A 2 per cent distance to $z = 0.35$ by reconstructing baryon acoustic oscillations – I. Methods and application to the Sloan Digital Sky Survey

Nikhil Padmanabhan,1* Xiaoying Xu,2 Daniel J. Eisenstein,3 Richard Scalzo,1,4 Antonio J. Cuesta,1 Kushal T. Mehta2 and Eyal Kazin5,6

1Department of Physics, Yale University, 260 Whitney Avenue, New Haven, CT 06520, USA
2Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, AZ 85721, USA
3Harvard-Smithsonian Center for Astrophysics, Harvard University, 60 Garden Street, Cambridge, MA 02138, USA
4Research School of Astronomy & Astrophysics, The Australian National University, Mount Stromlo Observatory, Cotter Road, Weston, ACT 2611, Australia
5Centre for Astrophysics and Supercomputing, Swinburne University of Technology, PO Box 218, Hawthorn, VIC 3122, Australia
6Center for Cosmology and Particle Physics, New York University, 4 Washington Place, NY 10003, USA

Accepted 2012 August 7. Received 2012 August 1; in original form 2012 February 1

ABSTRACT

We present the first application to density field reconstruction to a galaxy survey to undo the smoothing of the baryon acoustic oscillation (BAO) feature due to non-linear gravitational evolution and thereby improve the precision of the distance measurements possible. We apply the reconstruction technique to the clustering of galaxies from the Sloan Digital Sky Survey (SDSS) Data Release 7 (DR7) luminous red galaxy (LRG) sample, sharpening the BAO feature and achieving a 1.9 per cent measurement of the distance to $z = 0.35$. We update the reconstruction algorithm of Eisenstein et al. to account for the effects of survey geometry as well as redshift-space distortions and validate it on 160 LasDamas simulations. We demonstrate that reconstruction sharpens the BAO feature in the angle averaged galaxy correlation function, reducing the non-linear smoothing scale $\Sigma_{nl}$ from 8.1 to 4.4 Mpc $h^{-1}$. Reconstruction also significantly reduces the effects of redshift-space distortions at the BAO scale, isotropizing the correlation function. This sharpened BAO feature yields an unbiased distance estimate (<0.2 per cent) and reduces the scatter from 3.3 to 2.1 per cent. We demonstrate the robustness of these results to the various reconstruction parameters, including the smoothing scale, the galaxy bias and the linear growth rate. Applying this reconstruction algorithm to the SDSS LRG DR7 sample improves the significance of the BAO feature in these data from $3.3\sigma$ for the unreconstructed correlation function to $4.2\sigma$ after reconstruction. We estimate a relative distance scale $D_V/\alpha_{r_0}$ to $z = 0.35$ of 8.88 ± 0.17, where $r_0$ is the sound horizon and $D_V \equiv (D_A^2 H^{-1})^{1/3}$ is a combination of the angular diameter distance $D_A$ and Hubble parameter $H$. Assuming a sound horizon of 154.25 Mpc, this translates into a distance measurement $D_V(z = 0.35) = 1.356 \pm 0.025$ Gpc. We find that reconstruction reduces the distance error in the DR7 sample from 3.5 to 1.9 per cent, equivalent to a survey with three times the volume of SDSS.

Key words: cosmological parameters – cosmology: observations – dark energy – distance scale – large-scale structure of Universe.

1 INTRODUCTION

The baryon acoustic oscillation (BAO) method (see Weinberg et al. 2012 for a review) is a geometrical probe of the expansion rate of the Universe. Sound waves in the baryon–photon plasma are frozen as density fluctuations at recombination, with a characteristic scale set by the sound horizon (Sakharov 1966; Peebles & Yu 1970; Sunyaev & Zeldovich 1970; Bond & Efstathiou 1984, 1987; Hu & Sugiyama 1996; Hu, Sugiyama & Silk 1997; Eisenstein & Hu 1998). These sound waves manifest themselves as a peak in the matter (and therefore galaxy) correlation function at the scale of the sound horizon ($\sim 150$ Mpc for current concordance cosmologies) or equivalently as a series of oscillations in the power spectrum. Since the sound horizon is precisely calibrated by cosmic microwave background (CMB) measurements, BAO measurements may be used as a standard ruler (Tegmark 1997; Eisenstein, Hu & Tegmark 1998; Goldberg & Strauss 1998; Efstathiou & Bond 1999), mapping...
the angular diameter distance (with the ruler aligned perpendicular
to the line of sight) and the Hubble parameter (with the ruler parallel
to the line of sight) as a function of redshift (Blake & Glazebrook
2003; Hu & Jain 2003; Linder 2003; Seo & Eisenstein 2003).
There have now been multiple detections of the BAO feature in
galaxy surveys (Cole et al. 2005; Eisenstein et al. 2005; Blake
et al. 2007, 2011a,b; Padmanabhan et al. 2007; Percival et al. 2007,
2010; Gaztañaga et al. 2009a; Gaztañaga, Cabrê & Hui 2009b;
Kazin et al. 2010; Reid et al. 2010; Beutler et al. 2011; Sawangwit
et al. 2011; Chuang & Wang 2012; Ho et al. 2012; Seo et al. 2012)
and current (Hill et al. 2008; Schlegel, White & Eisenstein 2009;
Drinkwater et al. 2010) and planned (Laureijs et al. 2011; Schlegel
et al. 2011) BAO surveys are now a mainstay of experimental dark
energy programmes.
A key feature of the BAO method is the fact that the sound horizon
is much larger than the scales relevant for non-linear evolution and
galaxy formation. This scale separation protects the BAO feature
from large corrections due to these effects and therefore from sys-
tematic errors. There is now a considerable literature quantitatively
exploring these effects using both perturbative techniques and sim-
ulations (Meiksin, White & Peacock 1999; Seo & Eisenstein 2003;
Jeong & Komatsu 2006; Eisenstein et al. 2007b; Guzik, Bernstein
& Smith 2007; Huff et al. 2007; Angulo et al. 2008; Crocce &
Scoccimarro 2008; Seo et al. 2008, 2010; Smith, Scoccimarro &
Sheth 2008; Padmanabhan & White 2009; Mehta et al. 2011) and
consensus that systematic effects that might bias distance estimates
with the standard ruler are indeed small.

The dominant effect of the non-linear evolution of the density
field is to smooth the BAO feature in the correlation function. This
is equivalent to suppressing the oscillations in the power spec-
trum. While this smoothing does not bias the distance measure-
ments, it does reduce the contrast in the BAO feature and increase
the distance errors. This smoothing is well understood (Eisenstein
et al. 2007b; Crocce & Scoccimarro 2008; Matsubara 2008a,b;
Padmanabhan & White 2009; Seo et al. 2010) and is physically
caused by large-scale flows. This realization led Eisenstein et al.
(2007a) to suggest that this smoothing of the BAO feature may be
reversed, a process commonly referred to as ‘reconstruction’. They
provide a simple prescription for this process that has been shown
to sharpen the BAO feature and improve distance constraints (Noh,
White & Padmanabhan 2009; Padmanabhan, White & Cohn 2009;
Seo et al. 2010; Mehta et al. 2011). It is important to emphasize here
that this is not a deconvolution but rather uses information beyond
the two-point function that exists in the density field.

Although reconstruction has been studied with simulations in the
past, this paper represents the first application to data. The Sloan
Digital Sky Survey (SDSS) luminous red galaxy (LRG) sample
represents the current state of the art in low-redshift BAO measure-
ments and is a natural sample to implement reconstruction on. This
sample was analysed by Percival et al. (2010) using the power spec-
trum and by Kazin et al. (2010) using the correlation function, who
report an $\sim 3.5$ per cent distance measurement for the LRG sample
alone and an $\sim 2.7$ per cent measurement when combined with a
lower redshift sample of galaxies from the SDSS. This work is the
natural extension of these previous results. In addition to improving
the distance constraints from this sample, the performance of recon-
struction on the SDSS has important implications for the expected
performance of future surveys, most of which assume some level of
reconstruction.
This is the first in a series of three papers. This paper describes
the details of the reconstruction algorithm used, tests it on simulated
data and then presents the results for the Data Release 7 (DR7) data.

