Fables of reconstruction
Examining the evidence for dynamical dark energy

Robert Crittenden
Institute of Cosmology and Gravitation
University of Portsmouth

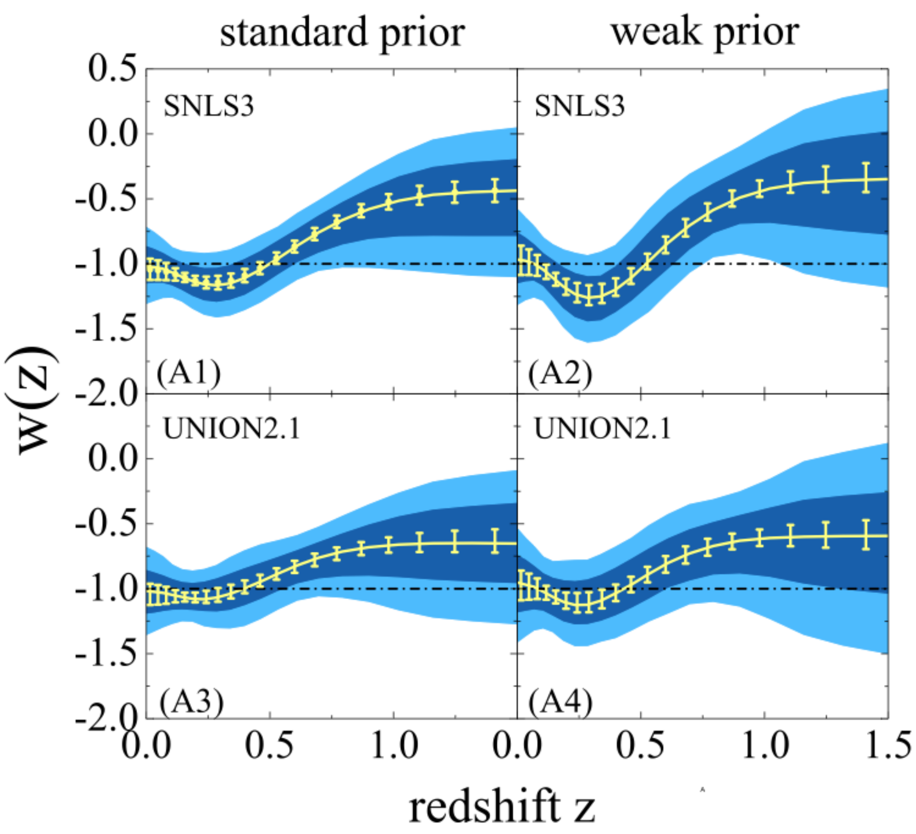
ICG Portsmouth

Work with G.B Zhao, L. Pogosian,
L. Samushia and X. Zhang

JCAP 1202 (2012) 048 and arXiv:1207.3804

Warning: Statistics ahead!

Summary: we have developed a new Bayesian method to better understand the possible dynamics of dark energy. Applying it to the most recent data sets suggests possible evolution in the equation of state, but the evidence is still marginal.



Zhao et al. 2012
arXiv 1207.3804

Growing dark energy evidence

We know the expansion of the Universe is accelerating from a wide variety of probes:

- Type 1A Supernovae as standard candles
- Angular size of CMB features
- Baryon acoustic feature in the large scale galaxy distribution
- *H(z)* from ages of passively evolving galaxies
- ISW signal seen in cross correlations of CMB and LSS
- Galaxy peculiar velocities, seen in redshift space distortions

Further evidence of low matter density from:

- Galaxy clusters
- Weak lensing measurements

Closing the net

By combining these data, we are beginning to learn about the dark energy properties.

- How much is there today?

$$\rho_{DE}(a=1) \text{ or } \Omega_{DE} = \rho_{DE} \times \frac{8\pi G}{3H^2}$$

- How is it evolving?

$$w(a) = \frac{p_{DE}}{\rho_{DE}} \quad \rho_{DE} \propto a^{-3(1+w)}$$

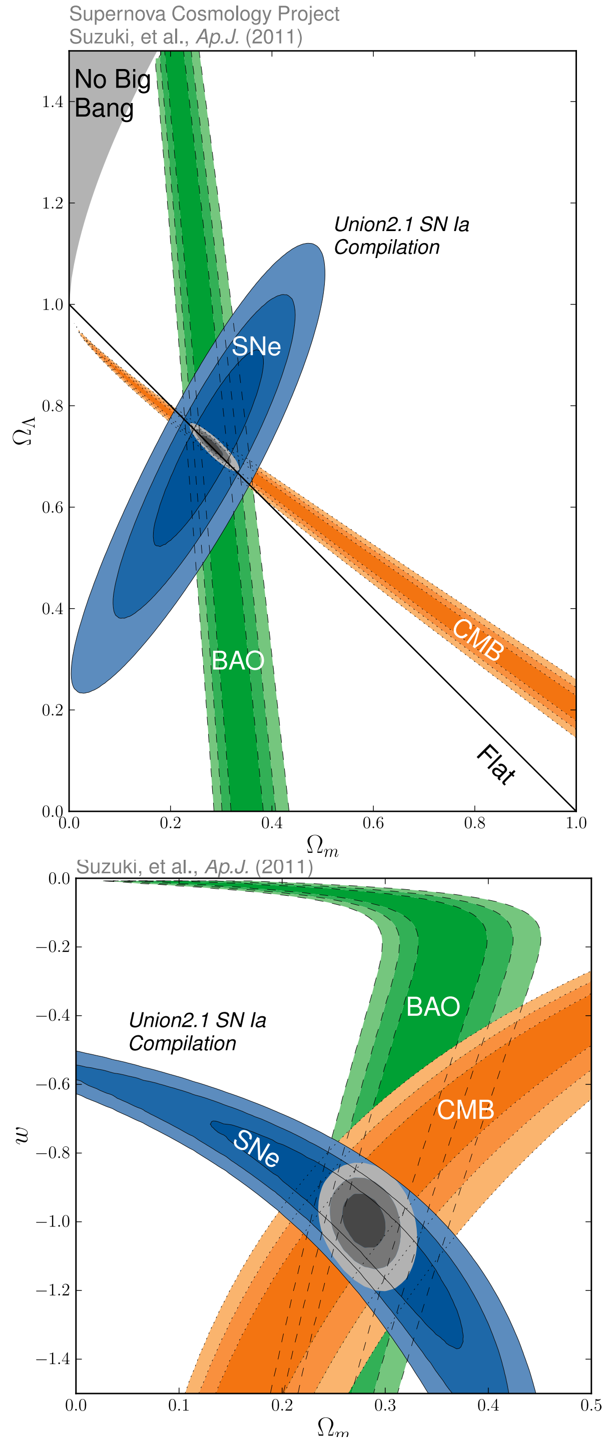
Equation of state

- Does it cluster?

$$\lambda = \int_0^{t_0} c_{DE}\, dt$$

Dark energy sound horizon

- Could the laws of gravity be wrong?



Understanding dark energy

Ultimately, we will use these measured properties to determine the nature of dark energy, from a range of ugly models:

- **Cosmological constant** – energy density of vacuum, constant in space and time.

$$w = \frac{p_\Lambda}{\rho_\Lambda} = -1$$

- **Quintessence** – scalar field rolling down some flat potential, similar to inflation; horizon scale sound horizon.

$$-1 < w < -\frac{1}{3}$$

- **Phantom models** – generic name for dark energy density which increases with time. (Leading to 'big rip' and other problems.)

$$w < -1$$

- **Modified gravity** – (e.g. f(R) or extra-dimensional models.) Background can look like cosmological constant, but structure growth is modified.

The big questions

How will we decide between these, or other possible models of dark energy (e.g. inhomogeneous models)?

- Is dark energy a true cosmological constant, or is it dynamical?
- If it is evolving in space and time, what's the best way to understand its nature and reconstruct its history?
- How much data is enough to understand dark energy?

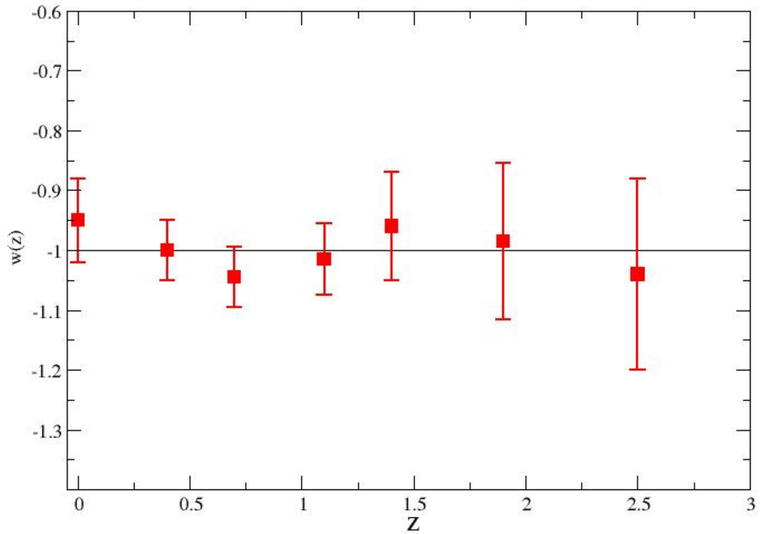
Answering these questions would be straightforward except for a fundamental problem:

Apart from the cosmological constant model, most models of dark energy are not well enough defined to allow us to pose clear statistical questions.

