLA*), we visually notice a more pronounced constraint on $w(z)$ and an increase of 0.49 nats when comparing *P* with *PBS*. In general, comparing *PBS* and *PabBS*, on either measure of information content or constraining power, shows an increase of 0.1 nats, which suggests that the Lyα data sets can complement the analysis even if not significantly changing the constraining power.

Fig. 5 shows the $\mathcal{D}_{\rm KL}$ as a function of redshift, calculated again using both the CosmoMC priors (solid lines; information content) and flat priors (dashed lines; constraining power). Comparing *PB* and *PS*, we observe that the peak information content of the *BAO* data set is significantly smaller than the *JLA* peak information content. Specifically, the *PB* data set has a peak of 0.6 nats at $z = 0.3$ which is of lower magnitude but later redshift than the *PS* peak of 1.2 nats around $z = 0.2$. The constraining power functions show that this information content is largely due to tightening posterior constraints, and we conclude that the *JLA* data set is more powerful in constraining the DE EoS than *BAO*. Reviewing the $\mathcal{D}_{\rm KL}(z)$ information content for the Lyα data set, combination *Pab* shows a large and broad peak of almost 2 nats at redshift 0.4, suggesting that the Lyα data set contains significantly more information than both the *BAO* or *JLA* data sets. However, this is due to a shift, and the constraining power has a significantly lower peak of only 0.6 nats but over a broad redshift range.

When analysing which data sets may primarily support deviations from ΛCDM, it is interesting to note that the addition of the Lyα data sets pushes the high-redshift constraints away from $w = -1$ further towards the supernegative. The *PB* combination plane reconstruction shows that $w = -1$ is on the 1σ contour over the range $1.5 < z < 2.0$, whilst the *PabB* combination disfavours $w = -1$ at more than 1σ for $z > 1.5$. This is similar for the *PS* and *PabS* comparison. Generally though, the plane reconstructions of most combinations either favour or approach a supernegative