The second paper in this series (Xu et al. 2012, hereafter Paper II)
describes how we robustly fit the galaxy correlation function to
obtain our distance measurements. The third paper (Mehta et al.
2012, hereafter Paper III) presents the cosmological implications of
these measurements.

This paper is structured as follows. Section 2 introduces the
basic principle behind reconstruction and then presents a detailed
description of the algorithm implemented in this work. We describe
the data and simulations used in Section 3. We discuss the impact
of reconstruction on the BAO feature and the derived distances on
simulated data in Section 4; we then apply it to the data in Section 5.
We summarize our conclusions in Section 6.

2 METHODS

We describe the various algorithms used in the reconstruction of
the BAO feature below. We start by describing the physical basis for
reconstruction and outline the reconstruction algorithm. We then
describe our approach to dealing with survey boundaries, as well
as our procedure for estimating a distance scale from a correlation
function.

2.1 Understanding reconstruction

Fig. 1 highlights the key aspects of reconstruction. As was first
emphasized by Eisenstein et al. (2007b), the erasure of the BAO feature
can be physically traced to the pairwise relative velocities of particles
separated by $\sim 100 \text{ Mpc} h^{-1}$. Fig. 1 highlights this, showing a slice
through an $N$-body simulation, with tracer particles represent-
ing the BAO feature highlighted. The dominant smoothing of the
BAO feature is due to the coherent flows that form the large-scale
structure, not the random motions of particles within gravitational
structures. This is the key insight behind reconstruction – the same
galaxy surveys used to detect the BAO feature also map the cosmic
structure responsible for its erasure. One can therefore use these
same surveys to infer the large-scale flow field and partially undo
the smoothing of the BAO feature.

Fig. 1 also makes it clear that reconstruction is actually work-
ing at the level of the density field, using information beyond what
exists simply in the two-point statistics of the field. The gains of re-
construction therefore cannot be achieved by simply forward mod-
ing the correlation function into the non-linear regime. While one
might attempt to recover this information by considering higher or-
der statistics of the density field, this is an awkward encoding of the
information required.

The steps outlined in the figures are a simplification of the re-
construction algorithm we use; we describe the detailed algorithm
next.

2.2 A reconstruction algorithm

We implement an extended version of the reconstruction algorithm
of Eisenstein et al. (2007a). The theoretical underpinnings of this
algorithm have been described in Padmanabhan et al. (2009) and
Noh et al. (2009), and the algorithm has been validated against
several different suites of simulations (Seo et al. 2010; Mehta et al.
2011). We outline the steps of this algorithm below and discuss
details specific to our implementation in subsequent subsections.

(i) Estimate the unreconstructed power spectrum $P(k)$ or corre-
lation function $\xi(r)$.
Figure 1. A pictorial explanation of how density field reconstruction can improve the acoustic scale measurement. In each panel, we show a thin slice of a simulated cosmological density field. Top-left panel: in the early Universe, the initial densities are very smooth. We mark the acoustic feature with a ring of 150 Mpc radius from the central points. A Gaussian with the same rms width as the radial distribution of the black points from the centroid of the blue points is shown in the inset. Top-right panel: we evolve the particles to the present day, here by the Zel’dovich approximation (Zel’dovich 1970). The red circle shows the initial radius of the ring, centred on the current centroid of the blue points. The large-scale velocity field has caused the black points to spread out; this causes the acoustic feature to be broader. The inset shows the current rms radius of the black points relative to the centroid of the blue points (solid line) compared to the initial rms (dashed line). Bottom-left panel: as before, but overplotted with the Lagrangian displacement field, smoothed by a 10 h⁻¹ Mpc Gaussian filter. The concept of reconstruction is to estimate this displacement field from the final density field and then move the particles back to their initial positions. Bottom-right panel: we displace the present-day position of the particles by the opposite of the displacement field in the previous panel. Because of the smoothing of the displacement field, the result is not uniform. However, the acoustic ring has been moved substantially closer to the red circle. The inset shows the new rms radius of the black points (solid), compared to the initial width (long-dashed) and the uncorrected present-day width (short-dashed). The narrower peak will make it easier to measure the acoustic scale. Note that the algorithm applied to the data is more complex than was just described, but this figure illustrates the basic opportunity of reconstruction.

(ii) Estimate the galaxy bias \( b \) and the linear growth rate \( f \equiv \frac{d\ln D}{d\ln a} \sim \Omega_m^{0.55} \) (Carroll, Press & Turner 1992; Linder 2005), where \( D(a) \) is the linear growth function as a function of scale factor \( a \) and \( \Omega_m \) is the matter density relative to the critical density. We hold the values of \( b \) and \( f \) fixed in our analyses to fiducial values (described below) and demonstrate that our results are robust to changes in these adopted values.

(iii) Embed the survey into a larger volume, chosen such that the boundaries of this larger volume are sufficiently separated from the survey.

(iv) Gaussian smooth the density field.

(v) Generate a constrained Gaussian realization that matches the observed density and interpolates over masked and unobserved regions (Section 2.3).

(vi) Estimate the displacement field \( \Psi \) within the Zel’dovich approximation (Section 2.4).

(vii) Shift the galaxies by \(-\Psi\). Since linear redshift-space distortions arise from the same velocity field, we shift the galaxies by an additional \(-f(\Psi \cdot \hat{s}) \hat{s}\) (where \( \hat{s} \) is the radial direction). In the limit of linear theory (i.e. large scales), this term exactly removes redshift-space distortions (Kaiser 1987; Hamilton 1998; Scoccimarro 2004). We denote these points by \( D \).

(viii) Construct a sample of points randomly distributed according to the angular and radial selection function and shift them by \(-\Psi\). Since these points have not been observed, they are not affected by redshift-space distortions. We do not therefore apply the additional redshift-space distortion correction as with the galaxies. We denote these points by \( S \).

(ix) The reconstructed correlation function \( \xi \) is then given by the Landy–Szalay estimator (Landy & Szalay 1993):

\[
\xi = \frac{DD - 2DS + SS}{RR},
\]

where \( DD \), etc. are the number of pairs at a given separation between various sets of points. The random points \( R \) are distributed randomly according to the angular and radial selection functions; these are assumed to be different from those to generate \( S \). We weight the points by an approximate minimum variance weight (Feldman, Kaiser & Peacock 1994),

\[
w_i = \frac{1}{1 + \tilde{n}(z_i)P(k_0)},
\]

© 2012 The Authors, MNRAS 427, 2132–2145
Monthly Notices of the Royal Astronomical Society © 2012 RAS

Downloaded from https://academic.oup.com/mnras/article-abstract/427/3/2132/1098449
by guest
on 26 July 2018
where \(\bar{n}\) is the redshift distribution at the galaxy redshift \(z\), and \(P(k_0) = 40,000 \text{ (Mpc}^{-1}\text{)}^3\) is the power spectrum approximately at the BAO scale for SDSS LRGs.

We note that we choose to use \(RR\) as opposed to \(SS\) in the denominator of the Landy–Szalay estimator. One can derive this form by noting that the reconstructed density field is the difference between the density defined by the shifted galaxies \(D\) and that defined by the shifted randoms \(\delta\). In both cases, the randoms that define the geometry are the same, yielding the form of the equation we use. However, we have also explicitly tested our results using \(SS\) in the denominator and found that none of our results were affected.