In particular, there generally is not a clear metric on the space of dynamical functions in most alternative models.

Lambda or something else?

Even given data consistent with a cosmological constant, the presence of noise means that there will be some dynamical model which provides a better fit.



How likely is that specific dynamical model in the space of all models?

As all good Bayesians know, the key is to look at the evidence ratios:

$$\int d\Lambda \mathcal{P}(\mathbf{D}|\Lambda)\mathcal{P}(\Lambda)$$

Prior on space of cosmological constant amplitude

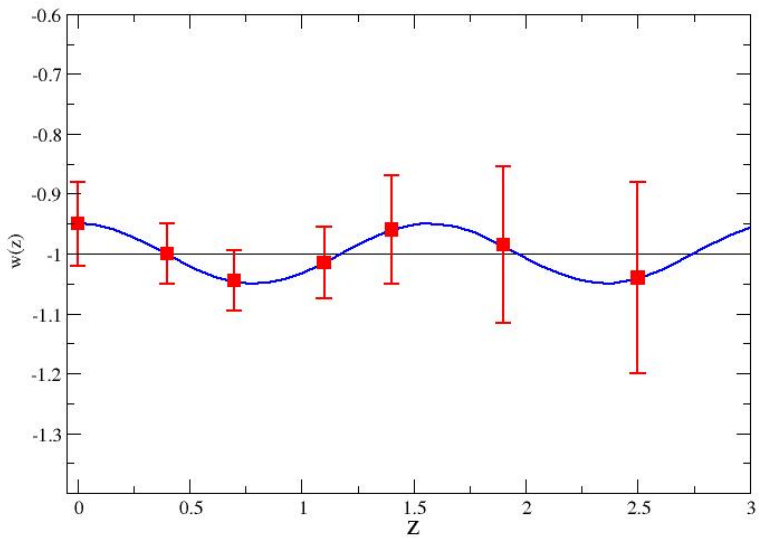
$$\int d^n\mathbf{w}\, d\rho_0 \mathcal{P}(\mathbf{D}|\mathbf{w},\rho_0)\mathcal{P}(\mathbf{w},\rho_0)$$

Prior on space of dynamical dark energy models

Lambda or something else?

The priors can provide a volume suppression to high dimensional models. But without knowing these theoretical priors, we cannot perform these evidence calculations.

And you never know, there may be a simple theory that could easily explain what you assumed was simply noise…



$$\int d\Lambda \mathcal{P}(\mathbf{D}|\Lambda)\mathcal{P}(\Lambda)$$

Prior on space of cosmological constant amplitude

$$\int d^n\mathbf{w}\, d\rho_0 \mathcal{P}(\mathbf{D}|\mathbf{w},\rho_0)\mathcal{P}(\mathbf{w},\rho_0)$$

Prior on space of dynamical dark energy models

Phenomenological approaches

Lacking guidance from theory, we need to make some choices for how to parameterize the dark energy behaviour and the prior on these parameters.

Which variables should we parameterise?

Many choices have been used in the literature, from the more phenomenological to more theoretical, e.g.:

$$H(z), \rho(a), D_A(z), w(a), \gamma(z), V(\phi)$$

Some choices can be tightly tied to particular observations, and are insufficient to connect to all observations, while other choices are so specifically model based, they can be difficult to connect to other theories.

We choose to work with *w(a)* because it connects to the expansion history and growth history (albeit assuming a simple sound speed) and is easy to relate to a range of theories.

Phenomenological approaches

Parametric or non-parametric?

Parametric approaches tend to have relatively few parameters, often motivated by underlying models.

$$w, [w_o, w_a], [w_i, w_0, w_t, \Delta a]$$

Parametric approaches are simple, but they lose information.

They can miss out on degrees of freedom which the observations are most sensitive to, and the results cannot be used to constrain any other parameterization.

Non-parametric approaches tend to chose some set of cosmological functions to expand in some set of basis functions.

E.g. binning, wavelets, Legendre polynomials

Given sufficient numbers of basis functions, they preserve information and can be a useful data compression step.
The cost: many poorly constrained degrees of freedom.

What could observations tell us?

Here, we focus on non-parametric methods to ensure we don't lose information that the observations are telling us.

To quantify this information, we can perform a Fisher matrix analysis for the projected experiments. This yields a Gaussian approximation for the expected likelihood as a function of the basis amplitudes.

$$\mathcal{F}_{ij} = -\left\langle \frac{\partial^2 \ln \mathcal{P}(\mathbf{D}|\mathbf{T})}{\partial T_i \partial T_j} \right\rangle$$

Fisher matrix describes how well an experiment constrains parameters

Principal components are the eigenmodes of the Fisher matrix, and tell us the combinations of amplitudes that will be constrained the best. These tend to be the modes that are low frequency and are localized where the data are measured.

Given enough basis functions, these principal components are largely independent of the choice of non-parametric method. But this gives a lower limit on the number of basis functions we can have without losing information.

What could observations tell us?

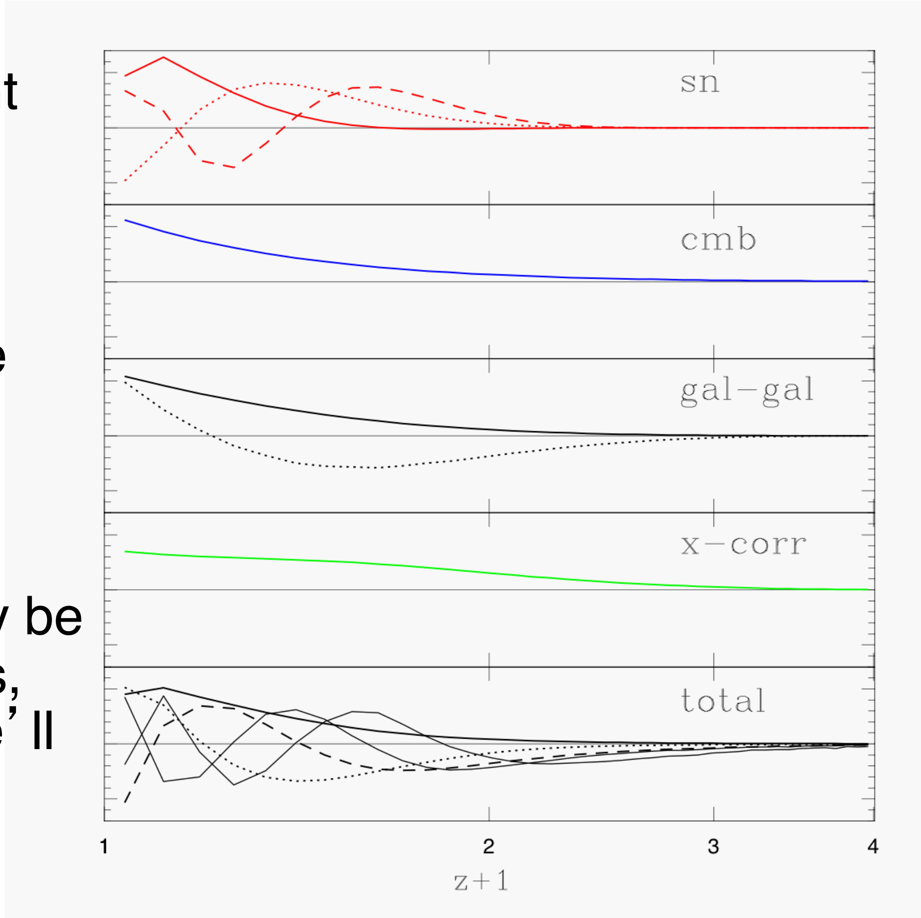
Principal components:

Each experiment measures a different combination of modes, and often more than one independent mode.

For example, SN constrain multiple modes at a reasonable level, while the CMB constrains only one (the CMB shift) but it measures it very well.

By combining data we may eventually be able to learn about 4-5 parameters, starting with low frequency, but we'll eventually get higher frequency modes.

Very sensitive to assumptions about systematic errors!



Crittenden, Pogosian & Zhao '08
Huterer & Starkman
Huterer & Linder, Knox et al.

What could observations tell us?

Spectra of eigenvalues from future experiments:

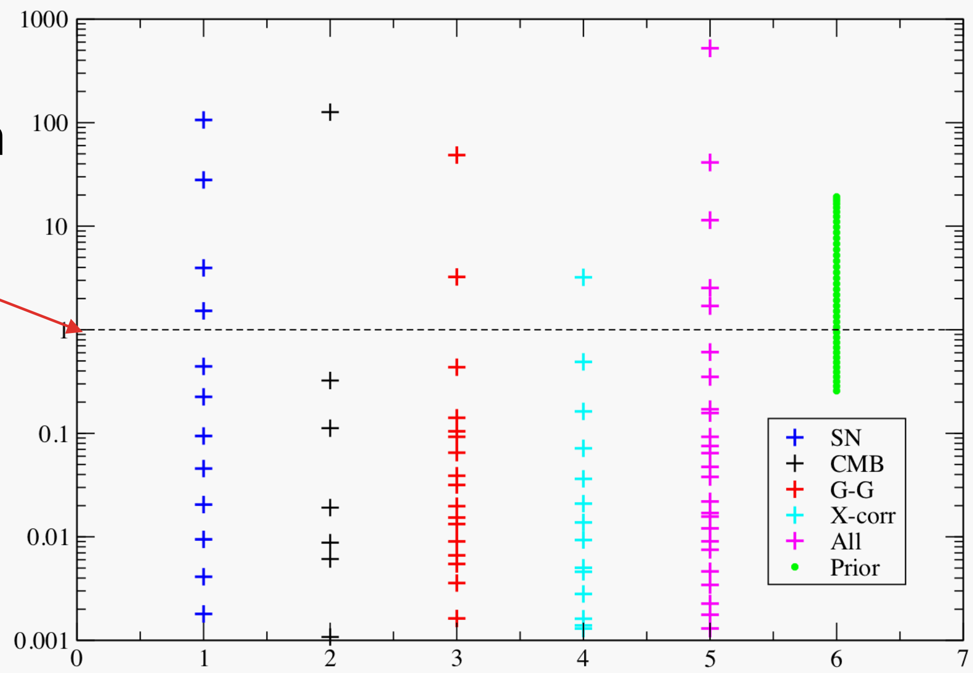
Higher ones are best determined $\sim 1/\sigma^2$

Where do we draw the line between the useful modes and the unhelpful ones?

