HAL open science

Euclid: Cosmology forecasts from the void-galaxy cross-correlation function with reconstruction

S. Radinovic, S. Nadathur, H.-A. Winther, W.J. Percival, A. Woodfinden, E. Massara, E. Paillas, S. Contarini, N. Hamaus, A. Kovacs, et al.

To cite this version:

S. Radinovic, S. Nadathur, H.-A. Winther, W.J. Percival, A. Woodfinden, et al.. Euclid: Cosmology forecasts from the void-galaxy cross-correlation function with reconstruction. Astronomy and Astrophysics - A&A, 2023, 677, pp.A78. 10.1051/0004-6361/202346121 . hal-04006508

HAL Id: hal-04006508

https://hal.science/hal-04006508

Submitted on 9 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

A&A 677, A78 (2023)
https://doi.org/10.1051/0004-6361/202346121
© The Authors 2023

Astronomy & Astrophysics

Euclid: Cosmology forecasts from the void-galaxy cross-correlation function with reconstruction★

S. Radinović[1], S. Nadathur[2], H.-A. Winther[1], W. J. Percival[3,4,5], A. Woodfinden[3,4], E. Massara[3,4], E. Paillas[3,4], S. Contarini[6,7,8], N. Hamaus[9], A. Kovacs[10,11,12], A. Pisani[13,14,15], G. Verza[16,17], M. Aubert[18], A. Amara[2], N. Auricchio[8], M. Baldi[6,8,7], D. Bonino[19], E. Branchini[20,21], M. Brescia[22], S. Camera[23,24,19], V. Capobianco[19], C. Carbone[25], V. F. Cardone[26,27], J. Carretero[28,29], M. Castellano[26], S. Cavuoti[30,31], A. Cimatti[32], R. Cledassou[33,34], G. Congedo[35], L. Conversi[36,37], Y. Copin[18], L. Corcione[19], F. Courbin[38], A. Da Silva[39,40], M. Douspis[41], F. Dubath[42], X. Dupac[36], S. Farrens[43], S. Ferriol[18], P. Fosalba[44,45], M. Frailis[46], E. Franceschi[8], M. Fumana[25], S. Galeotta[46], B. Garilli[25], W. Gillard[47], B. Gillis[35], C. Giocoli[8,7], A. Grazian[48], F. Grupp[49,9], S. V. H. Haugan[1], W. Holmes[50], A. Hornstrup[51,52], K. Jahnke[53], M. Kümmel[9], A. Kiessling[50], M. Kilbinger[43], T. Kitching[54], H. Kurki-Suonio[55,56], S. Ligori[19], P. B. Lilje[1], I. Lloro[57], E. Maiorano[8], O. Mansutti[46], O. Marggraf[58], K. Markovic[50], F. Marulli[6,8,7], R. Massey[59], S. Mei[60], M. Melchior[61], Y. Mellier[62,63,64], M. Meneghetti[8,7], E. Merlin[26], G. Meylan[38], M. Moresco[6,8], L. Moscardini[6,8,7], S.-M. Niemi[65], J. W. Nightingale[59], T. Nutma[66,67], C. Padilla[28], S. Paltani[42], F. Pasian[46], K. Pedersen[68], V. Pettorino[43], S. Pires[69], G. Polenta[70], M. Poncet[33], L. A. Popa[71], L. Pozzetti[8], F. Raison[49], A. Renzi[17,16], J. Rhodes[50], G. Riccio[30], E. Romelli[46], M. Roncarelli[8], C. Rosset[60], R. Saglia[9,49], D. Sapone[72], B. Sartoris[9,46], P. Schneider[58], A. Secroun[47], G. Seidel[53], S. Serrano[44,73], C. Sirignano[17,16], G. Sirri[7], L. Stanco[16], J.-L. Starck[69], C. Surace[74], P. Tallada-Crespí[75,29], I. Tereno[39,76], R. Toledo-Moreo[77], F. Torradeflot[75,29], I. Tutusaus[78], E. A. Valentijn[67], L. Valenziano[8,7], T. Vassallo[46], Y. Wang[79], J. Weller[9,49], G. Zamorani[8], J. Zoubian[47], and V. Scottez[62,80]

(Affiliations can be found after the references)

Received 10 February 2023 / Accepted 17 June 2023

ABSTRACT

We have investigated the cosmological constraints that can be expected from measurement of the cross-correlation of galaxies with cosmic voids identified in the *Euclid* spectroscopic survey, which will include spectroscopic information for tens of millions of galaxies over 15 000 deg^2 of the sky in the redshift range $0.9 \leq z < 1.8$. We have done this using simulated measurements obtained from the Flagship mock catalogue, the official *Euclid* mock that closely matches the expected properties of the spectroscopic dataset. To mitigate anisotropic selection-bias effects, we have used a velocity field reconstruction method to remove large-scale redshift-space distortions from the galaxy field before void-finding. This allowed us to accurately model contributions to the observed anisotropy of the cross-correlation function arising from galaxy velocities around voids as well as from the Alcock–Paczynski effect, and we studied the dependence of constraints on the efficiency of reconstruction. We find that *Euclid* voids will be able to constrain the ratio of the transverse comoving distance $D_{\rm M}$ and Hubble distance $D_{\rm H}$ to a relative precision of about 0.3%, and the growth rate $f\sigma_8$ to a precision of between 5% and 8% in each of the four redshift bins covering the full redshift range. In the standard cosmological model, this translates to a statistical uncertainty $\Delta\Omega_{\rm m} = \pm 0.0028$ on the matter density parameter from voids, which is better than what can be achieved from either *Euclid* galaxy clustering and weak lensing individually. We also find that voids alone can measure the dark energy equation of state to a 6% precision.

Key words. cosmology: observations – cosmological parameters – large-scale structure of Universe – surveys

1. Introduction

Euclid (Laureijs et al. 2011) is an upcoming space mission that aims to explore the nature of dark matter and dark energy through observations of galaxy clustering and weak lensing. Clustering statistics will mainly be extracted from the spectroscopic dataset, which will contain accurate spectroscopic redshifts for tens of millions of galaxies, spanning a wide redshift range ($0.9 \leq z < 1.8$) and covering almost a third of the sky. The primary analyses of such data usually focus on the galaxy auto-correlation statistics, in particular the two-point correlation function (2PCF) and its Fourier space counterpart, the power spectrum. These analyses allow for the measurement of the characteristic baryon acoustic oscillation (BAO) feature (e.g., Alam et al. 2017, 2021), together with the signatures of galaxy velocities which give rise to redshift-space distortions (RSD; Kaiser 1987). However, non-Gaussianity introduced by non-linear gravitational evolution means that two-point statistics do not fully characterise all of the information available from galaxy surveys. In addition, models based on perturbation theory start to break down at small scales (at $r \lesssim 25\,h^{-1}$ Mpc in the 2PCF, or $k \gtrsim 0.15\,h\,{\rm Mpc}^{-1}$ in the power spectrum); therefore, information from these smaller scales is often excluded from analysis to avoid systematic biases in the modelling (Beutler et al. 2017; Satpathy et al. 2017).

★ This paper is published on behalf of the Euclid Consortium.

Open Access article, published by EDP Sciences, under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
This article is published in open access under the Subscribe to Open model. Subscribe to A&A to support open access publication.

Several alternative clustering summary statistics have been used to extract additional information from the galaxy distribution, such as polyspectra (e.g., Verde et al. 2002; Scoccimarro et al. 2001; Gil-Marín et al. 2016; Philcox & Ivanov 2022) and N-point correlation functions (e.g., Wang et al. 2004; Nichol et al. 2006; Marín et al. 2013; Guo et al. 2015; Slepian et al. 2017). Other proposed statistics include marked correlation functions or power spectra (e.g., White 2016; Satpathy et al. 2019; Massara et al. 2023), counts-in-cells statistics (e.g., Yang & Saslaw 2011; Uhlemann et al. 2017), density-dependent clustering (e.g., Tinker 2007; Chiang et al. 2014; Bayer et al. 2021; Bonnaire et al. 2022; Paillas et al. 2023), anisotropic clustering (e.g., Paz et al. 2011; Kazin et al. 2012; Correa et al. 2019), wavelet-based methods (e.g., Valogiannis & Dvorkin 2022a,b), non-linear transformations of the galaxy field (e.g., Neyrinck et al. 2009; Carron & Szapudi 2014; Wolk et al. 2015), and density field reconstruction (e.g., Wang et al. 2022).

With the advent of large galaxy redshift surveys, cosmic voids – large low-density regions in the galaxy distribution – have emerged as interesting probes of cosmology in many contexts, for example using the void size function (e.g., Pisani et al. 2015; Nadathur 2016; Correa et al. 2019; Contarini et al. 2022), gravitational lensing by voids (e.g., Sánchez et al. 2017; Raghunathan et al. 2020; Bonici et al. 2023), secondary cosmic microwave background (CMB) anisotropies (e.g., Granett et al. 2008; Nadathur & Crittenden 2016; Alonso et al. 2018; Kovács et al. 2019), or void clustering (e.g., Chan et al. 2014; Kitaura et al. 2016; Zhao et al. 2022). Readers can refer to Pisani et al. (2019) for a comprehensive list.

The anisotropic distribution of galaxies around voids is a particularly interesting observable (Lavaux & Wandelt 2012) that has been shown to be a valuable source of cosmological information, providing parameter constraints that are highly complementary to those obtained from galaxy clustering (e.g., Nadathur et al. 2019b; Hamaus et al. 2022; Woodfinden et al. 2022). Individual voids have irregular shapes and arbitrary orientations on the sky, but, given a large enough sample of voids, the assumption of statistical isotropy of the Universe (together with some additional assumptions discussed in more detail in Sect. 2) should give rise to a spherically symmetric distribution of galaxies around voids on average (Ryden 1995). Two effects that spoil this symmetry are distortions due to the Alcock–Paczyński (AP) effect (Alcock & Paczynski 1979), arising from the choice of the fiducial cosmological model used to convert observed galaxy redshifts and angles on the sky into distances, and RSD arising from the peculiar velocities of galaxies around these voids.

As voids are underdense environments, the RSD contributions to the anisotropy can be relatively successfully modelled using linear theory (Hamaus et al. 2014; Cai et al. 2016; Nadathur & Percival 2019; Paillas et al. 2021) without needing to exclude small scales. This provides constraints on the growth rate of structure parametrised by $f\sigma_8$ or β (e.g., Paz et al. 2013; Hamaus et al. 2016; Hawken et al. 2017, 2020; Nadathur et al. 2019b; Achitouv 2019; Aubert et al. 2022; Woodfinden et al. 2022). More importantly, Hamaus et al. (2015) and Nadathur et al. (2019b) show that the anisotropies from RSD contributions can be easily distinguished from those arising due to the AP effect. This allows the stacked void profile – equivalent to the void-galaxy cross-correlation function (CCF) – to be used as a 'standard sphere' for the AP test, as originally advocated by Lavaux & Wandelt (2012). This observable can then be used to measure the AP parameter, which is the ratio of the transverse comoving distance $D_{\rm M}(z)$ to the Hubble distance $D_{\rm H}(z)$ as a function of redshift. The constraints obtained on this quantity by applying this method to the BOSS and eBOSS surveys are a factor of 1.7–3.5 more precise than the constraints from galaxy clustering and the BAO in the same data (Hamaus et al. 2016; Nadathur et al. 2019b, 2020b). This precise AP measurement underpins the power of the void-galaxy CCF as a cosmological probe: Nadathur et al. (2020a) showed that the combination of voids and BAO measurements alone provides greater than 10σ evidence for late-time cosmic acceleration, independent of the cosmic microwave background or supernovae.

In this work we aim to forecast the constraints on the AP parameter $D_{\rm M}/D_{\rm H}$ and the growth rate $f\sigma_8$ that can be obtained from measurements of the void-galaxy CCF by *Euclid*. This paper is one of a set of four companion papers published on behalf of the Euclid Consortium, which forecast the expected constraints from cosmic voids using different observables. Contarini et al. (2022) explored the use of the void size function, Bonici et al. (2023) investigated the cross-correlation of voids with weak lensing, while this paper and Hamaus et al. (2022) examine different methods of studying the void-galaxy CCF. The primary difference between our work and that of Hamaus et al. (2022) is that we use a velocity-field reconstruction method to approximately remove large-scale galaxy RSD effects before running the void-finding pipeline (Nadathur et al. 2019a) in order to mitigate against systematic errors due to anisotropic selection effects. On the other hand, instead of applying reconstruction, Hamaus et al. (2022) modify terms in the theoretical model used for the fits, and so measure a very different pattern of anisotropy in the CCF.

The structure of this paper is as follows: in Sect. 2 we provide a theoretical description of the void-galaxy cross-correlation function, as well as the AP test. In Sect. 3 we describe the Flagship galaxy mock and the reconstruction method we apply and which enables us to recover the approximate real-space galaxy positions. We also describe the void finder and detail the properties of the void catalogue. In Sect. 4 we describe our pipeline, obtaining the data vectors and translating this into constraints on parameters of interest. Sect. 5 describes the results we obtain, which are then summarised and put into a larger context in Sect. 6.