### 2.3 Generating constrained realizations

Since the gravitational potential (and therefore, the displacement) depends non-locally on the matter overdensity, it is sensitive to regions of space either masked out by the survey or not surveyed at all. To handle this complication, we embed the survey in a larger region where the boundaries of this larger region are sufficiently far from the true survey (in what follows below, we pad by 200 Mpc \(h^{-1}\)). Since the density field on large scales can be approximated by a Gaussian density field, the problem of ‘filling’ in the missing regions is then equivalent to the problem of generating constrained realizations of a Gaussian density field (Hoffman & Ribak 1991; see also Zaroubi et al. 1995, who demonstrate that the previous algorithm is equivalent to a Weiner filtering of the density field). Strictly speaking, one should marginalize over the ensemble of such realizations. However, as we demonstrate below, our results are insensitive to the details of this implementation. Therefore, for simplicity, we consider only a single realization.

We start by organizing the observed and constrained realization density fields into an \(N_{\text{obs}}\) element vector \(\delta\) and an \(N_{\text{embed}}\) element vector \(\tilde{\delta}\), related by the trivial projection \(P\) on to the observed points, \(\delta = P\tilde{\delta}\). For the analyses below, \(N_{\text{embed}}\) is a 512\(^3\) element grid, with a grid spacing of 4.7 Mpc \(h^{-1}\). The Hoffman–Ribak algorithm is then

\[
\delta = \delta_0 + CC^{-1}(\tilde{\delta} - P\delta_0),
\]

where \(\delta_0\) is an unconstrained Gaussian realization, with an assumed power spectrum \(P(k)\). The covariance matrices \(C\) and \(\tilde{C}\) are defined by the correlation function between pairs of pixels,

\[
C_{ij} \equiv \langle \delta(r_i)\delta(r_j) \rangle,
\]

where the correlation function is just the Fourier transform of the power spectrum. Note that \(C\) is an \(N_{\text{obs}} \times N_{\text{obs}}\) matrix and operates only on the observed pixels, while \(\tilde{C}\) has dimensions \(N_{\text{embed}} \times N_{\text{obs}}\) and relates the observations and constrained realization.

While this algorithm is straightforward in principle, a number of comments are in order regarding its implementation. The first concerns our choice of power spectrum for generating the constrained realization and covariance matrices. Unlike the algebraically similar problem of power spectrum estimation, we do have a measurement of the power spectrum, which we use. A related issue is that in redshift space, the pair correlation function is not isotropic and translation invariant. We ignore this subtlety for simplicity and use an isotropic correlation function. We demonstrate later that, for the SDSS DR7 geometry, the reconstructed correlation function is robust to these choices. Finally, we note that our prior power spectrum also includes a white noise component of amplitude \(\bar{n}^{-1}\).

The second issue is computational. The dimensions of \(\delta\) and \(\tilde{\delta}\) are \(N \sim O(10^6)\), making direct manipulation impossible. To proceed, we use the fact that multiplication by \(C\) is equivalent to a convolution by the correlation function (Padmanabhan, Seljak & Pen 2003). Since we have ignored the angular dependence of the power spectrum, we can implement this in \(O(N \log N)\) time using FFTs. The matrix inverse operations are implemented using a preconditioned conjugate gradient algorithm (Press et al. 1992) with the preconditioner being the convolution by the Fourier transform of \(1/P(k)\). For an unmasked survey, this preconditioner is the exact inverse. The above allows us to generate constrained realizations in \(O(10)\) iterations.

### 2.4 Solving for the displacement

To linear order, the displacement \(\Psi\) can be related to the density in redshift space by (Nusser & Davis 1994)

\[
\nabla \cdot F = \nabla \cdot \Psi + f \nabla \cdot (\Psi \delta) = -\frac{\delta_{\text{gal}}}{b},
\]

where \(\Psi = \Psi \cdot \delta\) is the displacement in the redshift direction and \(\delta_{\text{gal}}\) is the galaxy overdensity. Assuming that \(\Psi\) is irrotational, we write \(\Psi = \nabla \phi\). Substituting this above, we obtain

\[
\nabla^2 \phi + f \nabla \cdot (\nabla \phi \delta) = -\frac{\delta_{\text{gal}}}{b}.
\]

This equation resembles Poisson’s equation with an additional term for redshift-space distortions. However, the redshift-space term breaks the translational invariance of the problem and prevents us from solving this with FFTs.\(^1\) We solve this equation by converting all the derivatives to their finite difference counterparts and solve the resulting linear equation. We implement this using the parallel GMRES algorithm in the\textsc{petsc} (Balay et al. 1997, 2011a,b) toolkit. Having computed \(\phi\), we obtain the displacement field by finite differences. An advantage of this formulation is that the algorithm is easily extended to non-Cartesian coordinate systems, although we do not use this feature in this work.

### 2.5 Fitting the acoustic feature

We briefly describe our fitting procedure below; a detailed description and tests of this procedure is in Paper II. We define a fiducial fitting model:

\[
\xi_{\text{fit}}(r) = B^2 \xi_{\text{in}}(ar) + A(r),
\]

where

\[
\xi_{\text{in}}(r) = \int \frac{k^2dk}{2\pi^2}P_m(k)j_0(kr)e^{-k^2a^2r^2}
\]

and

\[
A(r) = \frac{a_1}{r^2} + \frac{a_2}{r} + a_3.
\]

The damping factor \(e^{-k^2a^2r^2}\) improves the convergence of the correlation function integral; we fix \(a = 1\text{ Mpc}^{-1}\) throughout, which does not affect the correlation function on any of the scales of interest.

Our template power spectrum \(P_m(k)\) (equations 7 and 8) is determined by interpolating between the linear theory power spectrum and one with the BAO feature erased (Eisenstein et al. 2007b):

\[
P_m(k) = \{P_{\text{lin}}(k) - P_{\text{smooth}}(k)\}e^{-k^2\xi_{\text{lin}}^2/2} + P_{\text{smooth}}(k).
\]

For convenience, we choose to normalize this template to the observed correlation function at \(r = 50\text{ Mpc}^{-1}\); this ensures \(B^2 \sim 1\).

\(^1\) Note that in the plane-parallel approximation, translational invariance is restored.
The $\Sigma_{\text{nl}}$ parameter smooths the BAO feature, modelling the degradation due to non-linear structure growth. The $e^{-k^2a^2}$ term is used to damp the oscillatory transform kernel $j_0(kr)$ at high $k$ to induce better numerical convergence in the integration. The $A(\alpha)$ term, with the associated $a_{1,2,3}$ nuisance parameters, is used to help marginalize out the unmodelled broad-band signal in the correlation function. This broad-band signal includes redshift-space distortions, scale-dependent bias and any errors we make in our assumption of the model cosmology which might bias the acoustic peak.

Our distance constraints are captured by the scale dilation parameter $\alpha$ which represents the shift in the acoustic peak. An $\alpha > 1$ indicates a shift towards smaller scales and an $\alpha < 1$ indicates a shift towards larger scales.

We obtain the best-fitting value of $\alpha$ by minimizing the $\chi^2$ goodness-of-fit indicator, varying the distance scale $\alpha$, the galaxy bias $B$ and the nuisance parameters in $A(\alpha)$. A detailed description of this is in Paper II.

In order to compute the likelihood function $p(\alpha)$, given the measured correlation function, we start with

$$\chi^2(\alpha) = [d - m(\alpha)]^T C^{-1} [d - m(\alpha)], \quad (11)$$

where $d$ is the measured correlation function and $m(\alpha)$ is the best-fitting model at each $\alpha$. We use a fiducial fitting range of 30–200 h$^{-1}$ Mpc. $C$ is a modified Gaussian covariance matrix described in detail in Paper II. We minimize this function for a grid of fixed values of $\alpha$; this step is exactly equivalent to marginalizing over the linear parameters $a_i$, and is a good approximation for the $B^2$ term as well. The likelihood distribution of $\alpha$ is then simply

$$p(\alpha) \propto e^{-\chi^2/2}, \quad (12)$$

where $\chi^2$ is the $\chi^2$ minimum at $\alpha$, and the proportionality constant is determined by ensuring that the probability integrates to 1. In addition, we impose a 15 per cent Gaussian prior on $\log(\alpha)$. While this prior does not affect the core of the probability distribution, it does suppress values of $\alpha < 1$ that correspond to the BAO scale being shifted to very large scales, reflecting our current state of knowledge about the background cosmology.