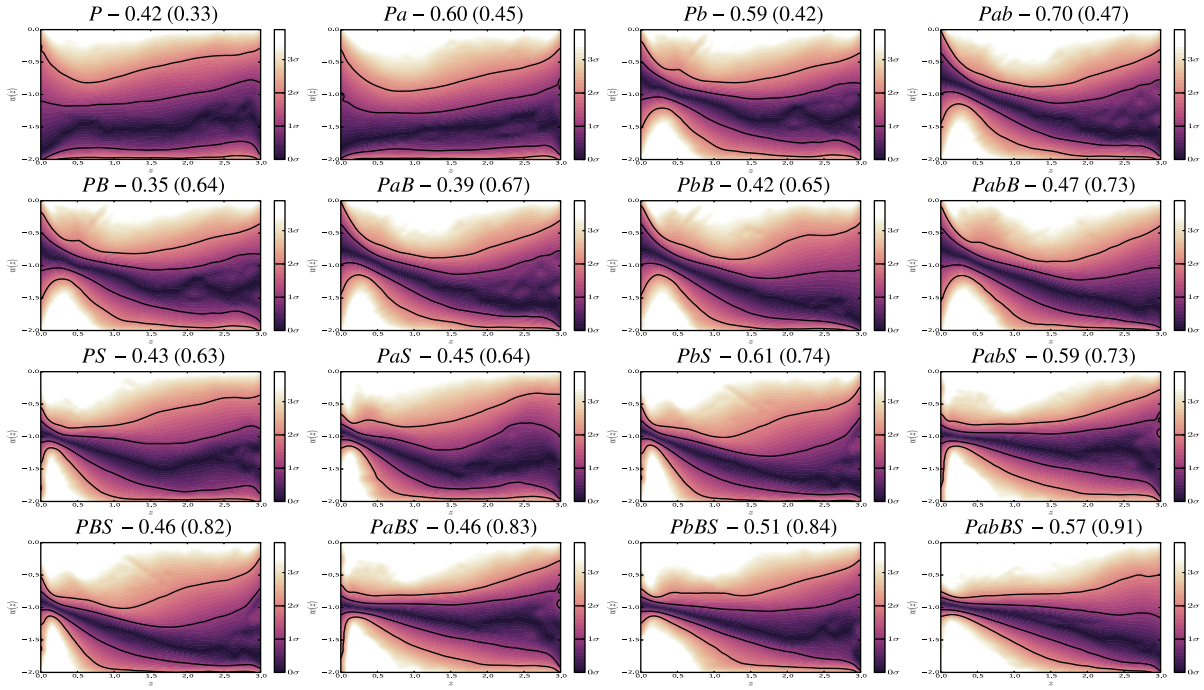

Figure 4. Plane reconstructions of $w(z)$ using the 2CDM model for *Planck* data with each possible combination of the $A_{\mathrm{Ly}\,\alpha}$, $B_{\mathrm{Ly}\,\alpha}$, *BAO* and *JLA* data sets (abbreviated to *P*, *a*, *b*, *B* and *S*, respectively). Results are laid out in a grid with columns of Lyα combinations (without any, with *a*, with *b* and with both) against rows of *BAO* and *JLA* combinations (without either *B*, *S* and both). $\mathcal{D}_{\mathrm{KL}}$ values for the $w(z)$ plane reconstructions, from 2CDM prior to each given posterior, are stated next to each data set combination to quantify the information gained when moving from prior to posterior due to the data. In brackets are the $\mathcal{D}_{\mathrm{KL}}$ values when moving from a flat reconstruction to the posterior, which capture the overall constraining of the posterior whilst ignore any shifts between prior and posterior peaks. Reviewing each row from left to right shows that the Lyα data sets add only some constraining power, whilst reviewing each column from top to bottom shows that *BAO* and *JLA* data sets are both numerically and graphically significant.