It depends on what we think we already know!

In the absence of any prior information, they are all informative. But we always know something, and at some point our prior knowledge is better than the weak observations.

Most informative



Least informative

Reconstructing w(a)

Various non-parametric methods have been suggested for reconstructing w(a) from the data:

- Simple maximum likelihood estimators tend to be very noisy, because some eigenvalues have very large variance. (Flat directions in the large parameter space, which take forever to converge in MCMC analyses.)

- Most methods implement some *ad hoc* smoothing aspect, intrinsically assuming the high frequency behavior is noise, rather than real behavior.

- An alternative is to keep only those principal components which are best determined (e.g. Huterer & Starkman 02). But how do we decide how many to keep?

Reconstructing w(z)

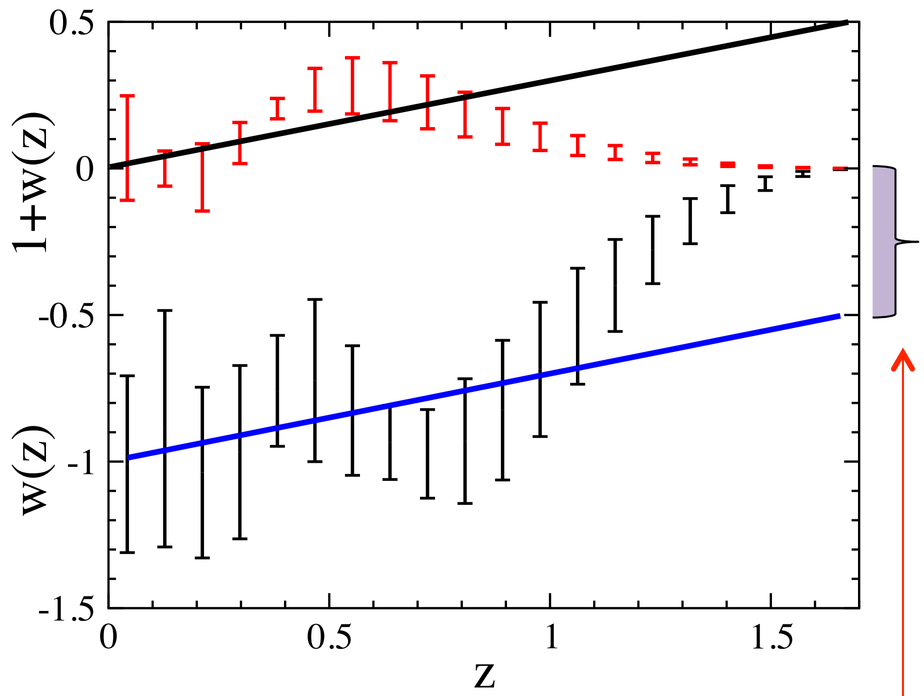
Truncating the principal components (Huterer & Starkman 02)

Keeping just a few principal components reduces the variance, but introduces another issue, **bias.**

- Where the data are poor at high redshifts, the result tends to the fiducial model (w=0.)
- The errors are also very small at high redshifts because none of the modes kept has any weight there.

Many modes ➔ large variance
Few modes ➔ large bias

- Optimal number depends on the assumed input model.



Large bias where the fit reverts to fiducial model

Bias and variance

Mean squared error (MSE) as a metric for reconstructions

$$\mathrm{MSE} \equiv \sum_i (w_i^{\mathrm{true}} - w_i^{\mathrm{recon}})^2$$

This provides a means of evaluating the quality of a reconstruction. Considering the ensemble of data consistent with a given model, we can define the average reconstruction: $w_i^{\mathrm{mean}} \equiv \langle w_i^{\mathrm{recon}} \rangle$

The difference between this and the true model is the bias. One can show:

$$\langle \mathrm{MSE} \rangle = \sum_i \underbrace{\langle (w_i^{\mathrm{mean}} - w_i^{\mathrm{recon}})^2 \rangle}_{\text{Variance}} + \underbrace{(w_i^{\mathrm{true}} - w_i^{\mathrm{mean}})^2}_{(\text{Bias})^2}$$

The challenge is to find a method which keeps both terms small.

Bayesian reconstruction

Many reconstruction methods in the literature are arguably *ad hoc*, because they arbitrarily decide on what level to smooth or how many modes to keep.

These decisions are made without reference to the evidence which might exist in the data, and usually without considering explicitly what theoretical models are more likely.

Bayesian methods can provide a simple and transparent means of reconstruction.

However, again the method requires a ***prior distribution*** for the expected *w(a)* behavior. The advantage is that any assumptions are made ***fully explicit.***

Bayesian reconstruction

Bayes' theorem:

Likelihood

Prior

Summarizes theory and earlier data

$$\mathcal{P}(\mathbf{T}|\mathbf{D}) = \frac{\mathcal{P}(\mathbf{D}|\mathbf{T})\mathcal{P}(\mathbf{T})}{\mathcal{P}(\mathbf{D})}$$

Posterior

Evidence

Normalizing factor, used to compare different models or parameterisations

We search for the model parameters at which the posterior is maximum, and use the prior to control the flat directions in the likelihood.

Defining a prior

The prior distribution should reflect the theoretical considerations or previous data. Usually the new data far outweighs earlier data, so the prior primarily reflects theoretical considerations.

Generically the prior is an arbitrary function on the space of non-parametric amplitudes.

We simplify by assuming that it is Gaussian. Then we simply need to specify:

- The mean of the distribution, or fiducial *w(a)* model.
- The covariance matrix describing how the amplitudes are correlated.

We further assume the correlations are translation invariant in some independent variable, like the scale factor. This means the covariance matrix is replaced by a 1-D correlation function.

The correlated prior

These assumptions imply we can define a correlation function:

$$\xi_w(|a-a'|) \equiv \left\langle [w(a) - w^{\rm fid}(a)][w(a') - w^{\rm fid}(a')] \right\rangle$$

From this we can derive a covariance matrix.

In principle, the correlation function should be derived from theoretical considerations. Here we explore phenomenological forms, which have a correlation length given by a_c.
For example,

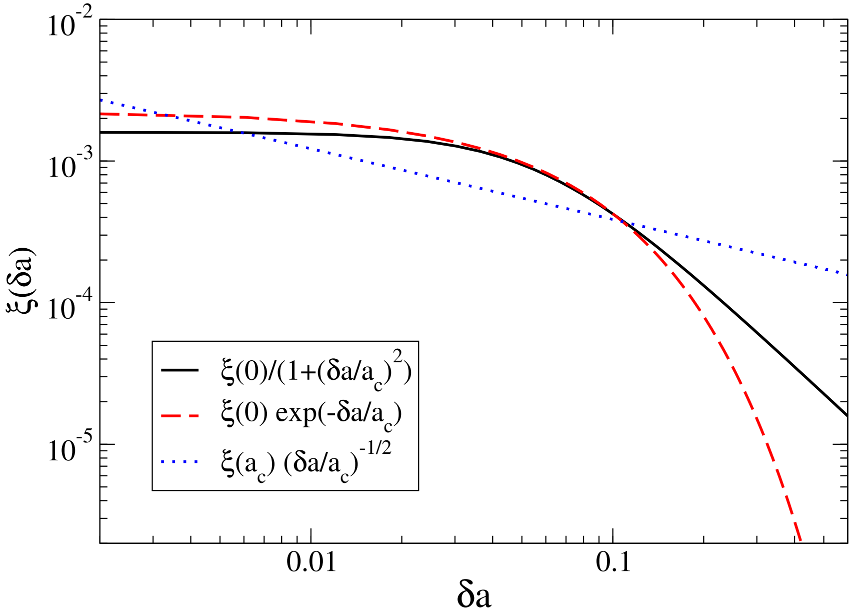
$$\xi_w(\delta a) = \frac{\xi_w(0)}{1 + (\delta a / a_c)^2} \ ,$$

If the correlation length is larger than the binning width, the precise choice of binning, or set of basis functions, becomes irrelevant.

Correlation functions

We considered a few possible correlation shapes.

These are all normalised to the same correlation length and the same constraint on the mean equation of state.