This probability distribution captures all the distance information of these data and is what we use when estimating cosmological parameters in Paper III. It is however convenient to summarize this information. We do so throughout this paper by reporting the mean $\langle \alpha \rangle$ and standard deviation $\sigma_\alpha$ for the data and each individual mock catalogue, where $\sigma_\alpha^2 = \langle \alpha^2 \rangle - \langle \alpha \rangle^2$ and

$$\langle \alpha^2 \rangle = \int d\alpha \, p(\alpha) \alpha^2. \quad (13)$$

As we see below for the data and is discussed in greater detail in Paper II, $p(\alpha)$ is well approximated by a Gaussian and therefore completely characterized by the mean and standard deviation. It is also worth emphasizing that this procedure yields an independent measurement of the distance error for every mock, capturing the effects of sample fluctuations. Our analyses must therefore not just explore the distribution of $\alpha$ but also of $\sigma_\alpha$ (see below).

3 DATA AND SIMULATIONS

3.1 The LRG sample

The SDSS (York et al. 2000) has imaged $\sim 10,000$ deg$^2$ of the sky, and obtained spectra of nearly a million of the detected objects. The imaging was carried out by drift scanning the sky in photometric conditions (Hogg et al. 2001) in the ugriz bands (Fukugita et al. 1996; Smith et al. 2002) with the Apache Point 2.5 m telescope (Gunn et al. 2006) using a specially designed wide-field camera (Gunn et al. 1998). These data were processed by completely automated pipelines that detect and measure the photometric properties of the objects and astrometrically and photometrically calibrate these observations (Pier et al. 2003; Ivezić et al. 2004; Tucker et al. 2006; Padmanabhan et al. 2008). Subsamples from the resulting photometric samples were selected (Eisenstein et al. 2001; Strauss et al. 2002) for spectroscopy using a 640-fibre spectrograph. The SDSS has had three phases: SDSS-I and SDSS-II completed the observations described above in 2009 and the data were released in a series of seven DRs, with the SDSS DR7 (Abazajian et al. 2009) being the final DR. The third phase of SDSS (SDSS-III; Eisenstein et al. 2011) started taking data in 2009 and includes the Baryon Oscillation Spectroscopic Survey (BOSS; Schlegel et al. 2009) as part of its science goals.

The sample that we consider for this analysis is the LRG sample. The motivation and selection of this sample is described in detail in Eisenstein et al. (2001) and we refer the interested reader to the description there. These galaxies are very luminous and therefore can probe cosmologically interesting volumes. Furthermore, they are generally old stellar systems with very uniform spectral energy distributions, characterized by a strong break at 4000 Å. This gives these galaxies a distinct colour–flux–redshift relation, which allows them to be uniformly selected over a broad redshift range. The LRG samples have been used for a number of cosmological analyses in the SDSS, including the BAO detection in the SDSS (Eisenstein et al. 2005; Blake et al. 2007; Padmanabhan et al. 2007; Percival et al. 2007, 2010; Gaztanaga et al. 2009a,b; Kazin et al. 2010; Reid et al. 2010; Sawangwit et al. 2011; Chuang & Wang 2012). We use a sample identical to that used in Kazin et al. (2010) and refer the reader there for a detailed description of its construction.

Fig. 2 shows the sky coverage of the sample that we consider. While the SDSS-II survey has data in both the northern and southern Galactic caps, we focus on the contiguous footprint in the north, with a total area of 7189 deg$^2$. Fig. 3 plots the number density of the LRG sample as a function of redshift. We truncate the redshift distribution at $z = 0.16$, since the colour selection of the LRG sample breaks down there. The redshift distribution is approximately constant out to a redshift of $z \sim 0.35$, falling off due to the magnitude limit of the sample beyond.

3.2 LasDamas mock galaxy catalogues

We test our implementation of the reconstruction algorithm on mock galaxy catalogues created from the Large Suite of Dark Matter...
Simulations (LasDamas; McBride et al., in preparation). Our goals are (a) to demonstrate that reconstruction yields an unbiased and improved distance scale measurement, (b) to test the robustness of the reconstruction algorithm to its input parameters and (c) to tune these parameters in a ‘blind’ manner.

We use the publicly available mock galaxy catalogues constructed by the LasDamas Collaboration, which are designed to simulate the SDSS LRG samples. We chose the gamma release of lrgFull (in the LasDamas nomenclature) which best corresponds to our data sample. LasDamas assumes a flat CDM cosmology, roughly consistent with the 5-year Wilkinson Microwave Anisotropy Probe (WMAP5), with \( \Omega_{\text{baryon}} = 0.04 \), \( \Omega_{\text{matter}} = 0.25 \), \( \Omega_{\Lambda} = 0.75 \), \( h = 0.7 \), \( n_s = 1 \) and \( \sigma_8 = 0.8 \). The LRG mocks were constructed from 40 ‘Oriana’ N-body realizations, which were each run in a large cubical volume (\( L = 2.4 h^{-1} \) Gpc) with 1280^3 particles and initialized using second-order Lagrangian perturbation theory at \( z = 49 \). The mock galaxy catalogues were constructed by populating dark matter halo catalogues, where the halo occupation parameters were varied to match observed galaxy clustering measurements. The mock galaxy catalogues modelled the realism of observed data by modelling redshift distortions, matching the angular selection function (survey boundaries and missing data within the survey due to bright stars, etc.) of the SDSS DR7 LRG sample and spanning the redshift range of \( z = 0.16–0.44 \). We make one modification to the public catalogues to better match our analysis: we down sample the radial selection function to match our specific LRG data. This is necessary to properly model the galaxy numbers in the flux-limited region of the LRG selection at redshifts greater than \( z = 0.36 \). We make use of the catalogues covering only the northern galactic cap of the SDSS footprint (7214.34 deg^2), which yield four mocks from each simulation for a total of 160 galaxy catalogues.

### 3.3 Fiducial cosmologies

We conclude this section by discussing two technical details – our adopted definition of the sound horizon and the fiducial cosmology assumed in our analyses. We follow Eisenstein & Hu (1998) and assume the sound horizon specified by equation (6) of that paper and defer a comparison of the different choices to Paper II. We use two fiducial cosmologies in our analyses below; while both of these are flat ΛCDM cosmologies, they differ in their choices of parameters.

For the LasDamas simulations, we use the cosmology assumed for the simulations – a baryon density of \( \Omega_b = 0.04 \), a matter density of \( \Omega_m = 0.25 \) and a Hubble constant of 70 km s^{-1} Mpc^{-1} (\( h = 0.7 \)). However, these choices differ significantly from the current best fit to the CMB data from the WMAP satellite (Komatsu et al. 2011, hereafter WMAP7), with \( \Omega_b = 0.0457 \), \( \Omega_m = 0.274 \) and \( h = 0.702 \). In particular, the sound horizon assuming the WMAP7 cosmology is 152.76 Mpc, compared to 159.68 Mpc for the LasDamas cosmology, a difference larger than our claimed statistical accuracy. While, as we discuss below, an incorrect cosmology does not bias our distance scale, it does change the errors on the distance scale. This makes it important to iterate and choose a cosmology close to the best fit in the analysis. As we demonstrate below, the data are well fitted by the WMAP7 cosmology and we use it as our fiducial model when analysing the SDSS DR7 data.