2. Theory

2.1. The void-galaxy cross-correlation function in redshift space

The void-galaxy CCF ξ represents the excess probability of finding a galaxy at a given separation from a void centre. We use the notation $\xi^{\rm rr}$ to denote this quantity when the cross-correlation is performed between the real-space positions of galaxies and the positions of voids identified in this real-space galaxy distribution. For convenience, we refer to $\xi^{\rm rr}$ as the real-space void-galaxy CCF. For the same fixed void centres, the cross-correlation with the redshift-space galaxy positions is then denoted $\xi^{\rm rs}$, and will be referred to as the redshift-space CCF. Here r refers to real space, s to redshift space, the first superscript refers to the voids, and the second to the galaxies. We denote the vectors from the void centre to the galaxy in real and redshift space by $\boldsymbol{r}$ and $\boldsymbol{s}$ respectively, and decompose them into their components perpendicular and parallel to the line-of-sight direction, $\boldsymbol{s} = \boldsymbol{s}_\perp + \boldsymbol{s}_\parallel$. Thus $\boldsymbol{s}_\perp = \boldsymbol{r}_\perp$ while

$$\boldsymbol{s}_\parallel = \boldsymbol{r}_\parallel + \frac{\boldsymbol{v}_\parallel}{aH}, \tag{1}$$

where a is the scale factor, H the Hubble rate, $v_\parallel$ is the component of the galaxy peculiar velocity along the line of sight (not the pairwise velocity; see e.g., Massara et al. 2022), and the line of sight is directed through the void centre.

Since, by construction, the number of void-galaxy pairs is conserved under the mapping between real and redshift space, $\xi^{\rm rs}$ and $\xi^{\rm rr}$ are related by a convolution with the (position-dependent) velocity distribution:

$$1 + \xi^{\rm rs}(\boldsymbol{s}) = \int [1 + \xi^{\rm rr}(\boldsymbol{r})]\, P(v_\parallel, \boldsymbol{r})\, {\rm d}v_\parallel. \tag{2}$$

This is recognisable as the streaming model (e.g., Peebles 1979), which was used to describe the void-galaxy CCF by Paz et al. (2013) and also, for example, by Cai et al. (2016) and Paillas et al. (2021). Although – for reasons explained below – we have chosen to hold fixed the void positions identified in the real-space galaxy field, Eq. (2) would be equally valid if both the CCFs were defined with respect to the void positions in the redshift-space galaxy field, that is if $\xi^{\rm rs}$ and $\xi^{\rm rr}$ were replaced by $\xi^{\rm ss}$ and $\xi^{\rm sr}$ respectively. Indeed this is the approach taken in a number of other papers (Paz et al. 2013; Hamaus et al. 2016; Achitouv 2017). However, the number of voids in a size-selected sample is not conserved under the RSD mapping (Chuang et al. 2017; Nadathur et al. 2019a; Correa et al. 2022), so Eq. (2) cannot be used to relate CCFs obtained from running void-finding independently in the two spaces, for example $\xi^{\rm ss}$ and $\xi^{\rm rr}$.

We can view the galaxy peculiar velocities as being composed of a term describing a coherent velocity outflow from the void centre, plus a random velocity component along a random direction. Given the assumption of statistical isotropy in real space, when averaged over sufficiently large numbers of voids, the coherent outflow $\boldsymbol{v}(r) = v_{\rm r}(r)\,\hat{\boldsymbol{r}}$ must be directed radially outwards from the void centre and be spherically symmetric about it. Thus the line-of-sight component of the galaxy velocity can be written as

$$v_\parallel(r, \mu_{\rm r}) = v_{\rm r}(r)\,\mu_{\rm r} + \tilde{v}_\parallel, \tag{3}$$

where $\mu_{\rm r} \equiv \cos(\theta)$ is the cosine of the angle θ between $\boldsymbol{r}$ and the line of sight, and $\tilde{v}_\parallel$ is a zero-mean random variable describing the projection of the stochastic velocity component along the line of sight.

Using Eq. (3) to change variables, Woodfinden et al. (2022) show that Eq. (2) is equivalent to

$$1 + \xi^{\rm rs}(\boldsymbol{s}) = \int [1 + \xi^{\rm rr}(\boldsymbol{r})]\, J_{\boldsymbol{r},\boldsymbol{s}}\, P(\tilde{v}_\parallel, \boldsymbol{r})\, {\rm d}\tilde{v}_\parallel, \tag{4}$$

where

$$J_{\boldsymbol{r},\boldsymbol{s}} \equiv \left|\frac{{\rm d}\boldsymbol{r}}{{\rm d}\boldsymbol{s}}\right| = \left[1 + \frac{v_{\rm r}}{r\,a\,H} + \frac{r\,v_{\rm r}' - v_{\rm r}}{r\,a\,H}\mu_{\rm r}^2\right]^{-1} \tag{5}$$

is the Jacobian of the transformation from real to redshift space, prime denotes derivatives with respect to r, and $P(\tilde{v}_\parallel, \boldsymbol{r})$ now describes the distribution of $\tilde{v}_\parallel$ about zero. Equation (4) is the same model described by Nadathur & Percival (2019). Several other works (e.g., Cai et al. 2016; Hawken et al. 2020; Aubert et al. 2022) use approximations to this expression, first by assuming a Dirac delta function form for the distribution $P(\tilde{v}_\parallel, \boldsymbol{r})$, which reduces the integral equation to

$$1 + \xi^{\rm rs}(\boldsymbol{s}) = [1 + \xi^{\rm rr}(\boldsymbol{r})]\, J_{\boldsymbol{r},\boldsymbol{s}}, \tag{6}$$

and then approximating the square brackets in Eq. (5) by a series expansion, which one can choose to truncate at some desired order. Hamaus et al. (2020) introduce additional nuisance parameters to Eq. (6), and Hamaus et al. (2022) further modify some terms of the Jacobian expansion.

A key assumption for all models based on Eq. (4) is that the coherent velocity term $v_{\rm r}(r)$ is spherically symmetric and radially directed, which depends crucially on the assumption that in real space there is no preferred orientation direction for the large number of individually asymmetric voids. This follows from statistical isotropy only provided there is no orientation-dependent selection effect introduced in creating the void sample. Nadathur et al. (2019a) argued that exactly such an effect is introduced for voids identified in redshift space, since regions with velocity outflows oriented along the line of sight direction appear to have lower densities in redshift space. This means that elongated voids aligned along the line-of-sight direction are more likely to be selected as void candidates if void-finding is performed on the redshift-space galaxy positions.

This orientation-dependent bias has a large effect on the measured CCFs (Nadathur et al. 2019a; Correa et al. 2022; Paillas et al. 2023). The CCF with real-space galaxy positions, $\xi^{\rm sr}$, becomes strongly anisotropic, the coherent velocity outflow $v_{\rm r}$ no longer has spherical symmetry, and the multipoles of the CCF with redshift-space galaxy positions, $\xi^{\rm ss}$, have a very different shape. The loss of spherical symmetry in particular introduces significant challenges for the modelling approach described above. To avoid this problem, in this paper we do not consider the case of voids identified in the redshift-space galaxy field. As the real-space galaxy positions are in general not known, we used the reconstruction technique described in Sect. 3.2 below to approximately recover them before performing void-finding, and used these void positions to approximately measure $\xi^{\rm rs}$ and $\xi^{\rm rr}$.

2.2. Modelling the velocity distribution

Given the stated assumptions, the convolution described by Eq. (4) – or, equivalently, Eq. (2) – is completely general and exact. In order to make specific predictions, however, we need to specify a model for the velocity field around voids to obtain the radial term $v_{\rm r}(r)$ and the distribution $P(\tilde{v}_\parallel, \boldsymbol{r})$.

The radial outflow velocity $v_{\rm r}$ itself is commonly modelled (e.g., Cai et al. 2016; Nadathur et al. 2019b; Woodfinden et al. 2022) using a a linearised form of the continuity equation (Peebles 1980):

$$v_{\rm r}(r) = -\frac{1}{3} f\,a\,H\,r\,\Delta, \tag{7}$$

where

$$\Delta(r) = \frac{3}{r^3}\int_0^r \delta(y)\,y^2\,{\rm d}y \tag{8}$$

is the fully non-linear enclosed density profile of dark matter, $\delta(r) = \rho/\bar{\rho} - 1$ is the matter overdensity, and f is the linear growth rate. Nadathur et al. (2019b) found that in simulated galaxy catalogues the dark matter profile of voids approximately scales as $\Delta \propto \sigma_8$, where σ_8 describes the amplitude of the matter density perturbations on $8\,h^{-1}$ Mpc scales[1]. When combined with Eq. (7), this implies a scaling $v_{\rm r} \propto f\sigma_8$.

[1] This scaling was demonstrated under conditions where the number density of the galaxies used to trace voids was held constant, and their bias was adjusted to keep $b\sigma_8$ – and thus the overall galaxy-clustering amplitude – fixed.

Equation (7) has been found to give a reasonably good description of the matter velocity field around voids by some studies (Hamaus et al. 2014; Nadathur & Percival 2019). However, other works (Achitouv 2017; Paillas et al. 2021) have found that it can be necessary to add correction terms. Moreover, it has been shown by Massara et al. (2022) that for some void finders, when void-finding is performed on sparse tracers of the matter density field such as galaxies, an effective velocity bias can arise between the matter velocity which satisfies Eq. (7), and the galaxy velocity appearing in Eq. (5). This can mean that even if Eq. (7) accurately describes the velocity field for matter, galaxy velocities can differ from this in the inner regions of voids. These considerations will need to be taken into account when improving the velocity modelling beyond that in Eq. (7). It is not our aim to attempt this task in this paper, but only to forecast the constraints from the void-galaxy cross-correlation function that could be obtained from such a model in the future. We therefore applied a template-based method that assumes only that the scaling inferred from Eq. (7) holds. That is, we measured the mean velocity profile $v_{\rm r}^{\rm sim}(r)$ from the Flagship simulation as described in Sect. 4.2 below, and assume that the velocity in other cosmological models is given by

$$\frac{v_{\rm r}(r)}{aH} = \frac{f\sigma_8}{f^{\rm sim}\sigma_8^{\rm sim}}\left(\frac{v_{\rm r}(r)}{aH}\right)^{\rm sim}. \tag{9}$$

For the velocity distribution $P(\tilde{v}_\parallel, \boldsymbol{r})$, we assume a Gaussian form,

$$P(\tilde{v}_\parallel, \boldsymbol{r}) = \frac{1}{\sqrt{2\pi\sigma^2_{v_\parallel}(r)}}\exp\left(-\frac{\tilde{v}_\parallel^2}{2\sigma^2_{v_\parallel}(r)}\right), \tag{10}$$

which has been shown to be a good approximation by Nadathur & Percival (2019) and Paillas et al. (2021). As with $v_{\rm r}(r)$, we assume the width of the dispersion $\sigma_{v_\parallel}(r)$ is spherically symmetric about the void centre and take a template-based approach to its modelling. We measured a template dispersion profile $\sigma_{v_\parallel}^{\rm sim}(r)$ from the Flagship simulation (Sect. 4.2). At large distances from the void centre, this profile asymptotes to a constant, $\sigma_{\rm v}^{\rm sim} \equiv \sigma_{v_\parallel}^{\rm sim}(r \to \infty)$ (see also Hamaus et al. 2015). To account for changes in other cosmologies, we renormalised this function by an amplitude factor $\sigma_{\rm v}$, taken to be a free parameter to be marginalised over:

$$\sigma_{v_\parallel}(r) = \frac{\sigma_{\rm v}}{\sigma_{\rm v}^{\rm sim}}\sigma_{v_\parallel}^{\rm sim}(r). \tag{11}$$

In summary, the template-based approach to modelling the velocity field around voids requires the specification of two template functions, $v_{\rm r}^{\rm sim}(r)$ and $\sigma_{v_\parallel}^{\rm sim}(r)$, and depends on two free parameters, $f\sigma_8$ and $\sigma_{\rm v}$.

2.3. The real-space correlation function

Although the velocity model has been specified as in Sect. 2.2, the RSD mapping of the void-galaxy correlation described by Eq. (4) only relates $\xi^{\rm rs}$ to its counterpart in real space, $\xi^{\rm rr}$. Though some empirical fitting formulae have been used (Paz et al. 2013; Hamaus et al. 2014; Hawken et al. 2017), models which attempt to predict $\xi^{\rm rr}$ directly from the cosmological parameters do not exist at present[2]. Previous studies have therefore taken different approaches to obtain the real-space CCF. Hamaus et al. (2020, 2022) apply an inversion algorithm (Pisani et al. 2014) to recover it from the redshift-space CCF projected along the line of sight (implictly assuming that the real-space correlation function is spherically symmetric about the void centre). Nadathur et al. (2019b, 2020b) and Woodfinden et al. (2022) instead measure the average $\xi^{\rm rr}$ from a large number of mocks that match the galaxy clustering seen in the data sample: this is effectively a template-based method similar to that described for $v_{\rm r}(r)$ and $\sigma_{v_\parallel}(r)$ above.