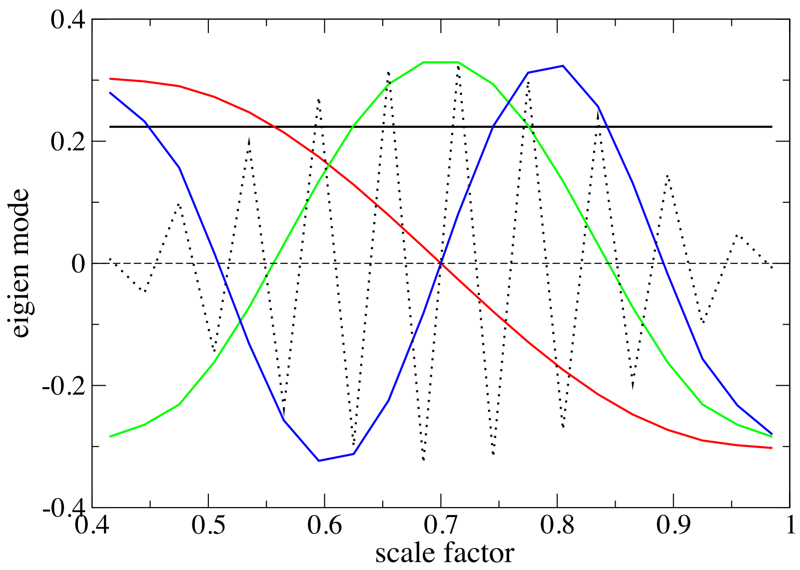
Functions which are divergent at small scales lead to diagonal covariance matrices, which means all modes are degenerate and independent. Flatter functions lead to more interesting correlations, so that higher frequency modes are better constrained.

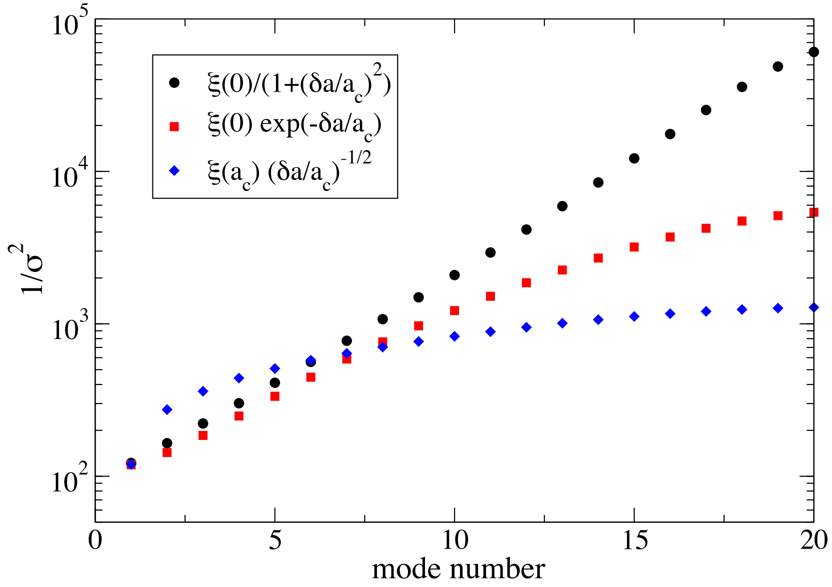
It is easy to discretize this to get the covariance matrix for a given basis.

Prior principal components

Perhaps surprisingly, the different correlation functions all have virtually the same eigenvectors, ordered in the same way. These are effectively the Fourier modes.

All that changes are the eigenvalues, but the highest frequency modes are always the most strongly constrained by the prior. This is precisely the opposite of the data constraints.



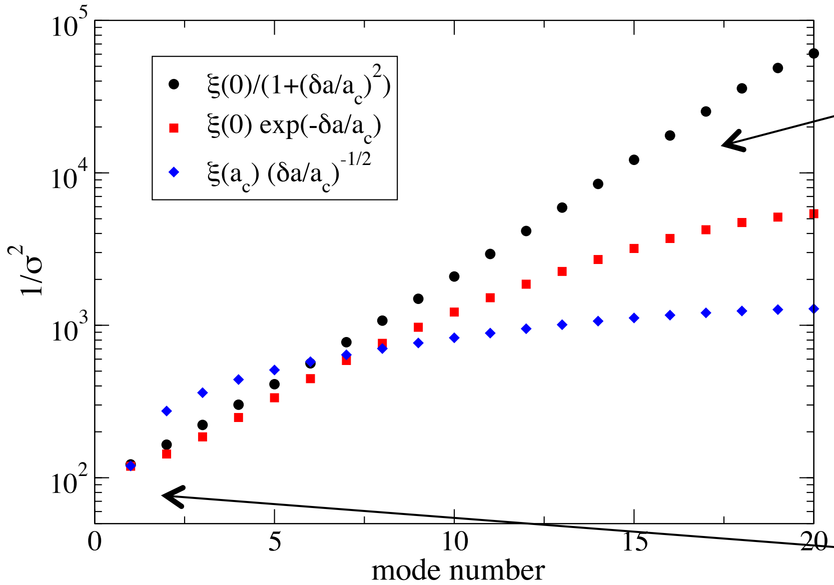


Interpreting the eigenvalues

For simplicity we focus on a simple correlation shape:

$$\xi_w(\delta a) = \frac{\xi_w(0)}{1 + (\delta a / a_c)^2}$$

High frequency



Low frequency
weakest constraints

Slope, or ratio of eigenvalues is determined by the correlation length a_C

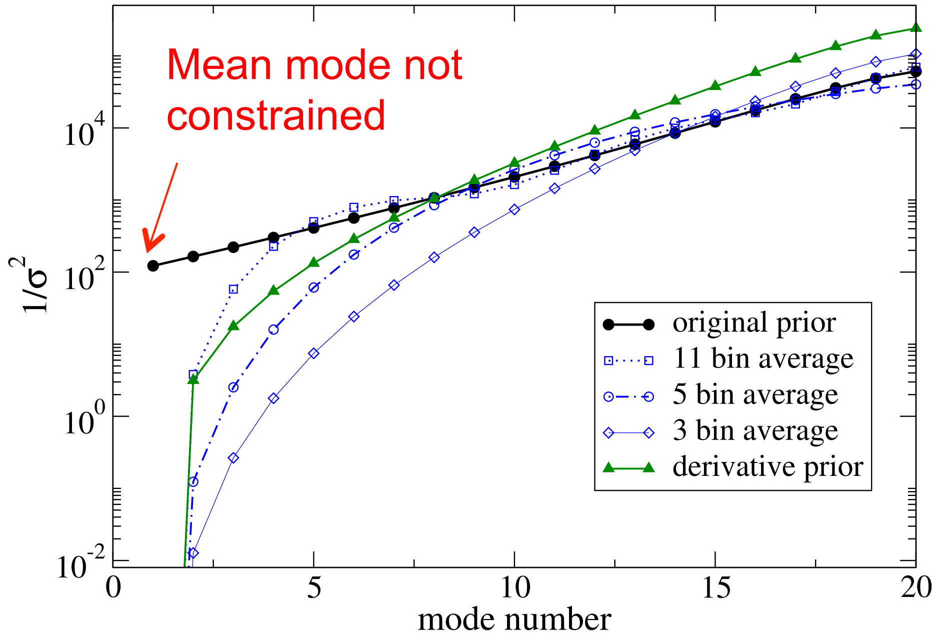
First mode is the bin average (zero frequency), the constrain the prior would put on a constant w model

Marginalising away the fiducial model

Since we are trying to determine if dark energy is dynamical, we don't want the prior to prefer any particular mean value.

We can remove this in many ways:

- Marginalizing over the fiducial value
- Subtracting a locally defined average
- Putting a prior instead on the derivative of w(a)



All of these methods just tweak the eigenvalues similarly.

They reduce the constraints on the long wavelength modes, ***particularly the global average***.

Wiener filtering

Assuming the distributions are Gaussian, the predictions for the ensemble of data can be solved analytically.

Prior distribution

$$\mathcal{P}_{\mathrm{prior}}(\mathbf{w}) \propto e^{-(\mathbf{w}-\mathbf{w}^{\mathrm{fid}})^T \mathbf{C}^{-1}(\mathbf{w}-\mathbf{w}^{\mathrm{fid}})/2}$$

Likelihood

$$\mathcal{P}(\mathbf{w}^{\mathrm{obs}}|\mathbf{w}) \propto e^{-(\mathbf{w}^{\mathrm{obs}}-\mathbf{w})^T \mathbf{F}(\mathbf{w}^{\mathrm{obs}}-\mathbf{w})/2}$$

Maximum posterior solution

$$\mathbf{w}^{\mathrm{recon}} = \mathbf{F}^{-1}(\mathbf{C}+\mathbf{F}^{-1})^{-1}\mathbf{w}^{\mathrm{fid}} + \mathbf{C}(\mathbf{C}+\mathbf{F}^{-1})^{-1}\mathbf{w}^{\mathrm{obs}}$$

High pass filter of fiducial model. (Removed)

Low pass filter of noisy peak of likelihood

Expected bias

$$\sum_i (w_i^{\mathrm{true}} - w_i^{\mathrm{mean}})^2 = ||\mathbf{F}^{-1}(\tilde{\mathbf{C}}+\mathbf{F}^{-1})^{-1}\mathbf{w}^{\mathrm{true}}||^2$$