### 4 RECONSTRUCTING SIMULATIONS

We start by discussing the impact of reconstruction on the LasDamas simulations. The top-left panel of Fig. 4 is the unreconstructed real-space correlation function, with the BAO ring clearly visible and the correlation function independent of angle. Turning on redshift-space distortions destroys the isotropy of the correlation function, with the maximal distortion, as expected, parallel to the line of sight. The distortions at small \( r \) and \( r_1 < 20 \) Mpc h^{-1} are due to virial motions inside haloes, the so-called Fingers of God (FoG). The bottom panels show the correlation functions after reconstruction, assuming a 15 Mpc h^{-1} smoothing length to estimate the displacement field and the true values of the logarithmic growth rate \( f \) and the galaxy bias \( b \). The most relevant feature for this paper is the enhanced BAO signal, apparent from the increased contrast of the BAO ring. Equally striking is the restored isotropy of the redshift-space correlation function, demonstrating that reconstruction is correcting for the large-scale redshift-space distortions. The breakdown on small scales is due to a combination of the fact that the model for the displacement field is based on linear theory and that reconstruction is imperfect on these scales. We also note that the FoG become more prominent, highlighting the tendency of reconstruction to blow up collapsed objects. It is important to emphasize that both these effects are restricted to small scales and have no effect on the acoustic scale.

We compress the 2D correlation functions by averaging over angle; the resulting correlation functions (both in real and redshift space) are in Fig. 5. In both cases, we observe the BAO feature sharpened after reconstruction. In the case of the redshift-space distortions, the overall amplitude of the correlation function is also reduced due to the removal of linear redshift-space distortions. We find that the reconstructed redshift-space correlation function does not match its real-space counterpart on small scales, indicating that the linear theory correction is breaking down on these scales. However, the agreement on large (\( r > 30 \) Mpc h^{-1}) scales is striking.

Fig. 5 has a useful interpretation as a redistribution of pairs of galaxies across different scales. Recall that \( r^2 \xi \) is proportional to the number of excess pairs (over a random distribution) in a spherical shell of width \( dr \) centred at \( r \). Since reconstruction does not change the total number of pairs but merely redistributes them over different scales, the area under these curves must be conserved. Comparing the correlation functions before and after reconstruction captures this redistribution of pairs. There are two effects worth noting, both of which are more easily noted in the real-space case. The first is a transferring of pairs from small scales (\( r < 20 \) Mpc h^{-1}) to intermediate scales (\( r \sim 50 \) Mpc h^{-1}), apparent in the fact that the

---

**Figure 3.** The redshift distribution of the DR7 LRG sample used in this paper. The dashed (red) line is a smooth fit to the redshift distribution used in the determination of the weights used in the correlation function.
Figure 4. The LasDamas galaxy correlation function, averaged over the 160 simulations, as a function of the separation perpendicular ($\perp$) and parallel ($||$) to the line of sight. The correlation functions have been scaled by $r^2$ to highlight the BAO feature. The top panels show the unreconstructed correlation functions, while the bottom panels show the reconstructed correlation functions; the left- and right-hand panels are real and redshift space, respectively. The BAO feature is visible as a ring at $\sim110\,\text{Mpc}\,h^{-1}$ in the top-left panel. Redshift-space distortions destroy the isotropy of the correlation function (top-right panel). Reconstruction both sharpens the BAO feature (highlighted in the bottom-left panel) and restores the isotropy (bottom-right panel) of the correlation function on the BAO scale.

Figure 5. Left-hand panel: the angle-averaged correlation function in real space, before (red circles) and after (blue squares) reconstruction and averaging over the 160 LasDamas simulations. The reconstruction algorithm assumes the default parameters described in the text. The acoustic feature is clearly sharpened after reconstruction. Right-hand panel: same as the left-hand panel, except in redshift space. Also shown for comparison is the average reconstructed real-space correlation (dashed line). In addition to sharpening the acoustic feature, the reconstruction algorithm also reduces the effects of redshift-space distortions on the correlation function.

unreconstructed correlation function is larger on small scales, with the trend reversed on intermediate scales. This is reconstruction reversing the infall of galaxies into overdensities. The second is that the unreconstructed correlation function is higher just before the BAO feature, due to pairs flowing out of the BAO feature. These flows are responsible for the smoothing of the BAO feature. The fact that the reconstructed correlation function is lower just before the BAO feature and then higher at the BAO peak is from the fact that reconstruction has moved these objects back into the BAO ring.

One metric to quantify the degree of reconstruction is to compare the values of $\Sigma_{nl}$ (see equation 10) before and after reconstruction. While $\Sigma_{nl}$ is poorly constrained in any single simulation, we can...
fit the average of the simulations before and after reconstruction. We find that $\Sigma_d$ decreases by close to 50 per cent from 8.1 to 4.4 Mpc $h^{-1}$. This improvement is in line with theoretical estimates (Padmanabhan et al. 2009) and corresponds well with assumptions made for future surveys.

Figs 6 and 7 quantify the impact of reconstruction on the inferred distance $\alpha$ in real and redshift space, respectively. Recall that $\alpha$ is the estimated distance relative to a fiducial distance, which, in the case of simulations, we choose to be the true comoving distance to the median redshift of the survey $z \sim 0.35$. Both the unreconstructed and reconstructed simulations yield unbiased distance estimates (i.e. $\langle \alpha \rangle = 1$), and the distances before and after reconstruction are clearly correlated with one another. Reconstruction, however, reduces the scatter in $\alpha$ from 3.0 to 2.0 per cent in real space and from 3.3 to 2.1 per cent in redshift space, an improvement of between a factor of 1.5 and 1.7. These figures also demonstrate that reconstruction noticeably reduces the number of outliers in the distance estimate, a direct effect of the increased significance of the BAO feature.

Before continuing, it is worth noting that the BAO scale in galaxies is expected to be biased at the $\sim 0.5$ per cent level (or lower) (Padmanabhan & White 2009; Mehta et al. 2011). These biases are understood to arise from second-order terms in perturbation theory (Crocce & Scoccimarro 2008; Padmanabhan & White 2009) and are expected to be reduced by reconstruction (Noh et al. 2009; Padmanabhan et al. 2009; Mehta et al. 2011). However, the amplitude of these effects are much below what we can expect to observe with a single SDSS-sized data set and are expected to be only marginally detectable with the full ensemble of simulations. Table 1 verifies this expectation, demonstrating that any bias in the distance scale is less than 0.2 per cent, much below our statistical precision. We therefore ignore these effects in all our subsequent analyses but note that they will become relevant as the statistical precision increases.

The above distance accuracies are ensemble averages. Given the still relatively low significance of the detection of a BAO feature, statistical fluctuations may make the BAO feature more or less prominent. Since our fitting procedure marginalizes out smooth components in the correlation function, a less prominent BAO feature would result in a significantly degraded distance measurement. It is therefore interesting to quantify the effect of reconstruction not just on the ensemble distance accuracies, but also on the distance accuracy estimated for each individual simulation. Figs 8 and 9 summarize this information for real and redshift space, respectively. The median errors agree well with the errors estimated from the ensemble distance measurements, confirming the validity of our

### Table 1. A summary of the effect of reconstruction on the distance estimates, in real and redshift space based on the 160 LasDamas mock catalogues. The $(\alpha - 1)$ and $\sigma_\alpha$ numbers are in per cent. For all these cases, we reconstruct using our default choices of parameters. The second column is the bias in the distance in per cent. The third column is the error in the distance estimated from the ensemble of the LasDamas simulations, while the fourth column is the median of the errors estimated per simulation, with the agreement between the two testing our error estimates. The last column shows the number of cases where the error in the distance is greater than 7 per cent. In all of these metrics, reconstruction improves the precision of the distance estimates.

| Type               | $\langle \alpha - 1 \rangle$ | $\sigma_\alpha$ | $\delta_\alpha$ | $n(\sigma_\alpha > 7\%)$ |
|--------------------|-------------------------------|-----------------|-----------------|--------------------------|
| Real, unrecon      | 0.1                           | 3.0             | 3.1             | 5                        |
| Real, recon        | -0.2                          | 2.0             | 2.2             | 0                        |
| Redshift, unrecon  | -0.1                          | 3.3             | 3.3             | 8                        |
| Redshift, recon    | 0.1                           | 2.1             | 2.3             | 0                        |

### Figure 8. The error on the distance scale before and after reconstruction, as individually estimated for each LasDamas simulation in real space. The short dashed line has slope 1; the fact that most points lie below the line demonstrates the efficacy of reconstruction. The horizontal and vertical lines mark the median error before and after reconstruction. Note also the large scatter in the estimated errors, due to the still relatively low significance of the BAO detection in the DR7 survey volume.
error estimates. We also observe that, for the majority of the simulations (∼98 per cent), reconstruction reduces the distance error. For the remaining ∼2 per cent, we find that the errors are very similar to those obtained before reconstruction. Furthermore, a number of these are cases where the distance scale itself is poorly measured. Paper II discusses these cases in more detail. Table 1 summarizes the above discussion, considering the recovered distances before and after reconstruction.