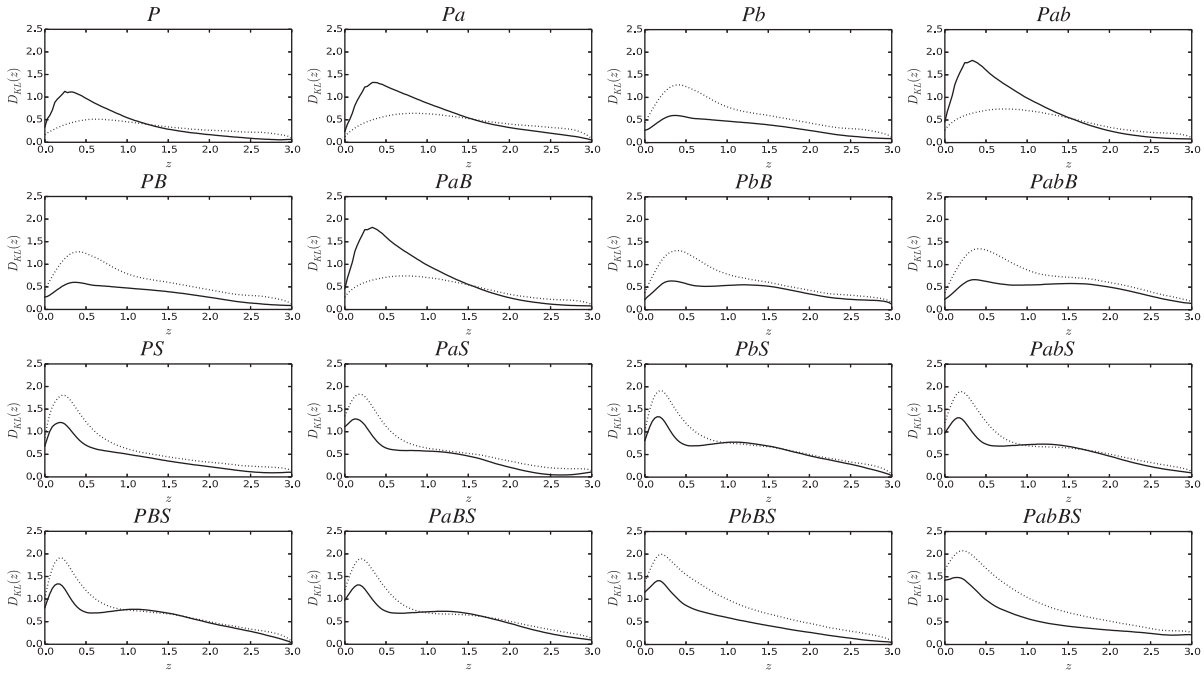

Figure 5. $\mathcal{D}_{\mathrm{KL}}(z)$ for all combinations of data sets, laid out as in Fig. 4, quantifying the constraining power observed qualitatively in the plane reconstructions. The solid lines use the CosmoMC priors when computing $\mathcal{D}_{\mathrm{KL}}(z)$ and demonstrate the additional information gained by using the data in updating our knowledge from the CosmoMC priors to the posteriors. The dashed lines use flat priors across the $w(z)$ plane when calculating $\mathcal{D}_{\mathrm{KL}}(z)$ and quantify more intuitively how constrained the plane appears visually, without including the effect of the posterior shifting from the CosmoMC prior peaks. Using the CosmoMC priors shows that the Lyα data sets add much information due to this shift, whilst the posteriors themselves are less tightly constrained than when using *BAO* and *JLA* data.

$w(z)$ for $z > 1.5$ at a 1σ level even without the Lyα data sets and the constraints often broaden out for $z > 2$ to be consistent with ΛCDM due to a lack of data (as can be observed by the trailing off in the $\mathcal{D}_{\rm KL}(z)$ plots at higher redshift). Therefore, we do not attribute supernegative behaviour strongly to any single data set when combining them. Another deviation from $w = -1$ can be observed in the combination *PaB* at low redshift, where this time $w > -1$ is favoured. Generally, the *BAO* data set seems to favour a less negative EoS for $z < 0.5$, whilst *JLA* is consistent with $w = -1$ at the same period and the Lyα data sets favour a supernegative w-value at all redshifts (which *Planck* does too).

Generally, from the dashed $\mathcal{D}_{\rm KL}$ plots, we conclude that for the Lyα data sets, a broad but small peak in $\mathcal{D}_{\rm KL}(z)$ at around $z = 1$ can be observed to complement the *BAO* and *JLA* data sets (when comparing *PB* with *PabB*, *PS* with *PabS* and *PBS* with *PabBS*) by increasing $\mathcal{D}_{\rm KL}(z)$ for $z > 1.5$.[1] Comparing the *PabBS* plane reconstruction figure (or any data set combination) with the corresponding $\mathcal{D}_{\rm KL}(z)$ plot shows good agreement with the qualitative conclusion that the data sets provide the most constraining power at redshift 0.1–0.5, and now provide a clear quantification of this effect together with a more precise conclusion: the constraining power for the *PabBS* data set and CosmoMC prior combination peaks at redshift 0.25 at 2.1 nats, whilst the data set maximizes information gain at redshift 0.2 with 1.5 nats.

5 CONCLUSIONS

We have presented a detailed Bayesian model selection analysis applied to the nodal reconstruction of $w(z)$, concluding that the Bayes factors on the Jeffreys scale 'slightly favour' ΛCDM when compared to wCDM and 'significantly disfavour' the tCDM, 1CDM, 2CDM and 3CDM models, with an error on the Bayes factors of around 0.29. Despite this favouring, a model averaging approach presents a bifurcation of the $P(w|z)$ plane reconstruction space which shows that, whilst $w = -1$ for all redshift is strongly favoured, a supernegative $w(z)$ EoS at redshift $z > 1.5$ within the 1.5σ confidence intervals of the posterior on $w(z)$ is supported by the data.

To understand this possible deviation, we analysed the constraining power of the data sets using the KL divergence ($\mathcal{D}_{\rm KL}$). We calculated a novel function $\mathcal{D}_{\rm KL}(z)$ to analyse the information gained when moving from the prior distribution of $w(z)$ to the posterior distribution, in slices of constant z, as well as a single $\mathcal{D}_{\rm KL}$ value for the whole plane. For each we used both CosmoMC priors and flat priors to observe information gain due to the data and the overall constraining power, respectively, and we analysed each permutation of data sets using the 2CDM model. We observed that the *BAO* and *JLA* data sets constrained the $w(z)$ plane much more strongly than the Lyα data sets used. These two data sets had a strong peak at redshifts <0.5, whilst the Lyα data sets peaked more broadly at $z = 1$. As expected, the combination of all data sets had the greatest constraining power, specifically the *Planck* data set alone had $\mathcal{D}_{\rm KL} = 0.33$ nats, the combination with *BAO* and *JLA* data sets had $\mathcal{D}_{\rm KL} = 0.82$ nats and the combination $Planck + BAO + JLA + \mathrm{Ly}\,\alpha$ had $\mathcal{D}_{\rm KL} = 0.91$ nats. The same data set combination had a maximum information gain at redshift 0.2 of 1.5 nats. Reviewing the plane reconstructions and $\mathcal{D}_{\rm KL}(z)$ functions showed that the Lyα data sets provided additional constraints at $z > 1.5$ that favours a supernegative EoS, with ΛCDM disfavoured at 1σ significance.