In this paper, we use the reconstruction-based estimators described in Sect. 3.2 to estimate the real-space CCF $\xi^{\rm rr}(\boldsymbol{r})$ from the simulation data. While we have argued that in the absence of orientation-dependent selection effects, statistical isotropy should imply spherical symmetry, $\xi^{\rm rr}(\boldsymbol{r}) = \xi^{\rm rr}(r)$, when measured in the true background cosmology, this will in general not be true in the presence of the AP distortions discussed in Sect. 2.4, and could also be spoiled by imperfections in the reconstruction-based RSD removal step. Our implementation of Eq. (4) therefore uses the full $\xi^{\rm rr}(\boldsymbol{r})$ without imposing an assumption of isotropy.

Our treatment of $\xi^{\rm rr}$ here differs from that in Nadathur et al. (2019b) and Woodfinden et al. (2022) because, unlike in those works, we currently only have access to the single representative Flagship simulation here, so we cannot construct a template based on the mean of thousands of mocks (this situation will change by the time of the final *Euclid* data analyses). This means that our estimates of $\xi^{\rm rr}$ are necessarily much noisier. As it is measured from the same data as the redshift-space correlation function $\xi^{\rm rs}$ we wish to fit, correlations between the theory model and data vector may also be introduced, and we consider the implications of this in Appendix A.

2.4. Alcock–Paczynski effect

Besides RSD, another source of anisotropies in the observed void-galaxy cross-correlation function is the AP effect. This arises from the conversion of observed galaxy angular positions and redshifts to distances in order to apply the modelling developed above. This conversion requires the assumption of a fiducial cosmological model which may differ from the true background and thus lead to incorrectly inferred distances, introducing anisotropies.

To parameterise this effect, we introduce two AP scaling parameters describing the ratio of distances parallel and perpendicular to the line of sight to those in the fiducial model,

$$\alpha_\parallel = \frac{D_{\rm H}(z)}{D_{\rm H}^{\rm fid}(z)}, \quad \alpha_\perp = \frac{D_{\rm M}(z)}{D_{\rm M}^{\rm fid}(z)}, \tag{12}$$

where $D_{\rm M}(z)$ is the comoving angular diameter distance and $D_{\rm H}(z) = c/H(z)$ is the Hubble distance at redshift z. In this model the correlation function scales as

$$\xi^{\rm rs}(s_\perp, s_\parallel) = \xi^{\rm rs,fid}\left(\alpha_\perp s_\perp^{\rm fid}, \alpha_\parallel s_\parallel^{\rm fid}\right), \tag{13}$$

where the superscript $^{\rm fid}$ denotes quantities calculated in the fiducial cosmology.

The parameters $\alpha_\parallel$ and $\alpha_\perp$ can also be recast into the following combinations:

$$\alpha = \alpha_\perp^{2/3}\alpha_\parallel^{1/3}, \tag{14}$$

which describes an isotropic volume dilation, and

$$\epsilon = \frac{\alpha_\perp}{\alpha_\parallel}, \tag{15}$$

[2] Any such model, if developed, would have to account for the selection criteria by which the void sample was chosen, and would thus be specific to a particular void finder, whereas Eq. (4) is completely general.

describing anisotropic distortions. The dilation parameter α can be measured using a standard ruler, such as the BAO scale (e.g., Blake & Glazebrook 2003; Seo & Eisenstein 2003), by comparing a measured distance scale to one predicted by fundamental physics models. In our analysis, however, the calibrated template functions used depend on the void size cut applied to the sample. This cut is numerically identical to the one applied to the data. However, in general if the template cosmology differs from the truth, due to AP dilation the same numerical cut might not select the same population of voids, leading to a shift in the mean void size. Therefore, we conservatively choose to avoid using the observed void size as an absolute ruler by marginalising over values of α in order to reduce the possible effects of the template cosmology. To achieve this we isotropically rescaled all of the template profiles $\xi^{\rm rr}(\boldsymbol{r})$, $v_{\rm r}(r)$ and $\sigma_{v_\parallel}(r)$ with α, as described by Nadathur et al. (2019b). This rescaling removes all sensitivity of the model to α. The model then depends only on the AP distortion parameter ϵ, via its effect on the quadrupole and (to a lesser extent) hexadecapole moments.

2.5. Multipole decomposition

The model described above applies generally to the full three-dimensional correlation function, which may be written as $\xi^{\rm rs}(s,\mu_{\rm s})$, where $s = |\boldsymbol{s}|$ and $\mu_{\rm s} \equiv s_\parallel/s$. In practice it is preferable to compress the information into a small number of Legendre multipoles, as

$$\xi^{\rm rs}_\ell(s) = \frac{2\ell+1}{2}\int_{-1}^{1} P_\ell(\mu_{\rm s})\,\xi^{\rm rs}(s,\mu_{\rm s})\,{\rm d}\mu_{\rm s}, \tag{16}$$

where $P_\ell(\mu)$ are the Legendre polynomials of order ℓ. Symmetry dictates that only the even multipoles are non-zero, and we restrict our attention to the monopole, quadrupole and hexadecapole moments ($\ell = 0, 2, 4$) only, as these contain almost all of the measurable information in the correlation.

The real-space cross-correlation function $\xi^{\rm rr}$ can also be decomposed into multipoles analogously to Eq. (16). As mentioned in Sect. 2.3, a commonly used approximation, justified by the assumption of spherical symmetry in real space, is that only the monopole term $\xi^{\rm rr}_0$ is non-zero. However, this assumption can be spoiled by the AP effect or imperfect reconstruction of real-space galaxy positions, so for generality we keep track of the high-order multipoles of $\xi^{\rm rr}$ as well.

3. Data

3.1. Galaxy catalogue

We used the *Euclid* Flagship galaxy mock (Castander et al., in prep.), which is constructed from an N-body simulation with 2 trillion particles of mass $m_{\rm p} \approx 2\times10^9\,h^{-1}M_\odot$. The simulation was run using `PKDGRAV3` (Potter et al. 2017) in a box of side $L = 3780\,h^{-1}$ Mpc, with a flat Lambda Cold Dark Matter (ΛCDM) cosmology and matter density $\Omega_{\rm m} = 0.319$, baryon density $\Omega_{\rm b} = 0.049$, dark energy density $\Omega_\Lambda = 0.681$, scalar spectral index $n_{\rm s} = 0.96$, Hubble parameter $h = H_0/(100\ {\rm km\,s^{-1}\,Mpc^{-1}}) = 0.67$, and the RMS value of density fluctuations on $8\,h^{-1}$ Mpc scales $\sigma_8 = 0.83$. These values were chosen to be very close to those obtained by Planck Collaboration VI (2020). The halo catalogue was constructed on the fly, with halos identified using `Rockstar` (Behroozi et al. 2013), and then populated with a halo occupation distribution (HOD) model which had been calibrated to reproduce observables such as galaxy luminosity and clustering statistics as a function of galaxy luminosity and colour. This was used to create a lightcone mock covering an octant (5157 deg^2) of the sky, which was then cut to match the expected observed Hα flux $f_{\rm H\alpha} > 2\times10^{-16}\,{\rm erg\,s^{-1}\,cm^{-2}}$, and expected redshift range $0.9 \le z < 1.8$ for the *Euclid* spectroscopic galaxy sample (Laureijs et al. 2011; Costille et al. 2018). To simulate the expected sample completeness of *Euclid*, we randomly downsampled the final catalogue, retaining 60% of all objects.

For each mock galaxy in this catalogue we record both the true redshift (corresponding to the background Hubble expansion) and the observed redshift, which includes the Doppler effect due to the galaxy peculiar velocities. We also consider the effects of redshift errors by adding an additional random component drawn from a Gaussian with RMS $\sigma_z = 0.001$[3]. The final resulting galaxy catalogue contains about 6.5×10^6 mostly central galaxies. For the analysis described below we divided this sample into four non-overlapping redshift bins as illustrated in Fig. 1, chosen to match those used in *Euclid* forecasts by Euclid Collaboration (2020). The effective redshift of each bin is generically defined as the weighted average redshifts of void-galaxy pairs in the bin,

$$z_{\rm eff} = \frac{\sum_{ij}\left(\frac{Z_i+z_j}{2}\right)W_i\,w_j}{\sum_{ij}W_i\,w_j}, \tag{17}$$

where Z_i is the redshift of the ith void centre, z_j is redshift of the jth galaxy, and W_i and w_j are optional weights associated with individual voids and galaxies when computing the CCF (as the Flagship mock does not include observational systematics, we do not use weights, that is they are all set to unity).The effective redshifts of the different bins are $z_{\rm eff} = 1.0$, 1.2, 1.4 and 1.64.

3.2. Reconstruction

As discussed in Sect. 2, in order to avoid an orientation-dependent selection bias in the void sample that would invalidate the theory model used, void-finding must be performed on the real-space galaxy distribution without large-scale RSD effects. To achieve this, we used a velocity-field reconstruction method, related to the reconstruction commonly used in BAO analysis (Eisenstein et al. 2007; Padmanabhan et al. 2012), to approximately remove galaxy RSD before the void-finding step (Nadathur et al. 2019a).

Reconstruction was performed using an iterative fast Fourier transform (FFT) algorithm (Burden et al. 2015) which solves the Zeldovich equation in redshift space

$$\nabla\cdot\boldsymbol{\Psi} + \frac{f}{b}\nabla\cdot[(\boldsymbol{\Psi}\cdot\hat{\boldsymbol{r}})\,\hat{\boldsymbol{r}}] = -\frac{\delta_{\rm g}}{b}, \tag{18}$$

for a Lagrangian displacement field $\boldsymbol{\Psi}$, where f is the growth rate, b is the linear galaxy bias, and $\delta_{\rm g}$ is the galaxy overdensity in redshift space. This step is implemented in the `pyrecon` package[4]. It is performed on a regular grid, and densities estimated on the grid are first smoothed with a Gaussian kernel of width $R_{\rm s}$ before solving for the displacement. Given the solution

[3] *Euclid* requirements conservatively specify that redshift error should be less than $\sigma_z = 0.001(1+z)$, but as the Near-Infrared Spectrometer and Photometer (NISP) instrument resolution will be determined by the point spread function and pixel size, the wavelength error and thus redshift error are expected to be constant with redshift. A fuller exploration of the effects of redshift errors, with less restrictive assumptions and including line misidentification, will be presented in future work.

[4] https://github.com/cosmodesi/pyrecon

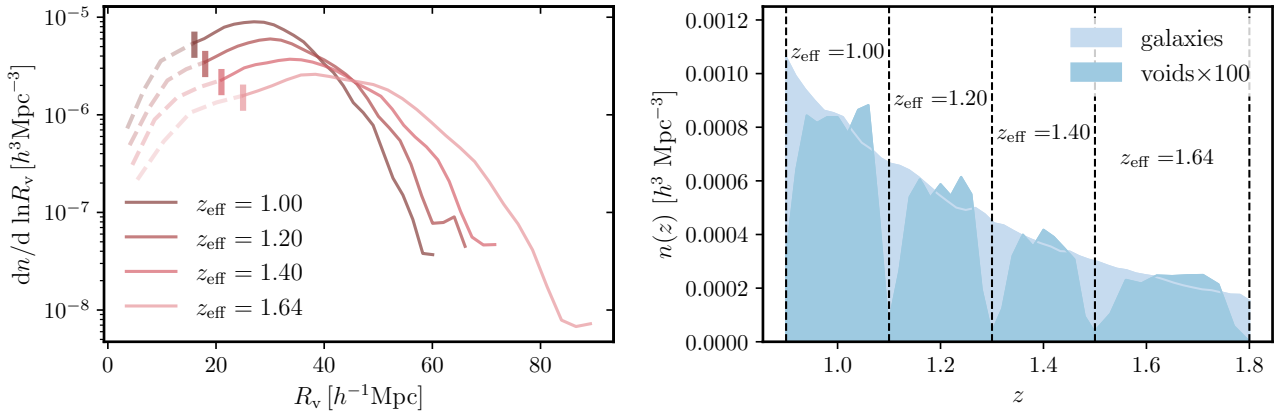

Fig. 1. Left: distribution of void sizes in each of the four redshift bins of the Flagship mock catalogue. The vertical bars indicate the void size cuts applied to each catalogue, and dashed lines represent the voids which are discarded. Right: comoving number density $n(z)$ of galaxies and voids as function of redshift in Flagship. For voids, $n(z)$ is multiplied by 100 for visibility. The edges of the redshift bins are shown by dashed vertical lines. The decrease in the number of voids near the edges of each bin arises because void-finding is performed on each bin individually as discussed in the text; some void candidates near the redshift edges of the bin are then missed.

to Eq. (18), individual galaxy positions are shifted by $-\boldsymbol{\Psi}_\mathrm{RSD}$ evaluated at their locations, where

$$\boldsymbol{\Psi}_\mathrm{RSD} = -f\,(\boldsymbol{\Psi}\cdot\hat{\boldsymbol{r}})\,\hat{\boldsymbol{r}}. \tag{19}$$

This shift subtracts the estimated RSD contribution to the observed galaxy redshifts, in order to approximately recover the real-space galaxy field. The algorithm for solving Eq. (18) is the same as that used for BAO reconstruction (e.g., in the eBOSS analyses Gil-Marín et al. 2020; Bautista et al. 2021) with the difference that the shifts applied to the galaxy positions only attempt to correct for large-scale RSD and not for non-linear evolution.