4.1 Robustness to reconstruction parameters

All of the above has assumed our default choices of parameters for reconstruction – a smoothing scale of 15 Mpc $h^{-1}$, the measured value of the galaxy bias ($b = 2.2$) and the input value of $f$. We also assume a concordance cosmology power spectrum in order to generate the constrained Gaussian realization. We explore the effects of varying these below. We use the simulations to determine the optimal smoothing scale and demonstrate that reconstruction is robust to variations in the other parameters.

The most important of the reconstruction parameters is the scale used to smooth the density field before estimating the potential. For scales too small, the fidelity of the potential reconstruction is affected by noise from the finite numbers of galaxies. At the other extreme, oversmoothing reduces the estimated displacements (an infinite smoothing scale leaves the galaxies at their original positions) and reduces the effectiveness of reconstruction. In the limit of a high number density of galaxies, the above argument suggests choosing a smoothing scale as small as possible; however, as argued in Eisenstein et al. (2007b), the bulk of the erasure of the BAO feature comes from velocities on scales between $k = 0.02$ and $0.2 \, h \, \text{Mpc}^{-1}$, suggesting that one reaches diminishing returns for scales smaller than $\sim 5 \, \text{Mpc} \, h^{-1}$. The number density of the LRG sample is $\leq 10^{-4} \, h^3 \, \text{Mpc}^{-3}$, which implies that the shot-noise power spectrum crosses at a scale of $\sim 0.15 \, h \, \text{Mpc}^{-1}$, suggesting smoothing on scales larger than $\sim 10 \, \text{Mpc} \, h^{-1}$.

Fig. 10 plots the reconstructed 2D correlation functions for different choices of the smoothing scale. For large smoothing scales, the degree of reconstruction is clearly degraded, although even in these cases, the BAO feature is still enhanced relative to the case of no reconstruction. This emphasizes the fact that it is large-scale flows that primarily shape the BAO feature.
that are responsible for the erasure of the BAO feature; even a large smoothing scale can effectively reverse these. At the other extreme at 10 Mpc h⁻¹, we find that reconstruction can strongly distort the correlation function on small scales. A useful picture to understand the distortions at small perpendicular separation is to remember that reconstruction effectively repels close pairs of particles. Since the smoothing scale averages the shot-noise in the input density field, insufficiently smoothing the field pushes apart noise fluctuations with an additional enhancement in the line-of-sight direction (generating its own FOG).

The second feature is the excess at r⊥ = 0. An examination of Fig. 4 shows that traces of this feature exist in the real-space correlation function as well. Unlike the redshift-space case, the reconstruction procedure in real space does not have a preferred direction. This feature can be traced back to our definition of the radial selection function. Since we do not know the mean density of the galaxies as a function of redshift, we simply define it by randomly resampling the observed galaxy redshifts. This suppresses radial density fluctuations and can create features in the transverse direction. As before, increasing the smoothing scale reduces both of these.

Fig. 11 plots the angle-averaged correlation function. We find that the correlation function is strongly distorted on small scales for the 10 Mpc h⁻¹ smoothing scale; increasing the smoothing scales removes these distortions. Given our desire to choose the smallest smoothing scale possible, we adopt 15 Mpc h⁻¹ as our fiducial choice. Table 2 summarizes the distance constraints as a function of smoothing scale and shows the same trends discussed above.

While the smoothing scale is the most important parameter input to the reconstruction algorithm, there are a number of other inputs – the galaxy bias b, the logarithmic growth rate of structure f and the power spectrum used for the constrained realizations. The galaxy bias and the input power spectrum are constrained by measurements of the unreconstructed correlation function, while f ∼ Ω₀M can be constrained by CMB measurements. While we adopt these fiducial values for our measurements, an immediate question is how sensitive reconstruction is to these choices. Table 3 summarizes the impact of changing these values. We vary the bias and f by ±20 per cent, and we consider two different choices for a fiducial power spectrum – a power spectrum with the BAO features erased (Eisenstein & Hu 1998) and a constant P(k) = 10⁴(Mpc h⁻¹)³ (corresponding to no clustering signal and only shot-noise). Of these, the bias has the strongest effect on reconstruction. An overestimate of the bias results in an underestimate of the density field, reducing the derived displacement field. On the other hand, underestimates of the bias result in an overestimate of the density field and overcorrects the displacements. While the distance scale obtained is still unbiased, Table 3 demonstrates that including the uncertainty in the galaxy bias in reconstruction would increase the statistical uncertainty of the measurement. However, as Table 3 also shows, even for large variations in the bias (recall that a 20 per cent uncertainty in the bias corresponds to a 40 per cent uncertainty in the normalization of the correlation function, much larger than the uncertainty in most estimates of the bias), this additional scatter is subdominant to our errors. We also observe that the impact of misestimating the bias on the inferred error in the mocks is much less than our quoted statistical errors.

One of the most prominent effects of reconstruction was the restoration of the isotropy of the correlation function. Since this depends on the choice of f, an immediate question is how sensitive reconstruction is to the particular choice. Fig. 12 plots the 2D correlation function with f = 0. This turns off both the modifications to the continuity equation in redshift space and the additional line-of-sight displacements that correct for the redshift-space distortions. We find that the BAO feature is still improved, although the correlation function is still strongly distorted from isotropy. Less drastic variations in the value of f (although larger than our current errors on Ω₀M) are in Table 3. Even more strongly than in the case of the bias, we find that this additional error is much smaller than our statistical error, demonstrating the robustness of reconstruction.

![Figure 11](image-url)  
Figure 11. The averaged reconstructed correlation functions in redshift space for the LasDamas simulations, as a function of the reconstruction smoothing scale. Also plotted for reference (dashed line) is the unreconstructed correlation function. Our fiducial smoothing length is 15 Mpc h⁻¹.

### Table 2

Analogous to Table 1, except as a function of the smoothing scale input to the redshift-space reconstruction. As before, the (α - 1) and σ_u numbers are in per cent. In all cases, the distance is unbiased, but as expected, we find the error increasing as a function of the smoothing scale. Reconstruction after smoothing at 10 Mpc h⁻¹ distorts the shape of the correlation function, and our model is no longer a good fit to the shape. We use 15 Mpc h⁻¹ as our fiducial smoothing scale.

| Smoothing (Mpc h⁻¹) | (α - 1) | σ_u | σ_u' | m(σ_u > 7 per cent) |
|---------------------|---------|------|------|---------------------|
| 15                  | 0.1     | 2.1  | 2.3  | 0                   |
| 20                  | 0.4     | 2.3  | 2.5  | 0                   |
| 25                  | 0.6     | 2.6  | 2.6  | 0                   |

### Table 3

The impact of changing the values of the galaxy bias, f and the input power spectrum for the constrained Gaussian from their fiducial values.

| Case | σ(Δα) | r | σ(Δσ_u) | r |
|------|-------|---|---------|---|
| b = 1.8 (~20 per cent) | 0.9 | 0.92 | 0.3 | 0.94 |
| b = 2.6 (+20 per cent) | 0.6 | 0.97 | 0.2 | 0.97 |
| f = 0.5 (~20 per cent) | 0.2 | 1.00 | 0.1 | 1.00 |
| f = 0.8 (+20 per cent) | 0.2 | 1.00 | 0.1 | 1.00 |
| P(k), no wiggle | 0.0 | 1.00 | 0.0 | 1.00 |
| P(k), shot-noise | 0.3 | 0.99 | 0.1 | 0.99 |
Finally, we consider varying the inputs to the constrained realizations. We consider two alternative power spectra—one with the BAO feature erased and the other with no clustering signal—and we find that reconstruction is robust to these choices as well. These results are dependent on the geometry of the survey and could possibly change for different geometries.