Generally, many of the data set combinations disfavoured ΛCDM at 1σ significance around $1.5 < z < 2$, with higher redshifts being too poorly constrained to draw conclusions. For redshifts below 1.5, the Lyα data sets favoured a supernegative $w(z)$, the *JLA* data set typically agrees with ΛCDM and the *BAO* data set tends towards $w > -1$ values (around 1σ significance at $z = 0.25$). Concluding on the higher redshift deviations, we do not attribute this supernegative favouring to a particular data set, but note that the inclusion of Lyα data adds prominence as it provides a small amount of much needed constraining power over that range.

In the future, the conclusions of an analysis with these techniques will strengthen as data quality improves. The nodal reconstruction has again been shown to be useful in constraining cosmological models and developing a model-independent data-driven analysis (Vázquez et al. 2012a,b; Aslanyan et al. 2014; Hee et al. 2015; Planck Collaboration XX 2016). In addition, the novel formalism introduced here of the KL divergence as a function of redshift provides a quantitative analysis of data set information content applied to specific cosmological problems. Future applications of this method with upcoming mission and survey data or for forecasting with mock data will provide useful insights into the value of data sets in constraining our cosmological models.

[1] Note that taking the difference of two $\mathcal{D}_{\rm KL}(z)$ graphs does not represent the information gained or lost between combinations, but the observed change in shape is what we are commenting on. The addition of *ab* raises $\mathcal{D}_{\rm KL}(1.5 < z < 2)$ slightly and tightens the plane reconstruction contours for higher redshifts.

ACKNOWLEDGEMENTS

This work was performed using the Darwin Supercomputer of the University of Cambridge High Performance Computing Service (http://www.hpc.cam.ac.uk), provided by Dell Inc. using Strategic Research Infrastructure Funding from the Higher Education Funding Council for England and funding from the Science and Technology Facilities Council (STFC). Parts of this work were undertaken on the COSMOS Shared Memory system at DAMTP, University of Cambridge operated on behalf of the STFC DiRAC HPC Facility; this equipment is funded by BIS National E-infrastructure capital grant ST/J005673/1 and STFC grants ST/H008586/1, ST/K00333X/1. SH and WJH thank STFC for financial support. We wish to thank the reviewer for their insightful additions.

REFERENCES

Anderson L. et al., 2014, MNRAS, 441, 24
Aslanyan G., Price L. C., Abazajian K. N., Easther R., 2014, J. Cosmol. Astropart. Phys., 8, 52
Aubourg É. et al., 2015, Phys. Rev. D, 92, 123516
Bayes T., Price R., 1763, An Essay towards Solving a Problem in the Doctrine of Chances. By the Late Reverend Thomas Bayes, F. R. S. Communicated by Richard Price, in a Letter to John Canton, A. M. F. R. S.
Betoule M. et al., 2014, A&A, 568, A22
Bridges M., Feroz F., Hobson M. P., Lasenby A. N., 2009, MNRAS, 400, 1075
Caldwell R., 2002, Phys. Lett. B, 545, 23
Caldwell R., Dave R., Steinhardt P., 1998, Phys. Rev. Lett., 80, 1582
Chevallier M., Polarski D., 2001, Int. J. Modern Phys. D, 10, 213
Delubac T. et al., 2015, A&A, 574, A59
Fang W., Hu W., Lewis A., 2008, Phys. Rev. D, 78, 087303