If we define a new variable $\boldsymbol{\psi} \equiv b\boldsymbol{\Psi}$, Eq. (18) can be rewritten as $\nabla\cdot\boldsymbol{\psi} + \beta\nabla\cdot[(\boldsymbol{\psi}\cdot\hat{\boldsymbol{r}})\,\hat{\boldsymbol{r}}] = -\delta_\mathrm{g}$, where $\beta = f/b$ is the RSD parameter. The shifts applied to the galaxy positions in order to remove RSD are then $\boldsymbol{\Psi}_\mathrm{RSD} = -\beta\,(\boldsymbol{\psi}\cdot\hat{\boldsymbol{r}})\,\hat{\boldsymbol{r}}$, and it is clear that the output of our reconstruction procedure depends only on the parameter β and not on f and b individually[5]. The fiducial growth rate for the Flagship cosmology can be calculated as $f(z) = \Omega_\mathrm{m}^\gamma(z)$ (Linder 2005), where Ω_m is the matter density parameter, and $\gamma = 6/11 \approx 0.55$. To determine the fiducial linear bias b, we measured the galaxy power spectrum in real-space $P_\mathrm{g}(k)$, calculated the theoretical matter power spectrum $P_\mathrm{m}(k)$ for Flagship using `CAMB` (Lewis et al. 2000), and fit for $b^2 = P_\mathrm{g}(k)/P_\mathrm{m}(k)$, using only large-scale modes at $k < 0.07\,h\,\mathrm{Mpc}^{-1}$.

The fiducial values b^fid and β^fid obtained in this way are reported in Table 1 for each redshift bin. However, the growth rate f is a parameter in any fit we perform, and so we do not want to fix β to this fiducial value, but instead to marginalise over it. To achieve this in a computationally feasible way, we precomputed the reconstruction on a grid of β values in the range $\left[\beta^\mathrm{fid} - 0.15, \beta^\mathrm{fid} + 0.15\right]$ around the fiducial value. The data vectors obtained from each of these approximations to the real-space galaxy field consist of the monopole, quadrupole and hexadecapole (all ranging from 0 to $120\,h^{-1}$ Mpc), and are interpolated over when varying β in the analysis.

As the Zeldovich approximation breaks down on small scales, reconstruction also depends on the width of the Gaussian smoothing kernel, R_s, which dictates the smallest scales used. Reconstruction with a smoothing radius that is too small will over-correct the RSD, due to contributions from non-linear small-scale fluctuations, while a smoothing radius that is too big will fail to remove all of them. To determine the optimal smoothing radius to use, we computed the quadrupole of the post-reconstruction galaxy power spectrum, $P_2(k)$, to use as a proxy. In real space this quantity is consistent with zero, while RSD introduce significant anisotropies in redshift space (Fig. 2). In each redshift bin we tested reconstruction with multiple smoothing radii at the fiducial value β^fid, and pick the one that results in a quadrupole which is roughly consistent with zero. This gives $R_\mathrm{s} = 8\,h^{-1}$ Mpc for the first two redshift bins, and $R_\mathrm{s} = 9\,h^{-1}$ Mpc for the last two.

The reconstruction method outlined here cannot perfectly remove all redshift-space effects in the galaxy distribution. We are interested in quantifying the impact of these inaccuracies, as well as those caused by redshift errors, on the void-finding and data analysis pipeline below. We therefore repeated our analysis pipeline in three different scenarios, labelled as: perfect reconstruction, using the true real-space galaxy positions from the simulation; realistic reconstruction, applying reconstruction to the observed redshifts to obtain approximate real-space galaxy positions; and realistic reconstruction with σ_z, applying reconstruction to observed redshifts including redshift errors.

3.3. Void finding

Voids were found in the real space or post-reconstruction galaxy distribution using the `voxel` algorithm, implemented in `REVOLVER`[6] (Nadathur et al. 2019c). `voxel` is a watershed-based void-finding algorithm that operates on a particle-mesh interpolation of the galaxy density field. In the particle-mesh assignment step, tracers (in our case galaxies) are first placed onto a grid to estimate the local density field, a step which is much faster than commonly used tessellation algorithms, and scales linearly with the number of tracers, $O(N)$ (as opposed to $O(N^{3/2})$ for tesselation). The mesh size is determined by requiring each cubic grid cell, or voxel, to have a side length $a_\mathrm{vox} = 0.5\left(\frac{4\pi\bar{n}}{3}\right)^{-1/3}$, where $\bar{n}$ is the approximate value for the

[5] We note that this conclusion applies to our procedure that aims to remove only the RSD effects by shifting by $\boldsymbol{\Psi}_\mathrm{RSD}$, and not when galaxy positions are shifted by the full displacement field $\boldsymbol{\Psi}$ as is done for BAO analyses.

[6] https://github.com/seshnadathur/Revolver

Table 1. Summary properties of the galaxy and void catalogues used.

$z_{\rm eff}$	$N_{\rm g}$	$N_{\rm v}$	$R_{\rm cut}[h^{-1}\,{\rm Mpc}]$	$f_{\rm fid}\sigma_{8,\rm fid}$	$b_{\rm fid}$	$\beta_{\rm fid}$	$\sigma_{\rm v}[{\rm km\,s^{-1}}]$
1.00	2.26×10^6	1.9×10^4	16	0.442	1.598	0.55	321
1.20	1.75×10^6	1.4×10^4	18	0.418	1.907	0.47	318
1.40	1.27×10^6	1.0×10^4	21	0.394	2.106	0.44	310
1.64	1.26×10^6	1.0×10^4	25	0.367	2.382	0.39	295

Notes. For each redshift bin we indicate the approximate total number of galaxies $N_{\rm g}$, the number of voids $N_{\rm v}$ after the void size cut $R_{\rm v} > R_{\rm cut}$ was applied, and the fiducial values for the parameters $f\sigma_8$, linear galaxy bias b, $\beta = f/b$ and velocity dispersion amplitude $\sigma_{\rm v}$. Void numbers refer to those in the perfect reconstruction case, and vary by a few percent in the other reconstruction cases.

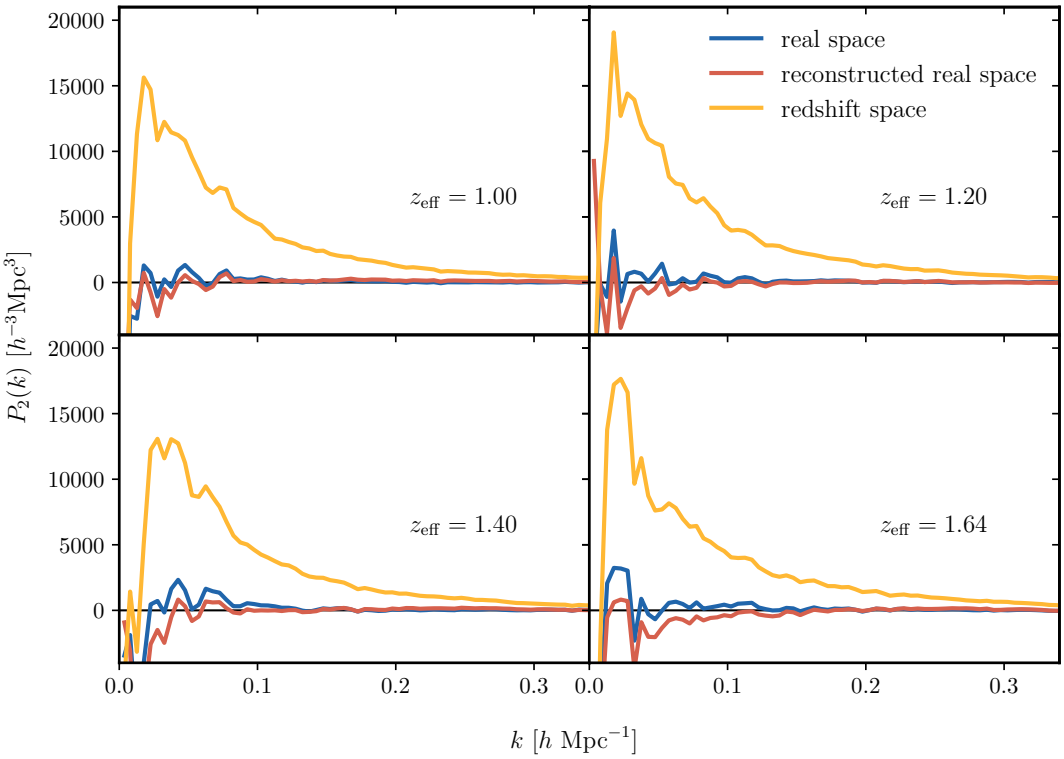

Fig. 2. Quadrupole moments of the galaxy-galaxy power spectrum in each redshift bin of the Flagship mock, measured in redshift space (yellow), real space (blue) and after applying reconstruction to remove RSD from the galaxy positions (red). Reconstruction was performed using the fiducial values of $\beta_{\rm fid}$ listed in Table 1 and using smoothing scales $R_{\rm s}$ of 8 and 9 h^{-1}Mpc for the first two and last two bins, respectively. The similarity of the red and blue lines and their proximity to zero is because reconstruction is successfully removing the large-scale RSD-induced anisotropy in the galaxy positions.

mean number density of galaxies. As $\bar{n}$ varies by a factor of about 5 over the Flagship redshift range, in order to maintain an appropriate resolution in the highest density bin while maintaining a reasonable number of voxels for computational efficiency, we decided to perform the void-finding separately in each of the four redshift bins, using a different $a_{\rm vox}$ for each. The `voxel` algorithm (like all void-finder algorithms) unavoidably misses some voids near the redshift edges of the galaxy sample, so this analysis choice leads to the drops in void numbers around redshifts $z = 1.1$, 1.3 and 1.5 seen in Fig. 1. An alternative approach could involve the use of overlapping redshift bins, which would require a more delicate treatment of correlation between redshifts. We have found that this does not improve our final constraints enough to justify the complexity of the analysis.

In order to correct for the survey window and redshift-dependent or angular selection effects, this density estimate is normalised by the density of a random unclustered catalogue of points which has the same window and selection effects imposed, projected on the same grid. We see from this that, in addition to being significantly faster, this method of estimating the galaxy density field can then more easily account for complex survey masks and systematic effects than popular tessellation-based density estimation methods used in other void-finding codes, such as those based on `ZOBOV` (Neyrinck 2008). For our purposes with the Flagship catalogue, we created a random catalogue with 50 times as many points as the galaxy catalogue, covering the Flagship octant and following the same redshift distribution $n(z)$.

The estimated galaxy density field is then smoothed with a Gaussian filter of smoothing length $r_{\rm s} = \bar{n}^{-1/3}$, in order to reduce shot-noise fluctuations on smaller scales. Voids are then identified as the sites of local minima in the smoothed density field, and their extents are determined using a watershed method, similar to that of `ZOBOV`, to create a final catalogue of non-overlapping voids. Each void is composed of a contiguous set of voxels, which define its volume. Individual voids have irregular

shapes and orientations, but we define an effective void radius as the radius of a sphere with the same volume,

$$R_{\rm v} = \left(\frac{3}{4\pi} N_{\rm vox} V_{\rm vox} \right)^{\frac{1}{3}}, \tag{20}$$

where $N_{\rm vox}$ is the number of voxels in the void and $V_{\rm vox} = a_{\rm vox}^3$ is the voxel volume. The centre of each void is identified as the location of the lowest-density voxel within it, being the position of the density minimum. This results in voids with very few galaxies in the centre.

For each reconstruction scenario considered, the full void catalogues obtained in this way contain about 87 000 voids (with small variations in number) across all redshift bins. However, we expect that smaller voids will not be adequately described by our approach to modelling the coherent velocity outflow $v_{\rm r}(r)$. This is due to a combination of factors. The dynamics of small voids are more likely to be dominated by the presence of neighbouring large density fluctuations so that the isolated void model is inappropriate (see also Schuster et al. 2023); they are more likely to be segments of larger voids that appear artificially small because they extend outside of the survey boundaries; and they are also more likely to be entirely spurious detections arising from shot-noise fluctuations (Cousinou et al. 2019). Therefore, we applied a void size cut to exclude them from the catalogue used for clustering measurements[7]:

$$R_{\rm v} > R_{\rm cut} = 1.5\, \bar{n}^{-\frac{1}{3}}. \tag{21}$$

Keeping only voids that survive this cut reduces the total numbers by about 40%. The radius cuts and the final sizes of the void catalogues are presented in Table 1.