5 RECONSTRUCTING DATA

We now apply reconstruction to the DR7 data set. Our fiducial choice of parameters is a 15 Mpc \( h^{-1} \) smoothing length, a galaxy bias of 2.2 and the WMAP7 cosmology. Fig. 13 plots the angle-averaged DR7 correlation function before and after reconstruction. We find that reconstruction on the DR7 data demonstrates the same features seen in the LasDamas simulations. The amplitude of the intermediate-scale correlation function decreases due to the correction of redshift-space distortions, while the transition into the BAO feature at \( \sim 80–100 \) Mpc \( h^{-1} \) is sharpened.

The correlated nature of the errors makes it difficult to quantitatively assess the impact of reconstruction on these data. Fig. 14 plots the \( \chi^2 \) surface for \( \alpha \) both before and after reconstruction. We note that the \( \chi^2 \) minimum after reconstruction is visibly narrower, indicating an improvement in the distance constraints. This improvement is also summarized in the first two lines of Table 4 which shows that reconstruction reduces the distance error from 3.5 to 1.9 per cent. These distance constraints are also consistent with the errors estimated from the LasDamas simulations.

Fig. 14 also plots the \( \chi^2 \) surface for a template without a BAO feature, using the ‘no-wiggle’ form of Eisenstein \& Hu (1998). The lack of a well-defined minimum either before or after reconstruction indicates that our distance constraints are indeed coming from the presence of a BAO feature and not any broad-band features in the correlation function. The difference in \( \chi^2 \) between the templates with and without a BAO feature also provides an estimate of the significance of the BAO detection in these data. Reconstruction improves this detection significance from 3.3\( \sigma \) (consistent with previous measurements) to 4.2\( \sigma \). This is not the only measure of the detection significance possible; Paper II discusses these in more detail.

As before, we would like to demonstrate the robustness of the results to the various parameters of the reconstruction algorithm. Table 4 lists the recovered distances varying the smoothing scale, input bias, growth rate (\( f \)) and prior power spectrum; for each of these cases, we recover distances consistent with the fiducial choices of parameters.

Our final test is the impact of the assumed fiducial cosmology. We consider two cases in Table 4: flat \( \Lambda \)CDM cosmologies with \( \Omega_m = 0.2 \) and 0.35. In both of these cases, we adjust the Hubble constant and the baryon density \( \Omega_b \) to keep the physical densities \( \Omega_b h^2 \) and \( \Omega_m h^2 \) equal to their WMAP7 values. This prescription leaves the CMB unchanged, but alters the distance–redshift relation. We find that the estimated values of \( \alpha \) are significantly different from the fiducial case. However, note that the physical observable is not \( \alpha \), but \( D_{\text{V}}(r_s) = \alpha(D_{\text{V}}/r_s)_{\text{fid}} \). Comparing this across the three cosmologies (second column, Table 4), we find it insensitive to the choice of cosmology.

The distance information from these BAO measurements may be summarized into a probability distribution \( p(D_{\text{V}}/r_s) \), plotted in Fig. 15 and summarized in the second column of Table 4. Unlike \( \alpha \), these measurements no longer make reference to a fiducial cosmology. One may however freely convert between \( p(\alpha) \) and \( p(D_{\text{V}}/r_s) \) by multiplying the latter by \( (D_{\text{V}}/r_s)_{\text{fid}} \). We use the results in Fig. 15 to explore the cosmological consequences of these measurements in Paper III. If we assume a perfectly measured sound horizon, these measurements can be converted into a distance measurement in Gpc. Using a sound horizon of 152.76 Mpc, we get a distance to \( z = 0.35 \) of 1.356 \( \pm \) 0.025 Gpc. Note that these numbers do

Figure 12. The 2D reconstructed correlation function, but without the redshift-space distortion corrections (i.e. setting \( f = 0 \) in the reconstruction algorithm). While the BAO feature is more prominent, redshift-space distortion still strongly distorts the BAO feature.

Figure 13. The unreconstructed (left) and reconstructed (right) DR7 angle-averaged correlation function. The error bars are the standard deviation of the 160 LasDamas simulations. These errors are however highly correlated from bin to bin and therefore no conclusions as to significance should be drawn from these figures. The solid line is the best-fitting model to these data. As in the simulations, the acoustic feature appears sharpened.
Our goal in this paper has been to understand the impact of reconstruction and we have therefore kept our sample as close to previously analysed samples. There has been significant recent work understanding systematics in the source catalogue arising from star–galaxy separation and photometric zero-point errors (Ross et al. 2011, 2012). These principally affect the shape of the correlation function and the above papers explicitly show that the BAO scale is unaffected. We therefore do not address these here, but note that quantifying these will be important as the precision of future surveys improve.

Table 4. The comoving distance to $z = 0.35$, expressed as (i) $\alpha \equiv D_v/r_s/(D_v/r_s)_{\text{fid}}$, (ii) $D_v/r_s$, and (iii) $D_v$ assuming the fiducial value of $r_s$ (and ignoring errors). The first group of numbers compares the unreconstructed and reconstructed cases (assuming our default reconstruction parameters). The distances obtained are consistent, but reconstruction reduces the error by a factor of 1.8, resulting in a distance precise to 1.9 per cent. The second group explores the impact of changing the various reconstruction parameters (as in Tables 2 and 3; both the distance and its error are robust to any of these changes). Finally, we consider changing the fiducial cosmology (keeping the physical densities fixed). In this case, we expect $\alpha$ to change, since our fiducial distance changes, but the derived distances (both absolute and relative to the sound horizon) are unchanged. Note that the errors do change, reflecting changes in the volume relative to the acoustic scale.

| Case       | $\alpha - 1 (\times 100)$ | $D_v/r_s$     | $D_v$ (Gpc)  |
|------------|---------------------------|---------------|--------------|
| Unrecon    | $1.3 \pm 3.5$             | $8.89 \pm 0.31$ | $1.358 \pm 0.047$ |
| Recon      | $1.2 \pm 1.9$             | $8.88 \pm 0.17$ | $1.356 \pm 0.025$ |
| Smoothing, 20 Mpc $h^{-1}$ | $0.9 \pm 2.1$             | $8.85 \pm 0.18$ | $1.352 \pm 0.028$ |
| $b = 1.8$ (–20 per cent) | $1.4 \pm 2.0$             | $8.89 \pm 0.18$ | $1.358 \pm 0.027$ |
| $b = 2.6$ (+20 per cent) | $1.4 \pm 1.9$             | $8.89 \pm 0.16$ | $1.359 \pm 0.025$ |
| $f = 0.5$ (–20 per cent) | $1.1 \pm 1.9$             | $8.87 \pm 0.16$ | $1.355 \pm 0.025$ |
| $f = 0.8$ (+20 per cent) | $1.5 \pm 1.9$             | $8.90 \pm 0.16$ | $1.360 \pm 0.025$ |
| $P(k)$, no wiggle | $1.2 \pm 1.9$             | $8.88 \pm 0.17$ | $1.356 \pm 0.025$ |
| $P(k)$, shot-noise | $1.5 \pm 1.9$             | $8.90 \pm 0.17$ | $1.360 \pm 0.026$ |
| $\Omega_m = 0.20$ | $15.9 \pm 2.4$             | $8.93 \pm 0.18$ | $1.377 \pm 0.028$ |
| $\Omega_m = 0.35$ | $-7.6 \pm 1.8$             | $8.93 \pm 0.17$ | $1.378 \pm 0.026$ |

Figure 14. The $\Delta \chi^2$ surface as a function of $\alpha$ before (left) and after (right) reconstruction. The solid (red) line uses our default template with a BAO feature in the correlation function, while the dashed (blue) line uses the ‘no-wiggle’ form of Eisenstein & Hu (1998). Reconstruction narrows the $\chi^2$ minimum, reflecting the improved distance constraints. The difference between the solid and dashed lines estimates the significance of the detection of the BAO feature; the horizontal dotted lines mark the 1σ to the 5σ significance level. Reconstruction increases the detection significance from 3.3σ to 4.2σ.