Combining with data

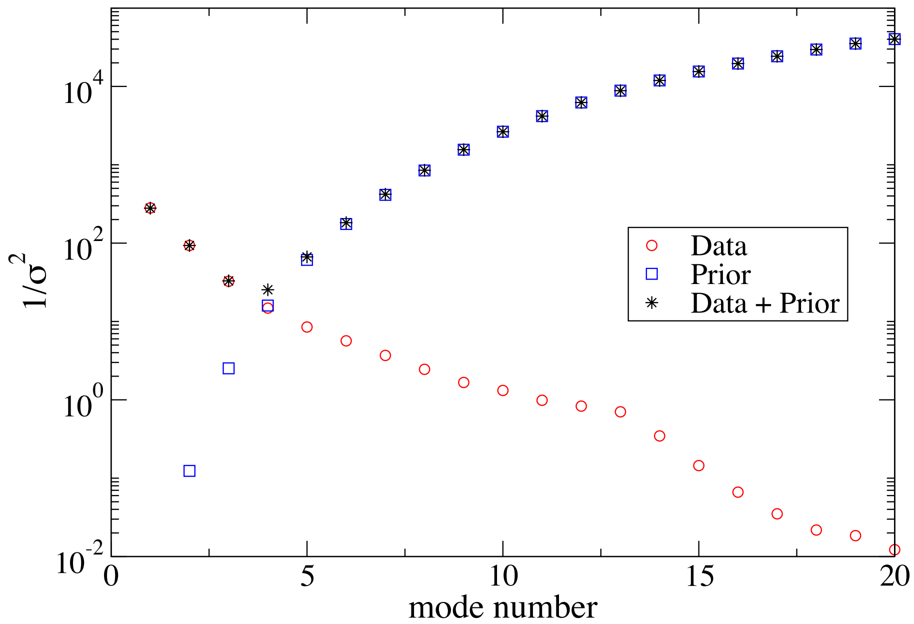
The posterior Fisher matrix is the combined prior + Fisher matrix covariance:

$$\mathcal{F}_{post} = \mathcal{F}_{data} + \mathcal{C}_{prior}^{-1}$$

The information from the data govern the long wavelength modes, while the prior sets the higher frequency modes.

In between, they compete to determine the best amplitudes.

In this example, four data modes survive to reconstruct the data.



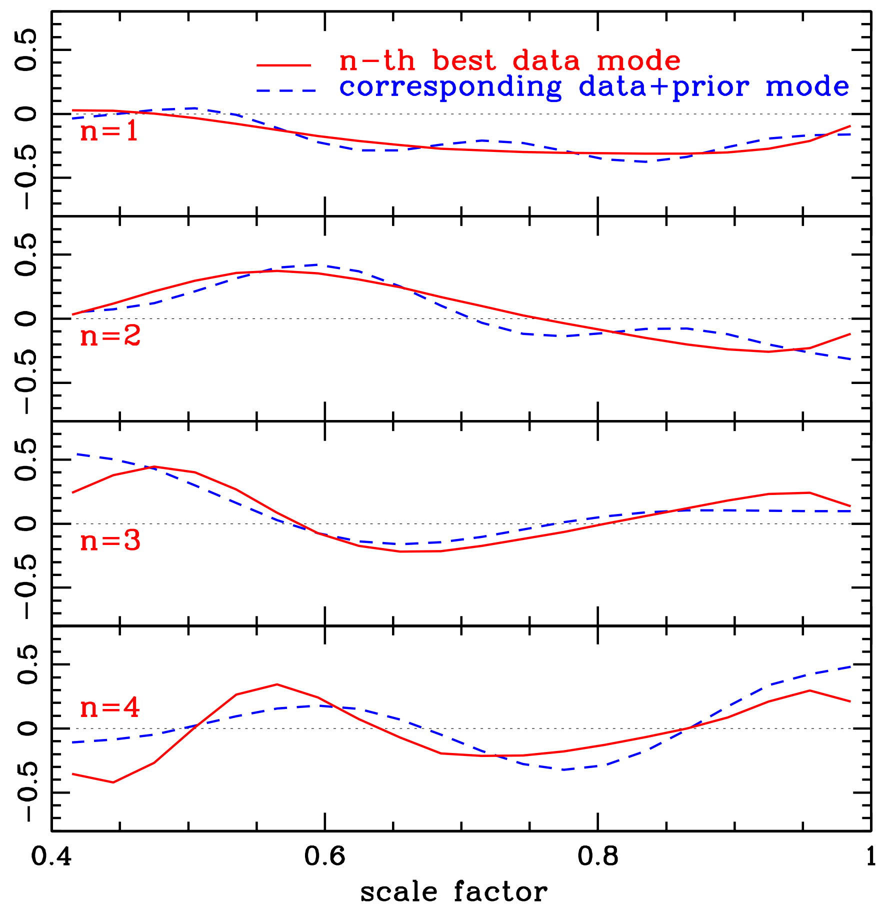
Low frequency High frequency

For the most part, the prior constrains those degrees of freedom that are noise dominated.

Impact of prior

The low frequency eigenmodes are very similar to the original ones defined by the data.

Any model which can be expressed with these surviving modes can be reconstructed with little bias, while the prior minimizes the variance.

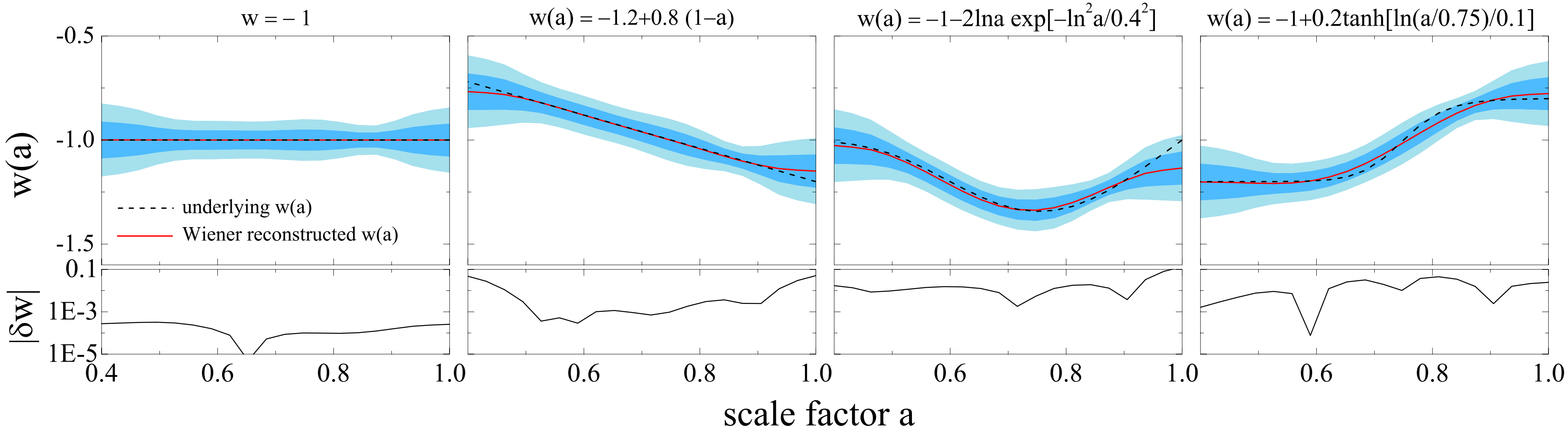


Gaussian projections

We have shown that the ***Wiener filtering*** predictions match numerical realizations.

We explored this for a range of potential models, assuming futuristic SN and H(z) data, and found they could well reproduce equation of state.

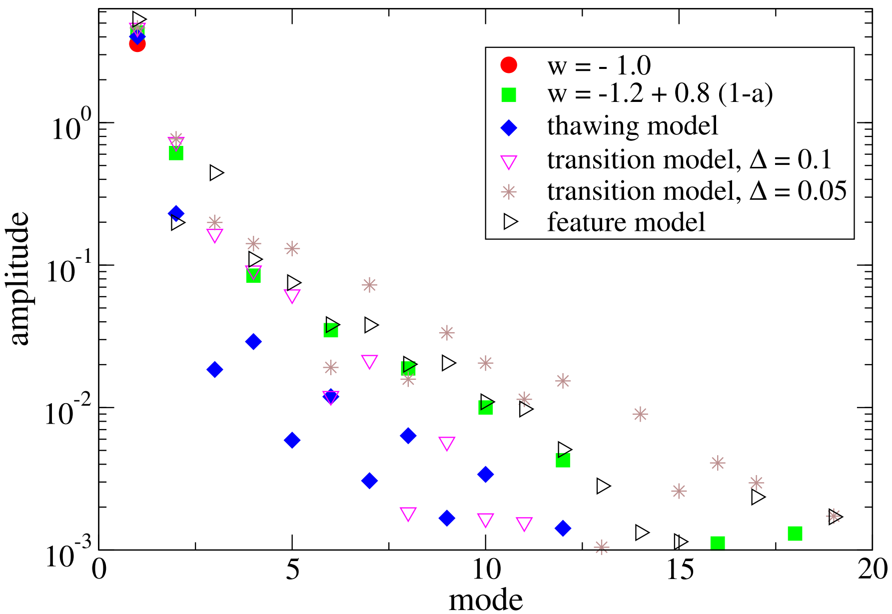
Higher frequency modes are smoothed out leading to biased reconstructions, but this was smaller than the variance contribution.