4. Method

4.1. Measuring the void-galaxy cross-correlation function

We measured correlation functions using the Python wrapper `pycorr`[8] for `Corrfunc` (Sinha & Garrison 2020; Sinha et al. 2019). `Corrfunc` is a pair-counting engine that counts the number of pairs of voids and galaxies within a given distance and angular separation bin. This was then compared to the equivalent pair counts computed using random unclustered catalogues to obtain an estimate of the CCF using the Landy-Szalay estimator (Landy & Szalay 1993):

$$\xi(r,\mu) = \frac{{\rm DD}_{12}(r,\mu) - {\rm DR}_{12}(r,\mu) - {\rm DR}_{21}(r,\mu) + {\rm RR}_{12}(r,\mu)}{{\rm RR}_{12}(r,\mu)}. \tag{22}$$

Here D and R refer to the data and unclustered random catalogues respectively, subscripts 1 and 2 are used to differentiate between voids and galaxies, and each term XY refers to the number of pairs of species X and Y in the (r,μ) bin, normalised relative to the total possible number of such pairs. Thus ${\rm DD}_{12}$ refers to normalised number of void-galaxy pairs, ${\rm DR}_{12}$ to the normalised number of pairs of voids and galaxy randoms, and so on. We created the unclustered random catalogues as described in Sect. 3.3, such that they have 50 times as many points as the respective data samples, are uniformly sampled from the Flagship octant footprint and follow the redshift distributions of voids and galaxies respectively (Fig. 1). We measured pair counts and correlation functions 30 radial bins over $0 < r < 120\,h^{-1}$ Mpc and in 200 angular bins over $-1 < \mu < 1$, keeping distances in units of h^{-1} Mpc rather than rescaling them in units of the void size, as is done in some void publications.

It is important to note that shot-noise fluctuations in the galaxy random catalogue used in the mesh density estimation step during void finding (Sect. 3.3) will be slightly correlated with the positions of identified voids. We therefore used an independent realisation of the galaxy random catalogue (generated using the same method but with a different random seed) for measuring correlations via Eq. (22). If instead the same random catalogue were used for both, small artefacts could be propagated to the correlation ξ in the regions very close to the void centres.

We used the estimator in Eq. (22) with the pair separations determined by the positions of voids and galaxies in different combinations of real and redshift space to measure the different correlation functions relevant to our analysis. For instance, in the idealised case of perfect reconstruction, voids are identified in the true real-space galaxy field, and their cross-correlation with the real and redshift-space galaxy positions give $\xi^{\rm rr}$ and $\xi^{\rm rs}$, respectively. On the other hand, when we use realistic reconstruction we use the superscript p to distinguish positions in this post-reconstruction approximation of the real-space galaxy field from the true real-space positions. Voids are found in the post-reconstruction field and, as before, their positions are held fixed to allow measurements of $\xi^{\rm ps}$ and $\xi^{\rm pp}$ with the galaxies in redshift space and approximate real space respectively.

Figure 3 shows the recovered monopole, quadrupole and hexadecapole moments for the true real-space void-galaxy CCF $\xi^{\rm rr}$ in the first Flagship redshift bin, and its comparison with the equivalent multipoles of $\xi^{\rm pp}$ in the realistic reconstruction scenarios with and without including redshift errors. Reconstruction recovers the real-space CCF multipoles well, and anisotropies in all cases are negligible, as evidenced by the quadrupole and hexadecapole moments being close to zero. This demonstrates that our void sample selection successfully avoids the problem of orientation-dependent selection bias discussed in Sect. 2.

4.2. Constructing template functions

As discussed in Sect. 2, we adopt a template-fitting approach to the analysis, and so we require templates for the three functions $v_{\rm r}(r)$, $\sigma_{v_\parallel}(r)$ and $\xi^{\rm rr}(\boldsymbol{r})$. These were measured from the Flagship simulation data as follows.

To measure the radial velocity profile $v_{\rm r}(r)$, we measured the mean radial component of galaxy peculiar velocities (given in the simulation box rest frame) in spherical shells around the void centre, where the average was taken over all void-galaxy pairs in that separation bin. We used 30 linearly spaced radial bins, up to $r = 120\,h^{-1}$ Mpc. The mean radial velocity profiles at different redshifts are shown in the top panel of Fig. 4. The velocity is positive over almost the entire radial range, corresponding to an outflow from the void centre as expected. The profile shows a peak in the vicinity of the edge of the void and at larger radii it goes to zero, showing that there is no net bulk motion far outside of the void, as expected.

We also measured the width of the line-of-sight velocity PDF, $\sigma_{v_\parallel}(r)$, from the distribution of the galaxy peculiar velocities in the same spherical shells. The full vector form of Eq. (3) can be written as

$$\boldsymbol{v}(\boldsymbol{r}) = v_{\rm r}(r)\,\hat{\boldsymbol{r}} + \tilde{v}\,\hat{\boldsymbol{n}}, \tag{23}$$

[7] This cut is similar to the one applied by Hamaus et al. (2022), who use a factor 1.86 instead of 1.5.

[8] https://github.com/cosmodesi/pycorr

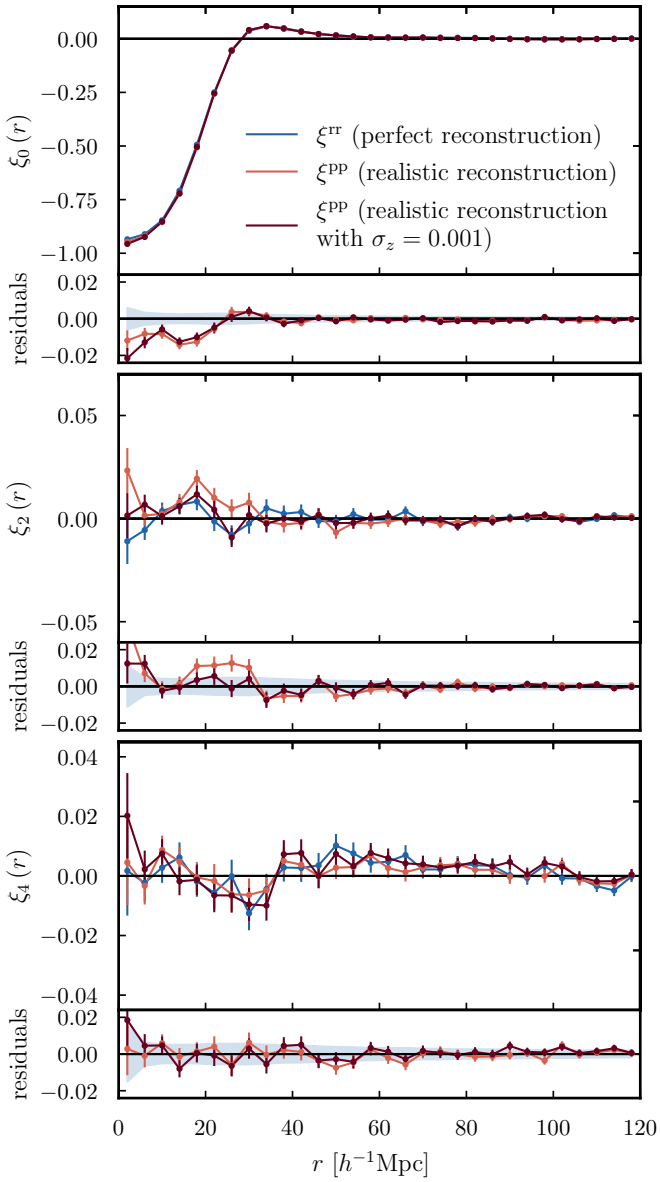

Fig. 3. Void-galaxy CCF in real space, measured in the Flagship simulation $z_{\rm eff} = 1.0$ redshift bin. From top to bottom, the panels show the monopole, quadrupole and hexadecapole moments of the CCF. The multipoles of $\xi^{\rm rr}$, corresponding to the perfect reconstruction scenario and measured using the true real-space positions are shown in blue, while light and dark red respectively show the multipoles of $\xi^{\rm pp}$ obtained using reconstruction without and with redshift errors. The inset panels show the residuals with respect to the perfect reconstruction case of $\xi^{\rm rr}_\ell$ for each multipole. In all cases ξ_2 and ξ_4 are close to zero, indicating that reconstruction successfully removes orientation-dependent selection bias.

where the stochastic velocity component has random magnitude $\tilde{v}$ and is directed along a random direction $\hat{\boldsymbol{n}}$. The quantity we wish to estimate is $\sigma^2_{v_\parallel}(r) = \langle \tilde{v}^2 (\hat{\boldsymbol{n}} \cdot \hat{\boldsymbol{r}})^2 \rangle$ and from Eq. (23) we obtain (Fiorini et al. 2022)

$$\sigma_{v_\parallel}(r) = \frac{\sqrt{\langle |\mathbf{v}|^2 \rangle - v_{\rm r}^2}}{\sqrt{3}}, \tag{24}$$

where $\langle |\mathbf{v}|^2 \rangle$ and $v_{\rm r}(r)$ are quantities independent of $\mu_{\rm r}$, and can be measured in radial bins.

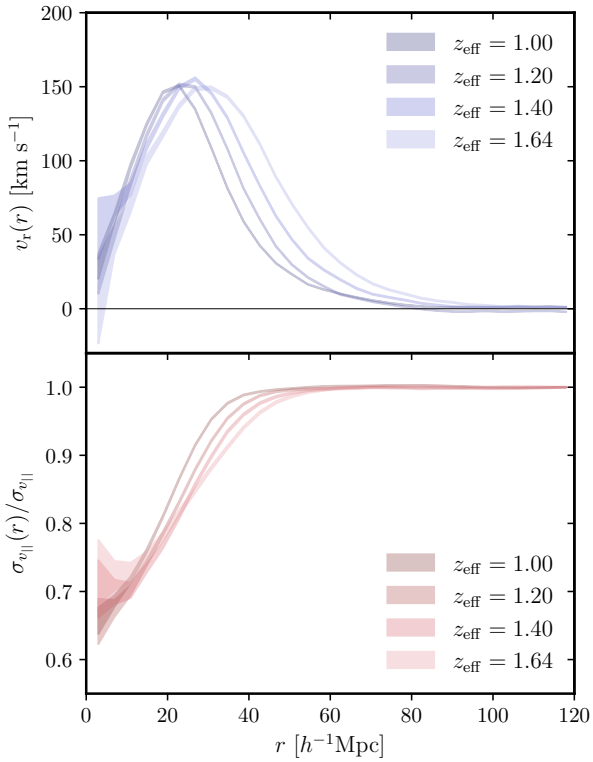

Fig. 4. Measured radial outflow velocity profile $v_{\rm r}(r)$ (top), and normalised line-of-sight velocity dispersion $\sigma_{v_\parallel}(r)$ (bottom) of galaxies around voids found in the four redshift bins of Flagship. Shading represents the 68.3% (1σ) measurement uncertainty coming from Poisson noise. The dispersion amplitude $\sigma_{\rm v}$ for the four bins can be found in Table 1.

The bottom panel of Fig. 4 shows the normalised template functions for $\sigma_{v_\parallel}(r)$ measured for each of the four redshift bins. The measured amplitudes at large distances, $\sigma_{v_\parallel}$, are listed in Table 1, where we see that the dispersion is higher at lower redshift, consistent with the growth of large-scale structure. From the figure we also see that the dispersion decreases in the void interior, where the potential is steep and the bulk outflow motion dominates.

The final template function required is that for $\xi^{\rm rr}(\boldsymbol{r})$. In the two realistic reconstruction scenarios we test, we assume that the true real-space information is not available, so that $\xi^{\rm rr}$ cannot be directly measured. Instead we used the quantity $\xi^{\rm pp}(\boldsymbol{r})$, which could be measured using the approximate real-space catalogues, as an estimator for $\xi^{\rm rr}(\boldsymbol{r})$. This is the same approach as used by Nadathur et al. (2019b, 2020b) and Woodfinden et al. (2022) except that here we do not have a large number of mock catalogues over which we can average this estimator. This greatly increases the noise in the estimate, which is especially relevant when attempting to determine the dependence of $\xi^{\rm pp}$ on the reconstruction parameter β. For the current work we therefore decided to fix this estimator to $\xi^{\rm pp}(\beta^{\rm fid})$ evaluated at the fiducial β value given in Table 1. This approach will be revisited for a future full *Euclid* analysis, by which time we expect a full complement of mocks to be available.

As mentioned in Sect. 2.3, our general code implementation seeks to preserve information on possible anisotropies in $\xi^{\rm pp}(\boldsymbol{r})$ which may be introduced by the AP effect or imperfections in reconstruction, without imposing the assumption of spherical symmetry, $\xi^{\rm pp}(\boldsymbol{r}) = \xi^{\rm pp}(r)$. In practice we achieve this by measuring the monopole $\xi^{\rm pp}_0(r)$, quadrupole $\xi^{\rm pp}_2(r)$ and hexadecapole $\xi^{\rm pp}_4(r)$ moments and using them to reconstruct $\xi^{\rm pp}(\boldsymbol{r})$.

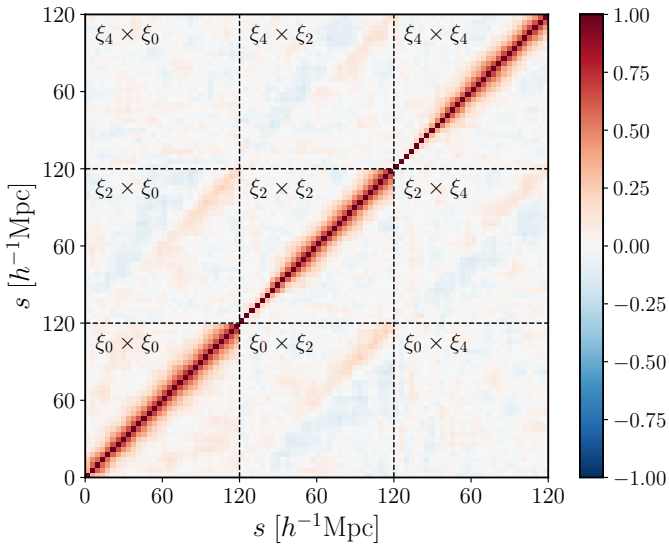

Fig. 5. Normalised covariance matrix $\mathsf{C}^{\rm data}$ for the data vector from the Flagship $z_{\rm eff} = 1.0$ redshift bin, estimated using jackknife resampling. Dashed lines differentiate between monopole, quadrupole and hexadecapole blocks of the data vector.