Figure 15. The probability of $\alpha$ before (blue dashed line) and after (red solid line) reconstruction. The mean and standard deviations of the two distributions are also listed. The shaded regions show the 1σ (dark shaded) and 2σ (light shaded) regions of the reconstructed probability distribution.

not have $h^{-1}$ factors in them. Of course, the sound horizon is not perfectly measured and its uncertainty must be taken into account when fitting for cosmologies. Paper III discusses the methodologies and results in detail.

A comment is in order on our particular choice of the fiducial redshift of 0.35. While we quote the distance to a single redshift, the analysis assumes a fiducial distance–redshift relation across the redshift slice; we simply choose to quote the result at the median redshift of the sample. A good test of the appropriateness of this choice is the results in Table 4 where we analyse the sample assuming a different cosmology – in both these cases, we recover the same constant ratio $D_v/r_s$ (within our statistical errors). An inappropriate choice here would have introduced an $\Omega_m$ dependence into the results here.

Finally, our treatment of systematics here has focused on the possible systematics introduced by reconstruction. Our goal in this

© 2012 The Authors, MNRAS 427, 2132–2145
redshift-space distortions and test it on the mock catalogues from the LasDamas suite of simulations. These mock catalogues have been designed to match both the SDSS survey geometry and the redshift distribution and clustering properties of the SDSS LRG sample.

(ii) We find that reconstruction both sharpens the BAO feature and restores the isotropy of the correlation function on large scales. The non-linear smoothing of the BAO feature decreases from 8.1 to 4.4 Mpc $h^{-1}$, in line with theoretical estimates (Padmanabhan et al. 2009).

(iii) We find that reconstruction improves the distance estimates, reducing the median error of the LasDamas simulations from 3.3 to 2.3 per cent. Furthermore, reconstruction also significantly reduces the number of outliers as a result of improving the detectability of the BAO feature.

(iv) We calibrate the smoothing scale input to the reconstruction algorithm and find that the optimal scale lies between $\sim15$ and $20\, \text{Mpc} h^{-1}$; we adopt $15\, \text{Mpc} h^{-1}$ as our fiducial value. Choosing too small a smoothing scale results in prominent artefacts in the correlation function, while too large a scale degrades the efficacy of reconstruction.

(v) We demonstrate that the reconstruction is robust to the choices of galaxy bias $b$, the rate of growth of structure $f$ and the fiducial power spectrum used to interpolate missing data and pad the survey edges.

(vi) Applying reconstruction to the SDSS DR7 LRG data, we measure a relative distance to $z = 0.35$ of $D_{v}/r_s = 8.88 \pm 0.17$, compared with $8.89 \pm 0.31$ before reconstruction. The two distances are consistent, but reconstruction reduced the error from 3.5 to 1.9 per cent, a factor of 1.8, equivalent to tripling the survey volume.

(vii) Reconstruction also improves the detectability of a BAO feature (relative to a model with feature erased) from 3.3 to 4.2.$\sigma$.

We can compare our results to the results of Percival et al. (2010) and Kazin et al. (2010), who analysed similar samples with different (albeit related) fitting methodologies. Percival et al. (2010) analyse a combination of the SDSS LRG data and lower redshift data from the SDSS main galaxy sample. Their primary distance constraints are therefore reported at a lower redshift from ours: $D_{v}/r_s(z = 0.275) = 7.19 \pm 0.19$, a 2.7 per cent measurement. Comparing these results with ours requires assuming a model to transform to a higher redshift. This comparison is done in detail in Paper III. We can however scale our results to this redshift (assuming that $\alpha$ does not change significantly with redshift) to obtain $7.15 \pm 0.13$. These results are also consistent with those of Kazin et al. (2010) who obtain $7.17 \pm 0.25$. We note that our error before reconstruction is larger than the Percival et al. (2010) results; this is however expected given the somewhat larger volume of that sample. Percival et al. (2010) also split their sample into two redshift slices with their higher redshift slice corresponding to our measurements. They obtain $D_{v}/r_s(z = 0.35) = 9.11 \pm 0.30$, consistent with our results before reconstruction.

Our results have important implications for current and future surveys. All of these surveys have assumed some level of reconstruction for their projected constraints. This work retires a major risk for these surveys, being the first application to data. Furthermore, the degree of reconstruction assumed (a reduction of the non-linear smoothing scale by 50 per cent) is consistent with what this work has achieved.

This work has also limited itself to the angle-averaged correlation function. One of the promises of the BAO method is the ability to measure both the angular diameter distance and the Hubble constant. These measurements are complicated by the loss of isotropy in the correlation function due to redshift-space distortions. As this paper has shown, reconstruction has the potential to undo the effects of redshift-space distortions and could significantly improve measurements of the anisotropic BAO signal. We will explore this in future work.

This paper has demonstrated that reconstruction is feasible on data and that it can significantly improve the distance constraints. We expect that reconstruction will become a standard method for analysing the BAO signal from large redshift surveys.

ACKNOWLEDGMENTS

Funding for the Sloan Digital Sky Survey (SDSS) and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the US Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society and the Higher Education Funding Council for England. The SDSS website is http://www.sdss.org/.

The SDSS is managed by the Astrophysical Research Consortium (ARC) for the Participating Institutions. The Participating Institutions are the American Museum of Natural History, Astrophysical Institute Potsdam, University of Basel, University of Cambridge, Case Western Reserve University, the University of Chicago, Drexel University, Fermilab, the Institute for Advanced Study, the Japan Participation Group, the Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and Cosmology, the Korean Scientist Group, the Chinese Academy of Sciences (LAMOST), Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, Ohio State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Observatory and the University of Washington.

We thank the LasDamas Collaboration for making their galaxy mock catalogues public. We thank Cameron McBride for assistance in using the LasDamas mocks and for comments on earlier versions of this work. We thank Martin White for useful conversations on reconstruction. NP and AJC are partially supported by NASA grant NNX11AF43G. DJE, XX and KTM were supported by NSF grant AST-0707725 and NASA grant NNX07AH11G. This work was supported in part by the facilities and staff of the Yale University Faculty of Arts and Sciences High Performance Computing Center.

REFERENCES

Abazajian K. N. et al., 2009, ApJS, 182, 543
Angulo R. E., Baugh C. M., Frenk C. S., Lacey C. G., 2008, MNRAS, 383, 755
Balay S., Gropp W. D., McInnes L. C., Smith B. F., 1997, in Arge E., Bruaset A. M., Langtangen H. P., eds, Modern Software Tools in Scientific Computing. Birkhäuser Press, Cambridge, MA, p. 163
Balay S. et al., 2011a, PETSc Users Manual. Tech. Rep. ANL-95/11-Revision 3.2, Argonne National Laboratory
Balay S. et al., 2011b, PETSc, http://www.mcs.anl.gov/petsc
Beutler F. et al., 2011, MNRAS, 416, 3017
Blake C., Glazebrook K., 2003, ApJ, 594, 665
Blake C., Collister A., Bridle S., Lahav O., 2007, MNRAS, 374, 1527
Blake C. et al., 2011a, MNRAS, 415, 2892
Blake C. et al., 2011b, MNRAS, 418, 1707
Bond J. R., Efstathiou G., 1984, ApJ, 285, L45
Bond J. R., Efstathiou G., 1987, MNRAS, 226, 655

© 2012 The Authors, MNRAS 427, 2132–2145
Monthly Notices of the Royal Astronomical Society © 2012 RAS