Predicting the expected bias

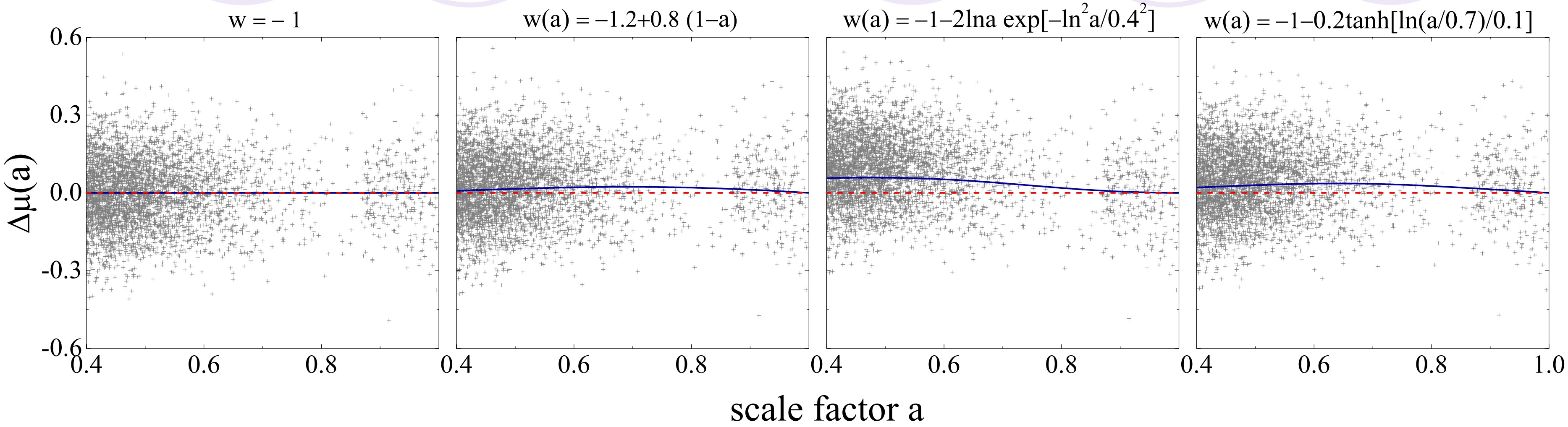
By expanding the models in the prior principal components, we can predict which models will be reproduced well, and which will be strongly biased.

Since the prior prefers smooth functions, they are reconstructed with the least bias.

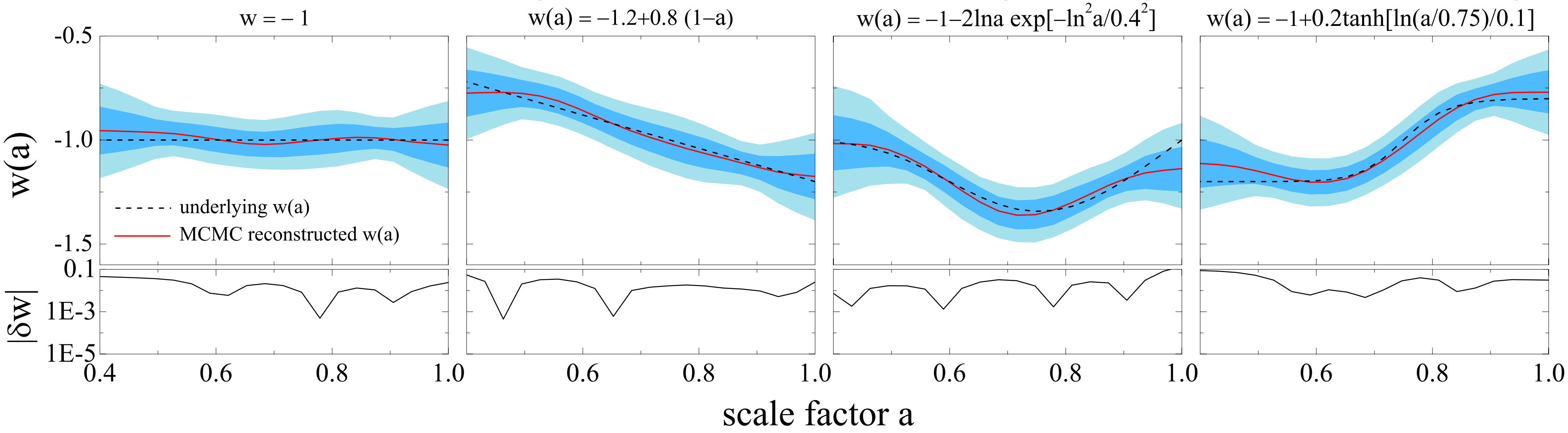


Here only the ‘feature’ model and the fast transition model have significant bias, which is to be expected given their sharp features.

More realistic simulations



These results are confirmed using MCMC reconstructions based on realisations of SN data. (Not quite the same things being plotted as before.)



Application to real data

Our reconstruction method appears to work well on simulated data, what about real data?

We have applied it to a collection of the latest data sets:

- Recent large SN data sets
  - Union 2.1 (Suzuki et al. 2012) 580 SNe
  - SNLS3 (Conley et al. 2011) 472 SNe
- Full WMAP CMB power spectrum
- Recent H(z) measurements (Moresco et al. 2012)
- BAO measurements (BOSS, WiggleZ, SDSS LRG, 6dF)
- Redshift space distortion measurements (BOSS, WiggleZ)

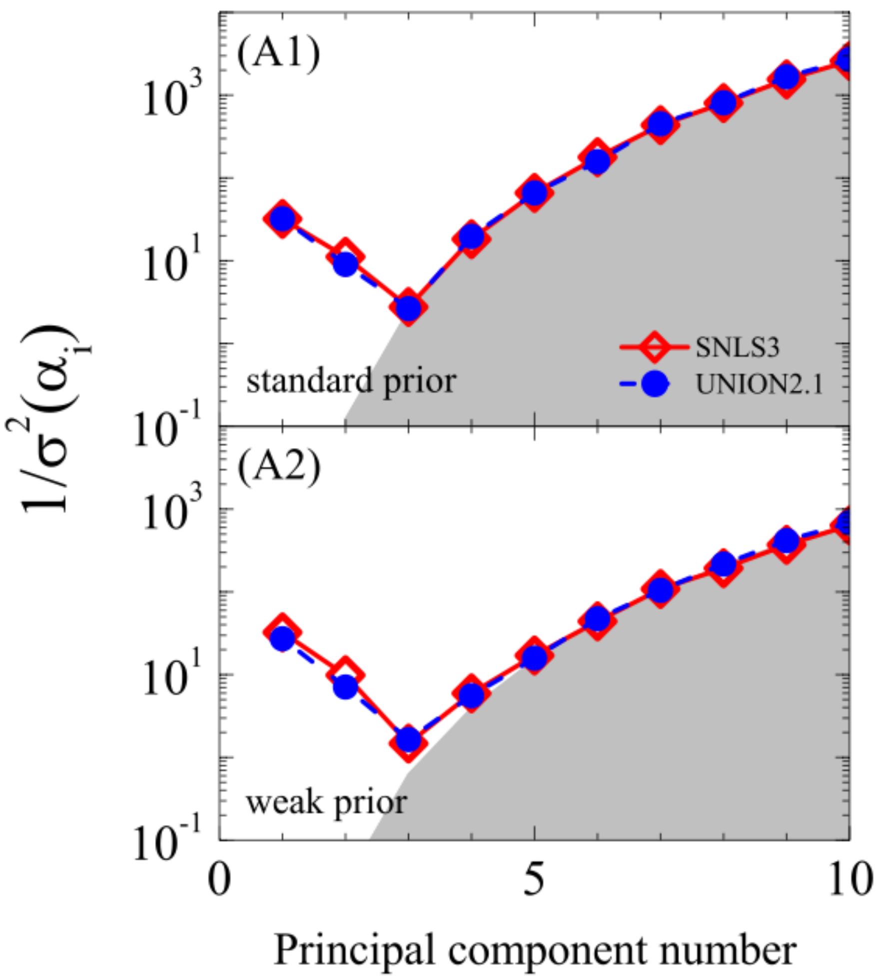
Principal Components

Which modes survive the prior to form the reconstruction?

This depends on the relative strengths of the data and the prior.

SNLS3 yields a slightly stronger data constraint, but the bigger difference is on the strength of the prior. **In either case, only 3 or 4 modes contribute.**

The relative strengths of the constraints on high frequency modes means that the priors would need to be shifted significantly for more modes to come through.