In the idealised perfect reconstruction scenario, things are much simpler and we just use the true real-space information from Flagship and directly measure the multipoles $\xi_0^{\rm rr}(r)$, $\xi_2^{\rm rr}(r)$ and $\xi_4^{\rm rr}(r)$ from the data.

4.3. Model fitting and parameter inference

We compare the model to the measured redshift-space CCF in terms of the compression to multipoles. In each redshift bin, we define the data vector $\boldsymbol{\xi}^{\rm s} \equiv \left(\xi_0^{\rm rs}, \xi_2^{\rm rs}, \xi_4^{\rm rs}\right)$ formed by concatenating the monopole, quadrupole and hexadecapole moments of $\xi^{\rm rs}(s, \mu_{\rm s})$, measured in 30 radial bins. In the two realistic reconstruction scenarios where the true real-space information is not used, we use the CCF $\xi^{\rm ps}(s, \mu_{\rm s})$ measured using the void positions identified in the post-reconstruction, approximate real-space galaxy distribution to construct $\boldsymbol{\xi}^{\rm s}$. In this case the data vector inherits a dependence on the reconstruction parameter β from the void positions.

To determine the uncertainty in this measurement of $\boldsymbol{\xi}^{\rm s}$, we used a jackknife resampling method to estimate the covariance matrix. For each of the four redshift bins we divided the Flagship data sample into $N_{\rm JK} = 900$ non-overlapping sub-regions of equal volume. We iterated over all sub-boxes, in each case estimating the cross-correlation function $\boldsymbol{\xi}^{\rm s(k)}$ by removing all the void-galaxy pairs for which both void and galaxy lie in the kth sub-box, and reweighting the ones where either void or galaxy lie within the sub-box using the matched weighting scheme prescribed by Mohammad & Percival (2022) and implemented in `pycorr`. We then used these realisations to estimate the covariance between the ith and jth bins of the data vector

$$\mathsf{C}_{ij}^{\rm data} = \frac{N_{\rm JK} - 1}{N_{\rm JK}} \sum_{k=1}^{N_{\rm JK}} \left(\boldsymbol{\xi}_i^{\rm s(k)} - \bar{\boldsymbol{\xi}}_i^{\rm s}\right)\left(\boldsymbol{\xi}_j^{\rm s(k)} - \bar{\boldsymbol{\xi}}_j^{\rm s}\right), \tag{25}$$

where $\bar{\boldsymbol{\xi}}^{\rm s}{}_i = \frac{1}{N_{\rm JK}} \sum_{k=1}^{N_{\rm JK}} \boldsymbol{\xi}_i^{\rm s(k)}$ is the mean of the $N_{\rm JK}$ jackknife realisations. The correlation structure of the resultant covariance matrix is shown in Fig. 5. Due to the limitations of having only the single Flagship mock, there will be an additional correlation between the model prediction and the measured data vector that has been neglected in this estimate of the covariance but is discussed in Appendix A.

At any point in parameter space, we computed the model prediction for $\boldsymbol{\xi}^{\rm s}$ as described in Sect. 2 using the public code `victor`[9]. We assume that the likelihood has a Gaussian form,

$$\log \mathcal{L} = -\frac{1}{2} \left(\boldsymbol{\xi}^{\rm s}{}_{\rm theory} - \boldsymbol{\xi}^{\rm s}{}_{\rm data}\right) \mathsf{C}^{-1} \left(\boldsymbol{\xi}^{\rm s}{}_{\rm theory} - \boldsymbol{\xi}^{\rm s}{}_{\rm data}\right)^{\rm T}. \tag{26}$$

In the realistic reconstruction scenarios we consider, this likelihood evaluation depends on four parameters: explicitly on $f\sigma_8$, $\sigma_{v_\parallel}$ and ϵ, and implicitly on the reconstruction parameter β via its effect on the recovered void positions in the post-reconstruction field and thus on the measured correlations. In the idealised case of perfect reconstruction there is no β dependence. We sampled the full parameter space using the interface between `victor` and the Monte Carlo Markov chain (MCMC) sampling code `Cobaya`[10] (Torrado & Lewis 2021). Repeating reconstruction, void-finding and CCF measurement at each β value along the chain would be too slow for the MCMC so instead we followed the method of previous works (Nadathur et al. 2019b, 2020b) and pre-computed the correlations on a grid of β values around the fiducial $\beta^{\rm fid}$ and interpolated the data vector at the time of evaluation of the likelihood[11]. To avoid extrapolating outside of this grid, the prior on β was limited to the same range as the interpolation; we used uninformative flat priors for all of the other parameters. Additionally, convergence of the MCMC chains is checked using the Gelman-Rubin $R - 1$ statistic, using the convergence criterion $R - 1 < 0.01$, and 20% of the initial steps were discarded as burn-in.

5. Results

5.1. Fits to Flagship data

In Fig. 6 we show the measured monopole, quadrupole and hexadecapole moments of the redshift-space CCF $\xi^{\rm ps}$ and the predictions for the best-fit model, together with the residuals, in each of the redshift bins for the realistic reconstruction case and for the catalogue with redshift errors. It can be seen that the model describes the observed multipoles well, although the hexadecapole moment does not contain much information except in the lowest redshift bin where the density of galaxies and voids is highest. This is the most realistic of the scenarios considered, but a similar quality of fit was obtained for the idealised perfect reconstruction scenario and when neglecting redshift errors.

The marginalised one-dimensional posterior constraints on the growth rate $f\sigma_8$ and the AP distortion parameter ϵ individually obtained from analysis of the Flagship mock are summarised in Table 2. For the idealised case of perfect reconstruction, we recover unbiased constraints on both $f\sigma_8$ and ϵ in all redshift bins. We find a measurement precision on ϵ of about 0.5–0.6% in each individual redshift bin, and the precision on $f\sigma_8$ varies between ~8.5% for the lowest redshift bin and ~12% in the two highest redshift bins. Using realistic reconstruction and adding redshift errors to the mocks does not significantly degrade these constraints, and the central values recovered are consistent to

[9] https://github.com/seshnadathur/victor
[10] https://github.com/CobayaSampler/cobaya
[11] Woodfinden et al. (2022) instead perform the interpolation at the level of the likelihood itself; both options are implemented in `victor` and give very similar results.

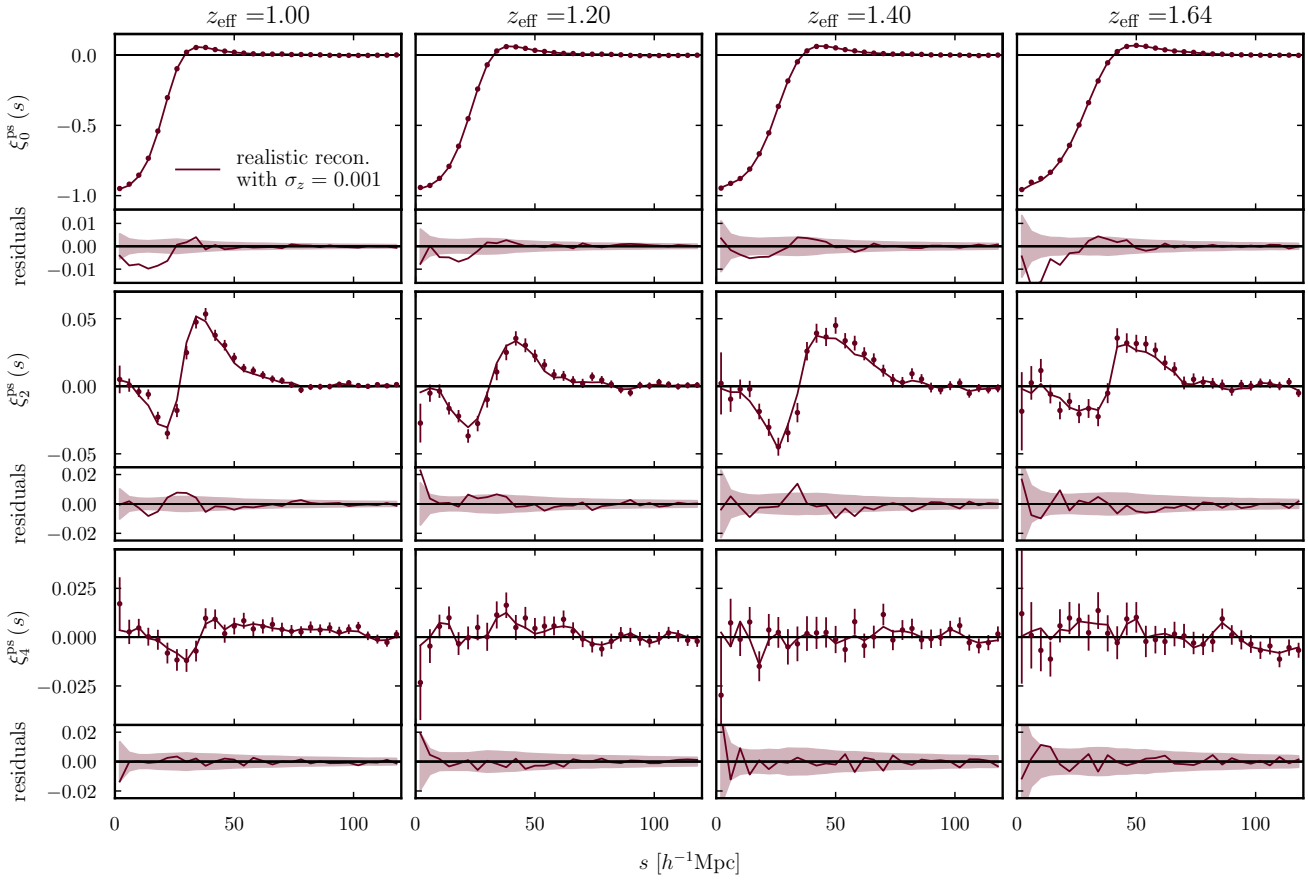

Fig. 6. Measured multipole moments of the redshift-space void-galaxy CCF, $\xi_\ell^{\rm ps}$, for realistic reconstruction applied to the catalogue including redshift errors, are shown as the data points. Error bars shown are determined from the diagonal entries of the covariance matrix $\mathsf{C}^{\rm data}$. The solid lines show the corresponding best-fit models, and insets show the residuals of the model with respect to the data. The rows correspond to (from top to bottom) the monopole, quadrupole and hexadecapole moments ($\ell = 0, 2, 4$). Columns correspond to the different redshift bins, labelled with the effective redshift $z_{\rm eff}$.

Table 2. Marginalised posterior constraints on the AP distortion parameter ϵ and growth rate $f\sigma_8$ obtained from fits to the Flagship mock data in four redshift bins.

	$z_{\rm eff}$	$f\sigma_8$	ϵ
Flat prior range		[0.05, 1.5]	[0.8, 1.2]
Perfect reconstruction	1.00	0.46 ± 0.04	1.002 ± 0.005
	1.20	0.41 ± 0.05	1.002 ± 0.006
	1.40	0.40 ± 0.05	1.003 ± 0.006
	1.64	0.41 ± 0.04	1.000 ± 0.006
Realistic reconstruction	1.00	0.49 ± 0.04	0.995 ± 0.005
	1.20	0.44 ± 0.05	1.000 ± 0.005
	1.40	0.40 ± 0.05	1.005 ± 0.006
	1.64	0.43 ± 0.05	0.999 ± 0.006
Realistic reconstruction with $\sigma_z = 0.001$	1.00	0.51 ± 0.04	0.997 ± 0.005
	1.20	0.44 ± 0.05	1.004 ± 0.005
	1.40	0.40 ± 0.06	1.005 ± 0.006
	1.64	0.44 ± 0.05	1.000 ± 0.006

Notes. We present results in three scenarios: idealised perfect reconstruction (using true real-space galaxy positions), realistic reconstruction and realistic reconstruction performed on a catalogue with added redshift errors. Fits are performed using covariance $\mathsf{C}^{\rm data}$ estimated for Flagship. The true value of ϵ is 1 as the data are analysed in the Flagship fiducial cosmology. The true values of $f\sigma_8$ in each redshift bin are as listed in Table 1.

within the stated uncertainty in each case. The reconstruction method used is therefore not causing a significant loss of information.

The marginalised 1D and 2D posteriors on $f\sigma_8$ and ϵ are shown in Fig. 7, where the dashed crosshairs indicate the fiducial values for the Flagship cosmology, which are consistently recovered. Comparing the perfect and realistic reconstruction contours, the largest shift (of $\sim 1\sigma$) is seen in the lowest redshift bin. The introduction of redshift errors does not cause significant additional shifts in any redshift bin.