Felice A. D., Nesseris S., Tsujikawa S., 2012, J. Cosmol. Astropart. Phys., 05, 29
Font-Ribera A. et al., 2014, J. Cosmol. Astropart. Phys., 05, 027
Grandis S., Seehars S., Refregier A., Amara A., Nicola A., 2016, J. Cosmol. Astropart. Phys., 5, 034
Green D. A., 2011, Bull. Astron. Soc. India, 39, 289
Handley W. J., Hobson M. P., Lasenby A. N., 2015a, MNRAS, 450, L61
Handley W. J., Hobson M. P., Lasenby A. N., 2015b, MNRAS, 453, 4384
Hee S., Handley W. J., Hobson M. P., Lasenby A. N., 2015, MNRAS, 455, 2461
Holsclaw T., Alam U., Sansó B., Lee H., Heitmann K., Habib S., Higdon D., 2010a, Phys. Rev. D, 82, 103502
Holsclaw T., Alam U., Sansó B., Lee H., Heitmann K., Habib S., Higdon D., 2010b, Phys. Rev. Lett., 105, 241302
Howlett C., Lewis A., Hall A., Challinor A., 2012, J. Cosmol. Astropart. Phys., 4, 27
Huterer D., Starkman G., 2003, Phys. Rev. Lett., 90, 031301
Jassal H. K., Bagla J. S., Padmanabhan T., 2004, MNRAS, 356, L11
Jeffreys S. H., 1961, The Theory of Probability. Oxford Univ. Press, Oxford
Kullback S., Leibler R. A., 1951, Ann. Math. Stat., 22, 79
Lazkoz R., Salzano V., Sendra I., 2012, Eur. Phys. J. C, 72, 2130
Lewis A., Bridle S., 2002, Phys. Rev. D, 66, 103511
Lewis A., Challinor A., Lasenby A., 2000, ApJ, 538, 473
Linder E., 2003, Phys. Rev. Lett., 90, 091301
MacKay D. J. C., 2003, Information Theory, Inference and Learning Algorithms. Cambridge Univ. Press, Cambridge
Parkinson D., Liddle A. R., 2013, Stat. Analysis Data Mining, 6, 3
Planck Collaboration I, 2016, A&A, 594, A1
Planck Collaboration XI, 2016, A&A, 594, A11
Planck Collaboration XIII, 2016, A&A, 594, A13
Planck Collaboration XIV, 2016, A&A, 594, A14
Planck Collaboration XX, 2016, A&A, 594, A20
Ratra B., Peebles P., 1988, Phys. Rev. D, 37, 3406
Sahni V., 2005, in Tamvakis K., ed., Lecture Notes in Physics, Vol. 653, The Physics of the Early Universe. Springer-Verlag, Berlin, p. 141
Sahni V., Starobinsky A., 2006, Int. J. Modern Phys. D, 15, 2105
Seehars S., Amara A., Refregier A., Paranjape A., Akeret J., 2014, Phys. Rev. D, 90, 023533
Seehars S., Grandis S., Amara A., Refregier A., 2016, Phys. Rev. D, 93, 103507
Seikel M., Clarkson C., Smith M., 2012, J. Cosmol. Astropart. Phys., 6, 036
Serra P., Cooray A., Holz D. E., Melchiorri A., Pandolfi S., Sarkar D., 2009, Phys. Rev. D, 80, 121302
Sivia D. S., Skilling J., 2006, Data Analysis: A Bayesian Tutorial. Oxford Univ. Press, Oxford
Skilling J., 2004, in Fischer R., Preuss R., Toussaint U. V., eds, AIP Conf. Ser. Vol. 119, Nested Sampling. Am. Inst. Phys., New York, p. 1211
Skilling J., 2006, Bayesian Analysis, 1, 833
Trotta R., Feroz F., Hobson M., Roszkowski L., de Austri R. R., 2008, J. High Energy Phys., 0812, 024
Tsujikawa S., 2013, Classical Quantum Gravity, 30, 214003
Vázquez J. A., Bridges M., Hobson M., Lasenby A., 2012a, J. Cosmol. Astropart. Phys., 6, 6
Vázquez J. A., Bridges M., Hobson M., Lasenby A., 2012b, J. Cosmol. Astropart. Phys., 9, 20
Zhao G.-B., Huterer D., Zhang X., 2008, Phys. Rev. D, 77, 121302
Zunckel C., Trotta R., 2007, MNRAS, 380, 865

This paper has been typeset from a TEX/LATEX file prepared by the author.