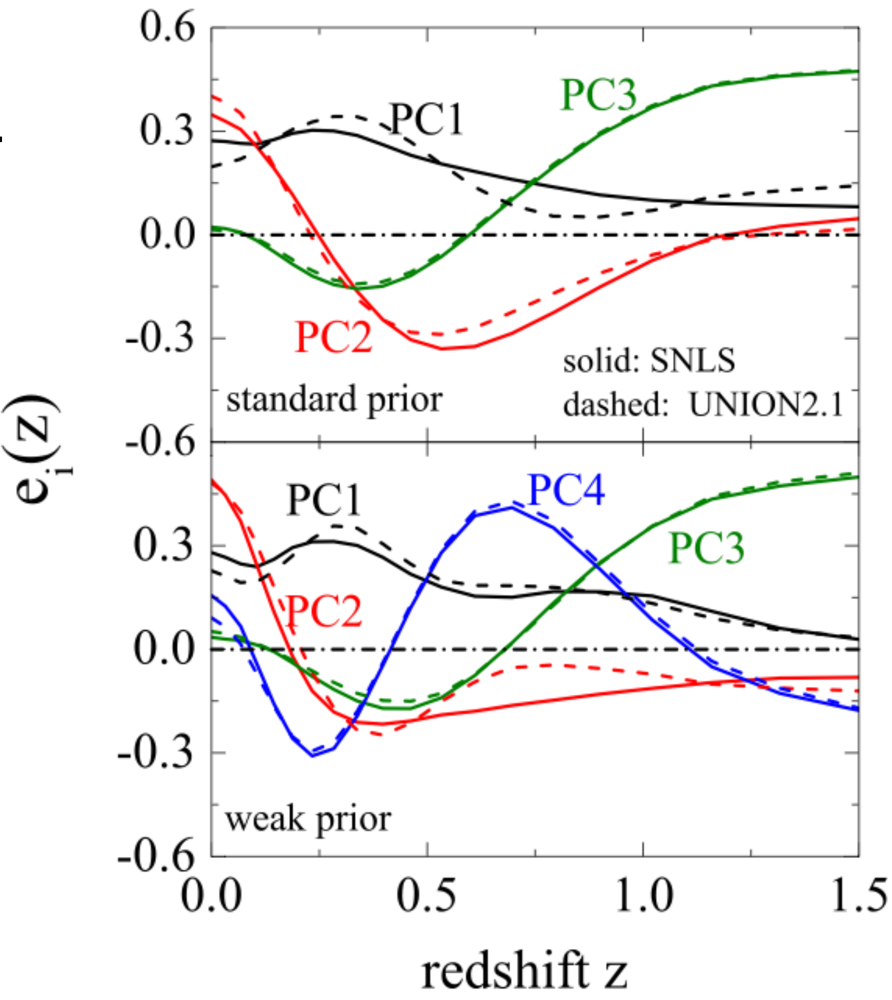
Principal Components

Which modes survive the prior to form the reconstruction?

This depends on the relative strengths of the data and the prior.

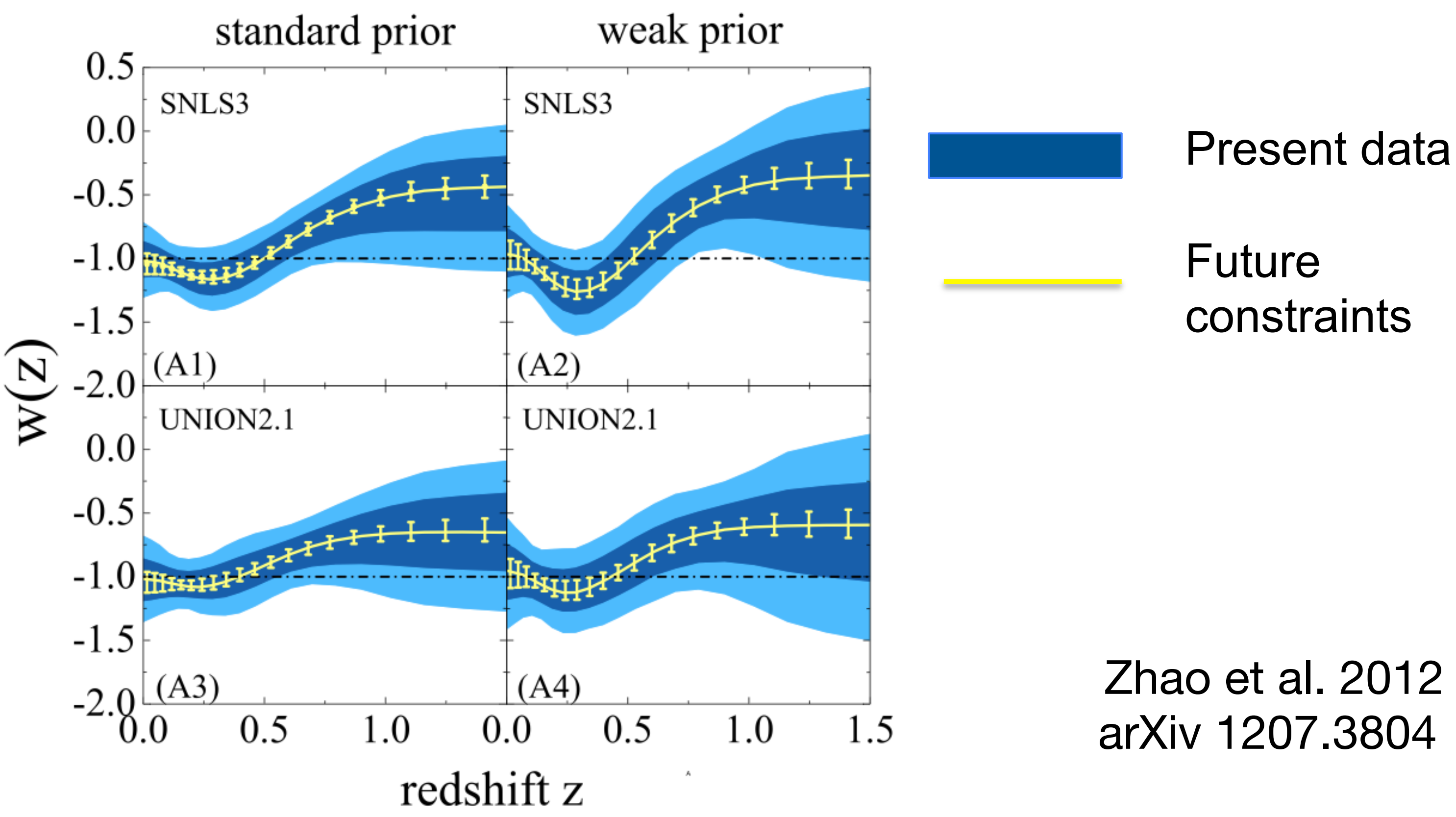
SNLS3 yields a slightly stronger data constraint, but the bigger difference is on the strength of the prior. **In either case, only 3 or 4 modes contribute.**

The relative strengths of the constraints on high frequency modes means that the priors would need to be shifted significantly for more modes to come through.



Reconstruction results

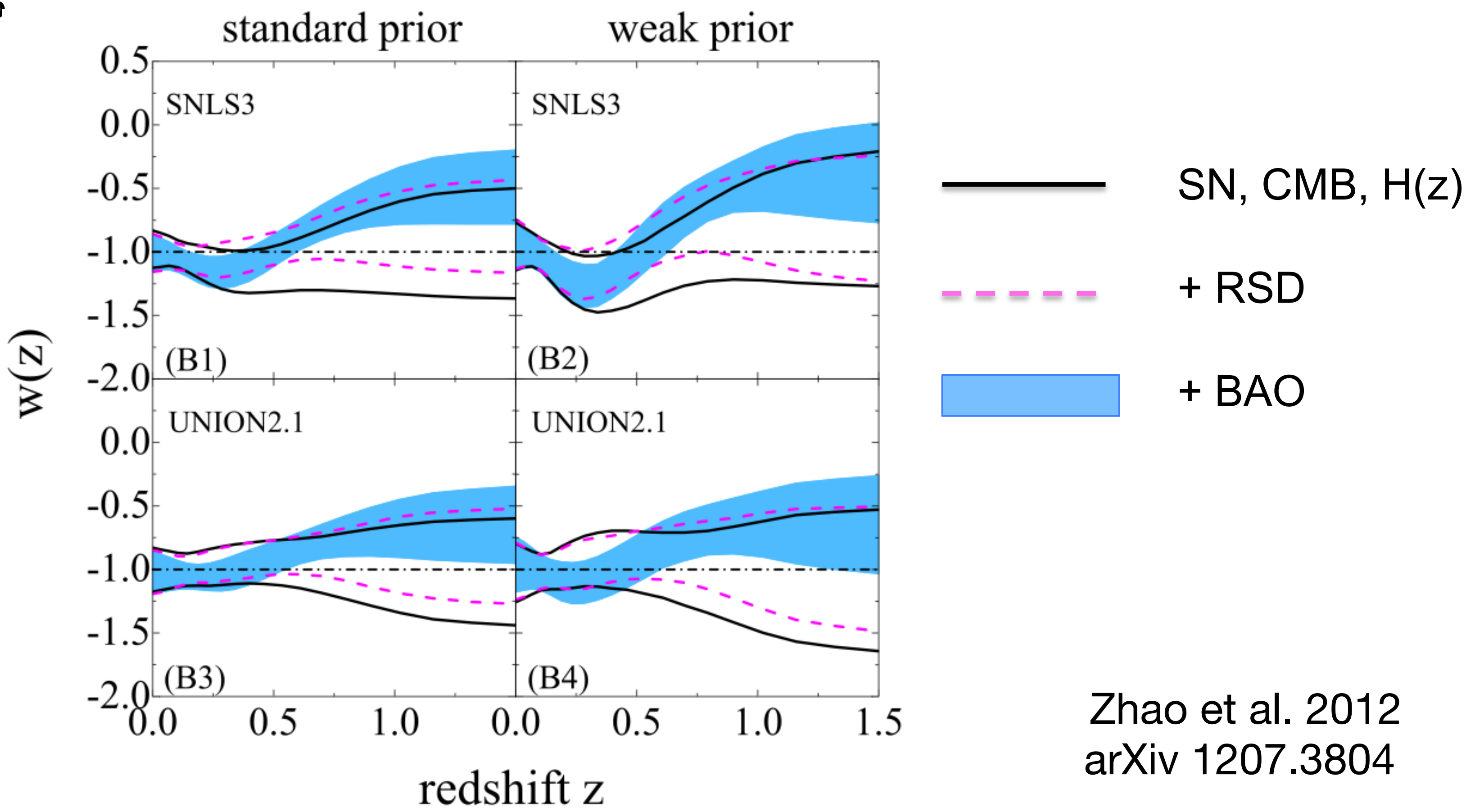
The reconstruction is largely consistent with cosmological constant behaviour, but there is a trend towards increasing w at higher redshifts, but this depends on the choice of data sets and is more pronounced as the prior is weakened.