Figure 8 shows the marginalised 1D measurements of the growth rate and ratio of the transverse comoving distance $D_{\rm M}$ to the Hubble distance $D_{\rm H}$, derived from the ϵ constraints, from the Flagship data as a function of redshift. These measurements are compared to the 68.3% (1σ) and 95.5% (2σ) confidence intervals on these quantities derived from extrapolating the CMB constraints from Planck Collaboration VI (2020) down to these redshifts while assuming a ΛCDM cosmological model. Results for ϵ itself are shown in the inset plot.

5.2. Full Euclid volume

The planned final footprint of *Euclid* is 15 000 deg^2, almost three times larger than that of the Flagship mock used in the analysis above. The constraints we expect from using the full *Euclid*

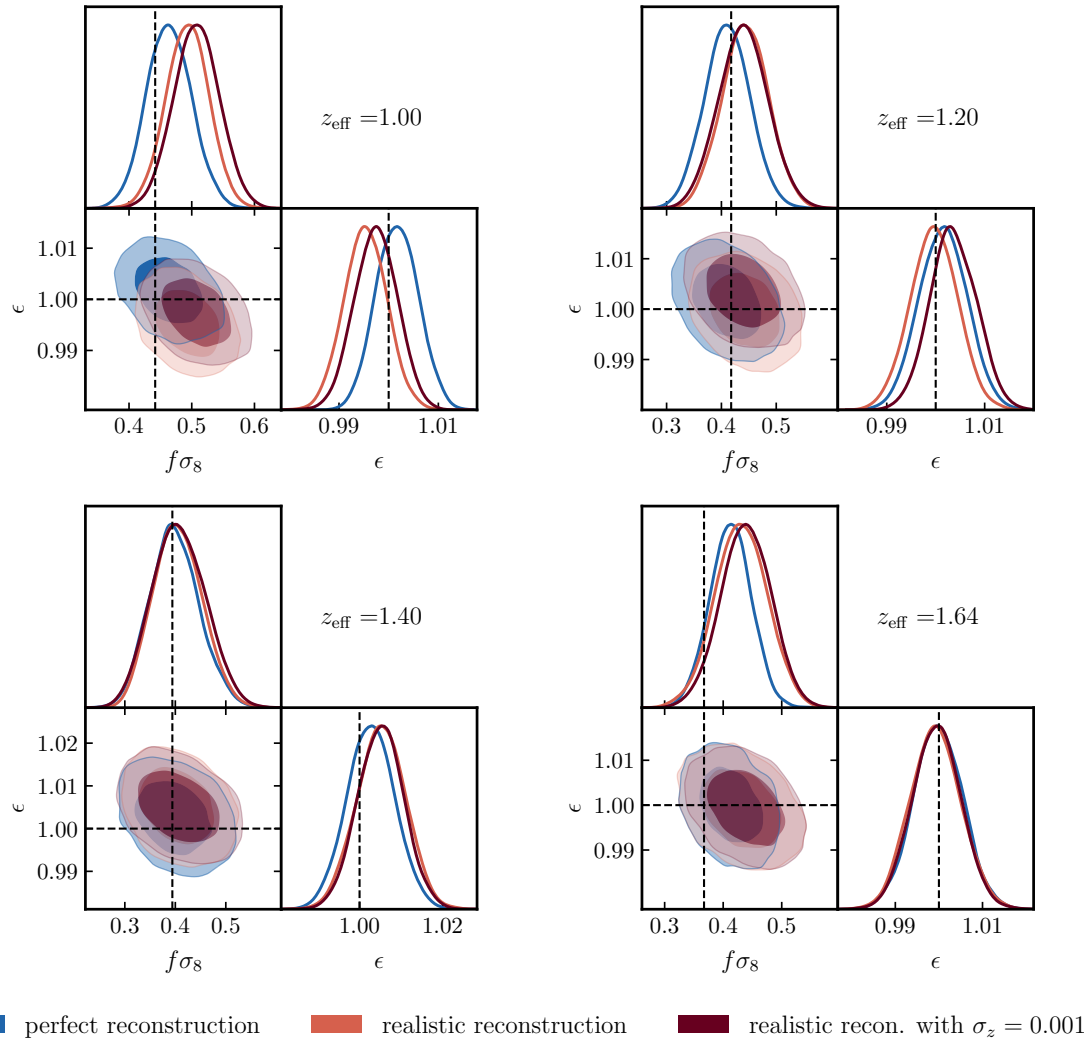

Fig. 7. Marginalised 1D and 2D posterior constraints for parameters $f\sigma_8$ and ϵ from fits to the Flagship void-galaxy CCF measurements in four redshift bins. Contours show the 68.3% (1σ) and 95.5% (2σ) confidence intervals obtained in the case of perfect reconstruction (blue), realistic reconstruction (light red), and realistic reconstruction with added redshift error (dark red). Dashed crosshairs indicate the true values for the Flagship cosmology.

volume will therefore be correspondingly better. A simple scaling of the covariance matrix suggests that it should shrink proportional to the increase in the volume, that is by a factor of 2.91. This corresponds to measurement uncertainties on the binned multipole moments that should be over 40% smaller than those shown in Fig. 6. We first confirmed this scaling with survey volume by computing covariance matrices for eight volume subsections of the Flagship simulation, ranging from 0.125 to 0.875. We confirmed that the terms of C scale roughly linearly with the survey volume. We therefore obtained the expected covariance matrix C_{ij}^{Euclid} for the full *Euclid* survey by scaling the one obtained in Sect. 4.3 by a factor of 1/2.91.

To test how this reduction in uncertainties in the measurement of the void-galaxy CCF propagates through to the recovered cosmological parameter constraints, we repeated the parameter inference step with this rescaled covariance and using a synthetic data vector generated assuming this level of measurement noise. To do this, we considered the idealised perfect reconstruction scenario. Using the real-space CCF multipoles $\xi_\ell^{\rm rr}(r)$ measured from Flagship, we computed the fiducial model prediction, $\boldsymbol{\xi}^{\rm s,th}$, for the redshift-space data vector. We then created a synthetic data vector

$$\boldsymbol{\xi}^{\rm s,mock} = \mathsf{L}\boldsymbol{Z} + \boldsymbol{\xi}^{\rm s,th}, \tag{27}$$

where $\mathsf{L}\mathsf{L}^{\rm T} = \mathsf{C}^{Euclid}$, L is obtained by a Cholesky decomposition of the scaled covariance matrix C^{Euclid}, and $\boldsymbol{Z}$ is a vector of independent standard normal random variables. We then repeated the parameter inference of Sect. 4.3 fitting to $\boldsymbol{\xi}^{\rm s,mock}$ using the rescaled covariance matrix.

The results of this fitting procedure provide the forecast for *Euclid* constraints from the void-galaxy CCF measurement, and form the main result of this paper. The forecasts are summarised in Table 3 and shown as a function of redshift in Fig. 9, compared to existing constraints from the application of the same void-galaxy method on current data. As expected, we recover the fiducial cosmological parameters up to the 68.3% (1σ) statistical uncertainty in each redshift bin. Comparing Tables 2 and 3 shows that while the forecast uncertainty in ϵ scales approximately as the $1/\sqrt{2.91}$ factor expected from the simple volume scaling, the forecast constraints on $f\sigma_8$ are slightly more pessimistic. Our final results suggest that with *Euclid* voids we will be able to measure the distance ratio $D_{\rm M}/D_{\rm H}$ to a precision of

0.3–0.4% in four redshift bins over the range $0.9 \leq z < 1.8$. This is comparable even to the extrapolated constraints from *Planck* which assume ΛCDM. Voids will also provide a precision of about 5–8% in measurement of the growth rate over these redshifts, a large increase in the statistical power of this method compared to current surveys, and a useful consistency check for results from galaxy clustering.

Figure 9 highlights the key point that AP constraints from voids will be competitive with those derived from model-dependent extrapolations to low redshift from the CMB (and will exceed those that can be obtained from galaxy clustering alone). AP measurements therefore present the best potential for information gain from using the void-galaxy CCF. On the other hand, while the constraints on $f\sigma_8$ are weaker than those that can be obtained from galaxy clustering alone and do not add much cosmological information, they can serve as useful consistency checks.

5.3. Cosmology forecasts

We now investigate the impact on cosmological model forecasts of the expected measurement precision that can be achieved with *Euclid* voids, summarised in Table 3. Euclid Collaboration (2020) have presented forecasts for the constraints on model parameters that can be achieved with the primary *Euclid* probes, namely spectroscopic galaxy clustering, weak lensing, photometric galaxy clustering, and their cross-correlations. In this work, we compare voids to these benchmarks as a separate standalone probe, and do not investigate the combination of voids with other observables (see, for instance, Nadathur et al. 2019b, 2020b; Bonici et al. 2023 for examples of such combinations).

In this scenario, voids probe cosmology through the AP measurement of $D_{\rm M}(z)/D_{\rm H}(z)$ and the stringent constraints this places on the background expansion history. In contrast, the growth rate constraints are weaker than those that can be obtained from galaxy clustering. These measurements can be used as tests of consistency of the model in different regimes, but in the absence of any discrepancy they do not add much cosmological information, even when voids are combined with other probes (Nadathur et al. 2020b). When treating voids as a standalone probe as we do here, they do not provide sufficient constraints to be of interest. We therefore neglect the $f\sigma_8$ results and focus entirely on the geometrical AP constraints. To do so, we centred the $D_{\rm M}(z)/D_{\rm H}(z)$ values at their expected values for the Flagship cosmology and used the measurement uncertainties from Table 3.

Within the flat ΛCDM model scenario, measurement of $D_{\rm M}/D_{\rm H}$ at a given redshift directly translates to a constraint on the matter density parameter $\Omega_{\rm m}$ (or, equivalently, on $\Omega_\Lambda = 1-\Omega_{\rm m}$). The precision that we can achieve on $D_{\rm M}(z)/D_{\rm H}(z)$ with voids corresponds to $\Omega_{\rm m} = 0.3183 \pm 0.0028$ (at the 68.3% confidence level), a relative statistical uncertainty of ±0.009. The corresponding forecast relative uncertainty from *Euclid* spectroscopic galaxy clustering in the optimistic (pessimistic) forecast scenario is 0.013 (0.021), from weak lensing is 0.012 (0.018), and from the combination of the two is 0.006 (0.009) (Euclid Collaboration 2020). Thus for this model and parameter, *Euclid* voids alone seem to outperform both galaxy clustering and weak lensing individually, and look to be competitive with their combination. It is worth noting, however, that this forecast is based only on the statistical constraints, obtained by rescaling a Flagship covariance up to the full *Euclid* volume.

We also consider a minimal one-parameter extension to the ΛCDM model known as wCDM, in which the cosmological constant Λ is replaced by a dark energy component with a constant equation of state that is not restricted to $w = -1$. Fig. 10 shows the forecast constraints obtained on $\Omega_{\rm m}$ and w in the dark red contours. We recover $w = -1.00^{+0.06}_{-0.05}$ and $\Omega_{\rm m} = 0.3183 \pm 0.0028$, corresponding to a relative precision of 6% on w. For comparison, the yellow contours in Fig. 10 show the current constraints on these parameters from SDSS at multiple redshifts between $z = 0.15$ and $z = 2.33$, comprising the BAO results from different galaxy and quasar samples reported by Alam et al. (2021) together with recent improved high-redshift AP measurements from the 'full shape' of the Lyman-α forest clustering (Cuceu et al. 2023), which give $w = -0.90 \pm 0.12$. Due to a well-known geometrical degeneracy which is broken by low-redshift AP measurements, the constraints on this model from the CMB (including CMB lensing) are relatively weak, $w = -1.57^{+0.50}_{-0.40}$ (Planck Collaboration VI 2020). For completeness, Fig. 10 also shows two sets of forecasts for the *Euclid* void-galaxy CCF obtained by Hamaus et al. (2022) through a different method for modelling and measurement to that used in this work.

These results highlight the excellent promise of the void-galaxy CCF as a standalone cosmological probe. When removing the assumption of flatness and allowing more general dark energy models with additional degrees of freedom, the void measurement of $D_{\rm M}/D_{\rm H}$ is no longer sufficient to provide useful model constraints on its own. However, the higher precision of the AP measurement still provides large gains in information and improves the figure of merit for dark energy in these models when voids are used in combination with other probes, as shown by Nadathur et al. (2020a). Thus this technique retains great potential to enhance *Euclid* science.

6. Conclusions

We have presented a forecast of the cosmological constraints that can be obtained from measurement of the anisotropic void-galaxy cross-correlation function (CCF) in the spectroscopic sample of the upcoming *Euclid* galaxy survey. The pattern of anisotropies in the redshift-space CCF can be used to simultaneously fit for AP distortions caused by differences between the fiducial analysis model and the true background cosmology, and for RSD caused by the peculiar motions of galaxies around voids. Similar cosmological analyses of the void-galaxy CCF have previously been applied to data from the Sloan Digital Sky Survey (Nadathur et al. 2019b, 2020b; Hamaus et al. 2020; Woodfinden et al. 2022), and have been shown to provide significantly better AP measurements than obtained from traditional galaxy-clustering analyses, including BAO. Precise AP measurements, particularly at high redshift, provide very powerful probes of cosmological models of dark energy (Nadathur et al. 2020a; Cuceu et al. 2023).