Zhao et al. 2012
arXiv 1207.3804

Reconstruction results

The primary effect seems to come with the addition of the BAO data, which pulls down *w(a)* at low redshifts. This is compensated at high redshifts to match with the integrated constraint from the CMB

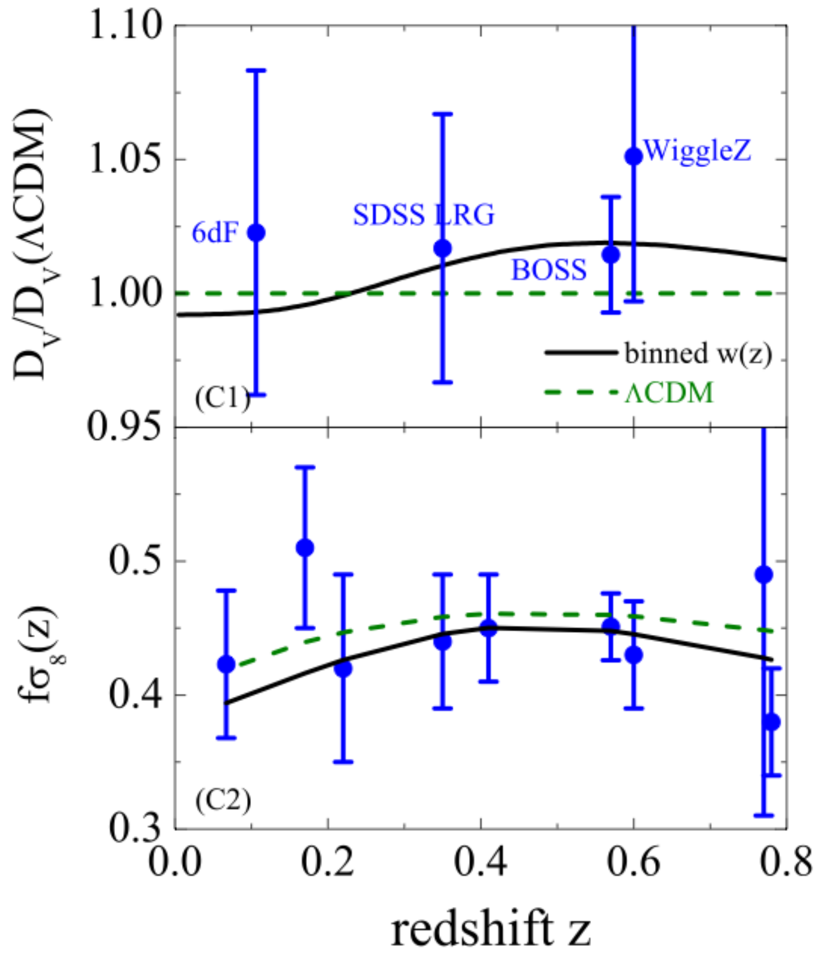


Zhao et al. 2012
arXiv 1207.3804

Reconstruction results

The reconstructed w(a) helps address the tension in the BAO measurements, which seem to consistently measure distances in excess of that for a cosmological constant.

The RSD fits are improved at high z, but worse at lower redshifts.



Zhao et al. 2012 arXiv 1207.3804

Is it significant?

Not surprisingly, with more degrees of freedom the fits improve relative to the cosmological constant model:

SNLS3 $\Delta\chi^2_{data} = 7.0$ Union 2.1 $\Delta\chi^2_{data} = 3.9$

But does is this improvement enough to compensate for the less predictive theory? For this we need the evidence, which for a Gaussian distribution is simply:

$$E \propto \left(\frac{\det \mathcal{C}_{\rm post}}{\det \mathcal{C}_{\rm prior}}\right)^{1/2} e^{-\chi^2_{\rm bf}/2}$$

Predictiveness – What fraction of prior volume is consistent with observations?

Fit quality – How well can the model fit the observations?

Evidence ratios

Relative fit quality

Relative predictiveness

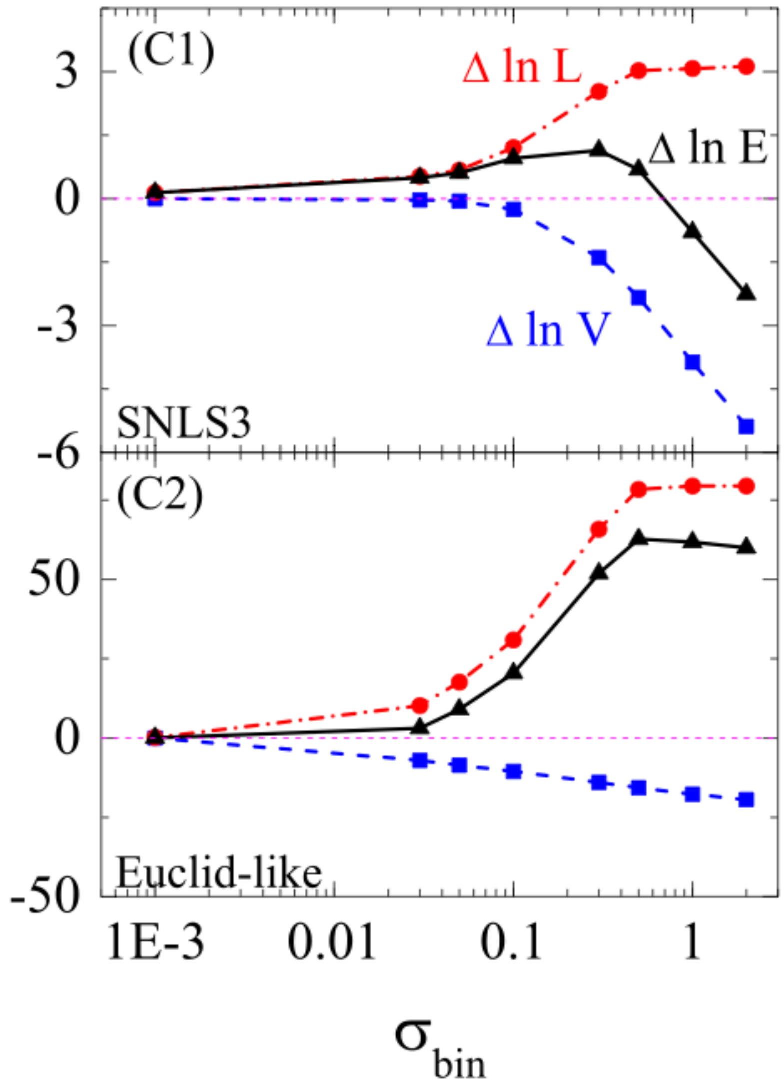
We consider a family of priors which has a cosmological constant limit:

$$\mathcal{C}^{-1}_{\rm prior} \to \mathcal{C}^{-1}_{\rm prior} + \sigma^{-2}_{\rm bin}\,\mathcal{I}$$

We plot the logs of the evidence ratios, compared to the cosmological constant limit.

Surprisingly, for some priors the evidence weakly prefers dynamical dark energy! (Ad hoc?)

Future data could do this easily for a broad range of prior choices.



Conclusions

We have developed a Bayesian method for reconstructing the equation of state which has many advantages:

- The method is transparent and the assumptions are explicit.
- Given an input model, we can predict the expected bias and variance of its reconstruction.
- Implementing it is easy and numerically inexpensive. It only requires choosing a prior covariance matrix and evaluating:

$$\chi^2_{\rm prior} = \mathbf{w}^T \tilde{\mathbf{C}}^{-1} \mathbf{w}$$

We have focused on w(z) as the means for describing dark energy, but it could be applied to any basis, such as the quintessence potential. (Or other problems, like power spectra estimation.)

Applying it to the present data, there is some evidence that *w(a)* was larger in the past, arising from BAO and SN data. However, this not strongly preferred compared to a cosmological constant model.

Gaussian processes

This method has many similarities to a 'Gaussian Processes' (GP) approach, which has recently been applied to the dark energy context:
Holsclaw et al, 2010; Shafieloo et al 2012; Seikel et al 2012.

The primary philosophical difference is that rather than choosing a prior on theoretical grounds, based on minimizing the bias of reconstruction in 'typical' models, GP describes the priors in terms of hyperparameters which are marginalised over.

How do we describe dark energy?

Parametric approaches:

Assume some particular functional form for ρ_{DE}, w_{DE} or some other way of describing dark energy, and constrain the parameters of that form.

Phenomenological: $\Omega_{DE}, w, (w_0, w_a), (w_0, w_1, z_t, \Delta z)$

More theoretical (e.g. quintessence potential): $V(\phi), \phi_0$

Potential issues:

Loss of information

Can't be used to constrain any other parametric model

Could potentially miss interesting features because they are not allowed in your parameterization!

How do we describe dark energy?

Non-parametric approaches:

Choose some cosmological function or functions to expand in a set of basis functions.

E.g. $$H(z), \rho(a), D_A(z), w(a), \gamma(z), V(\phi)$$

Basis functions could be binning, or other orthogonal functions which allow arbitrary behavior.

Potential issues:

How many basis functions do we use?

Too few, and your answers are sensitive to this choice.

Too many, and there will be many parameter degeneracies (flat directions in likelihood); hard to converge MCMC methods.

Also, what should we use as independent variable?