Our forecast is based on an analysis of the lightcone Flagship galaxy mock catalogue, which was calibrated to match the expected properties of the *Euclid* spectroscopic galaxy sample, and covers 5157 deg^2 of the sky between the redshifts $0.9 \leq z < 1.8$. We analysed this mock data using the same method previously used by Nadathur et al. (2019b) and Woodfinden et al. (2022). A key component of this method is the use of velocity field reconstruction to approximately remove large-scale RSD from the galaxy distribution before identifying voids. This step is necessary to remove a selection bias in the construction of the void sample that depends on the orientation of the voids to the line-of-sight direction, thus ensuring the validity of the theoretical modelling. To assess the efficiency of the particular

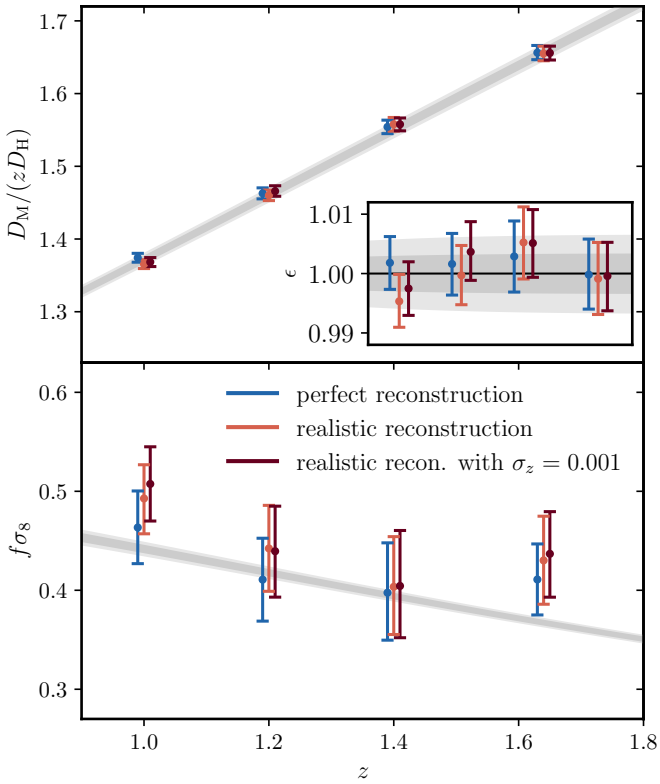

Fig. 8. Redshift dependence of the measured cosmological parameters obtained from fits to Flagship simulation data. The top panel shows constraints on the ratio of comoving transverse and Hubble distances divided by redshift z, $D_M/(zD_H)$, which is derived from ϵ (inset). The bottom panel shows measurements of the growth rate $f\sigma_8$. Coloured data points indicate the different reconstruction scenarios considered, plotted at slightly shifted redshift locations for clarity. The grey shaded bands represent the 68.3% (1σ) and 95.5% (2σ) confidence intervals from fits to Planck Collaboration VI (2020), extrapolated down to these redshifts assuming ΛCDM, and centred on the Flagship cosmology.

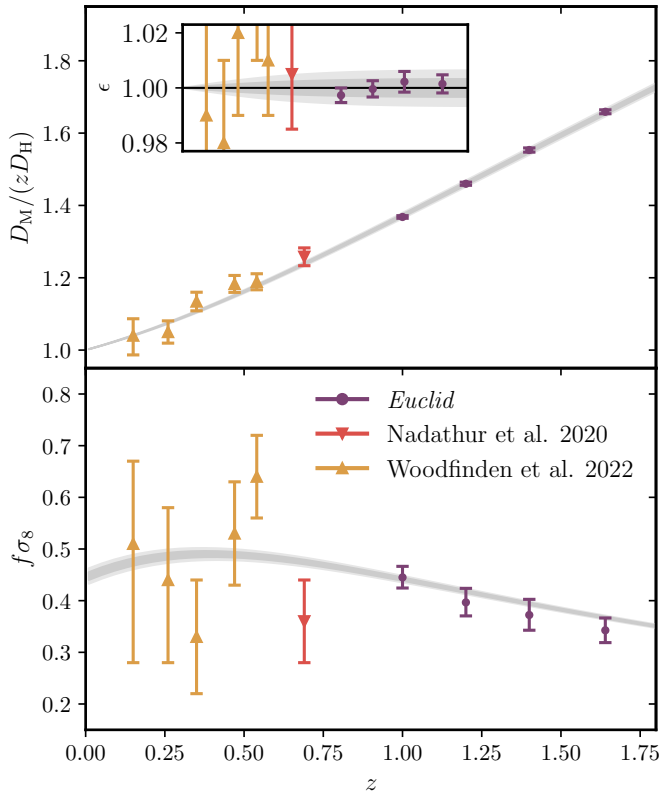

Fig. 9. Redshift dependence of the forecast measurements that could be obtained from the 15 000 deg^2 *Euclid* survey (purple data points). These are generated from a random realisation of the *Euclid* CCF (Sect. 5.2) and so the central values can be shifted relative to the Flagship results in Fig. 8. Low redshift measurements from current SDSS data obtained using the same analysis method by Nadathur et al. (2020b) and Woodfinden et al. (2022) are included for comparison.

Table 3. Forecast marginalised posterior constraints on ϵ and $f\sigma_8$ from the full *Euclid* survey, obtained from fitting a synthetic data vector using an appropriately scaled data covariance matrix.

$z_{\rm eff}$	$f\sigma_8$	ϵ	ρ
1.00	$0.445^{+0.022}_{-0.021}$	0.9973 ± 0.0027	−0.28
1.20	$0.397^{+0.028}_{-0.027}$	$0.9996^{+0.0031}_{-0.0030}$	−0.23
1.40	$0.372^{+0.031}_{-0.030}$	1.002 ± 0.004	−0.25
1.64	0.343 ± 0.024	$1.0014^{+0.0034}_{-0.0033}$	−0.16

Notes. Uncertainties in the two parameters are correlated, with correlation coefficient ρ.

algorithm used to perform this reconstruction, we compared the results obtained using the true real-space galaxy field – corresponding to the idealised perfect reconstruction scenario, that will not be achievable in practice with *Euclid* data – to those obtained with realistic reconstruction, with and without including spectroscopic redshift errors in the mock data.

We performed the analysis in four non-overlapping redshift bins, compressing the anisotropic CCF information to its first three even-order multipoles – monopole, quadrupole, and hexadecapole – which contain all of the practically measurable information. We found that using a realistic reconstruction algorithm did not significantly increase the statistical uncertainty of the recovered marginalised constraints on the cosmological parameters ϵ and $f\sigma_8$ compared to the idealised scenario, and also did not introduce significant systematic offsets at the level of statistical precision available from the Flagship mock. Realistic spectroscopic redshift errors were found to have a similarly minor effect on the results. However, we caution that the analysis here was limited by the availability of only a single simulation realisation, so we can not guarantee that systematic offsets will not be present when dealing with *Euclid* data. A more detailed study of systematic errors should be performed when a larger number of *Euclid* mocks become available.

We then extrapolated from the results obtained from the Flagship mock to the full *Euclid* survey, covering almost three times larger survey volume, to obtain our forecast for the achievable statistical precision. We found that void-galaxy CCF measurements with *Euclid* should be able to achieve a statistical precision of down to 0.3% on the measurement of the distance ratio $D_M(z)/D_H(z)$ over the redshift range $0.9 \leq z < 1.8$, and 5–8% on the growth rate parameter $f\sigma_8$. Within the context of flat ΛCDM cosmological models, the statistical uncertainty we achieve on $D_M(z)/D_H(z)$ translates to a relative statistical uncertainty of 0.009 on the matter density parameter Ω_m which is ∼30% better than forecasts of what is achievable from both the *Euclid* spectroscopic galaxy clustering and weak lensing probes individually, and is comparable to their combination (Euclid Collaboration 2020). In the wCDM model scenario,

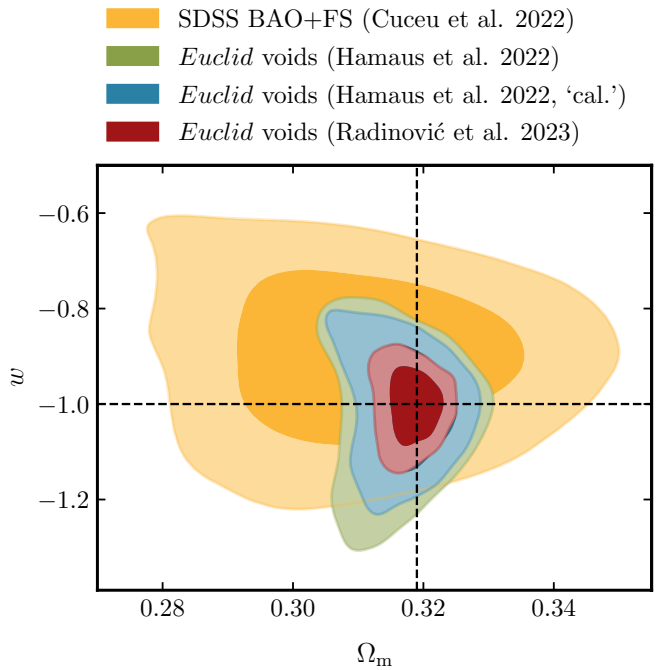

Fig. 10. Forecast constraints on the matter density parameter Ω_m and the dark energy equation of state w in a wCDM cosmological model from a measurement of the void-galaxy CCF with *Euclid* as outlined in this work (red contours), corresponding to $w = -1.00^{+0.06}_{-0.05}$. We also show previous forecasts for the *Euclid* void-galaxy CCF – obtained using a different model and measurement method – by Hamaus et al. (2022), for the 'independent' (green contours) and 'calibrated' (blue, labelled 'cal.' here) cases in that paper, corresponding to whether nuisance parameters are marginalised over or fixed. For context, we show current measurements from current BAO and 'full-shape' measurements in the SDSS galaxy, quasar and Lyman-α datasets from the MGS, BOSS and eBOSS surveys (yellow) presented by Cuceu et al. (2023).

we found that voids alone can measure the dark energy equation of state to a relative precision of 6%, $w = -1.00^{+0.06}_{-0.05}$.

It should be noted here that the modelling method adopted for this forecast is the template-fitting approach used in several previous works. In this approach template functions describing the real-space CCF, the void velocity profile (or equivalently, matter density profile) and the velocity dispersion profile are first constructed by calibrating against simulation results and then allowed to vary with cosmological parameters, as described in Sect. 2. The recent work of Massara et al. (2022) suggests that improvements to this template-fitting approach may be required in the future, but our aim is to forecast the constraints possible with currently available techniques so we have not considered these here. We have also not accounted for possible systematic effects which may enter this analysis if the simulations from which the templates are constructed differ strongly from the true cosmological model. A quantitative study of these effects (such as conducted in the context of SDSS by Woodfinden et al. 2022) would require mock catalogues constructed at different cosmologies, but at present we only have access to a single realisation of the *Euclid* Flagship mock at a fixed cosmology. A large number of mock catalogues would also allow a better estimate of the covariance matrix than the jackknife resampling technique used here.

In this work, the template we used for the real-space CCF was necessarily estimated from the single realisation of the Flagship data itself. This increases the noise in this estimate, in a manner that is correlated with measurement noise in the data vector. When the final *Euclid* data are analysed, we anticipate that large numbers of mock realisations will be available, so that while a similar procedure could be used in the final analysis, it will not be mandatory. If indeed both real-space and redshift-space CCFs are measured from the same data, it would be necessary to account for the correlation between them for a fully self-consistent analysis. We describe how to do this in Appendix A, although we choose not to include this in the main analysis presented above. However, as shown in Appendix A, the effect of accounting for this correlation is to *decrease* the effective statistical uncertainties on the recovered cosmological parameters. The forecasts using the larger statistical uncertainties presented in our headline analysis are therefore expected to be conservative.

Our treatment of possible observational systematics is quite simple and ignores some effects that we expect will be present in the *Euclid* data. In particular, although we have incorporated the effects of a uniformly low completeness of the spectroscopic sample and a Gaussian distribution of redshift errors, we have not considered the purity of the sample and errors due to spectral line misidentification. These effects will be explored in detail in a future work. This paper is one of several companion papers forecasting the constraining power from different void observables in *Euclid* (Hamaus et al. 2022; Contarini et al. 2022; Bonici et al. 2023) which all use the same common baseline of assumptions, to allow a consistent comparison across methods.

Hamaus et al. (2022) have also investigated forecasts for constraints obtained from the void-galaxy CCF with Flagship, using a different approach. Unlike in our work, that paper did not use reconstruction to try to remove the orientation-dependent sample selection bias in the void catalogues, and thus obtained sharply different measurements of the CCF than those we find here. They then introduced additional nuisance parameters and modified the growth-dependent terms of the theoretical model derived from the conservation equation in Sect. 2 in order to describe the Flagship data. Other minor differences in approach include the assumption that the matter density profile around voids is simply related to the galaxy profile by the large-scale linear galaxy bias b (instead of our template method for predicting the mean galaxy velocity), and neglecting velocity dispersion around the mean.

Despite these differences, the forecasts of Hamaus et al. (2022) are broadly compatible with ours. In their baseline model (which they refer to as 'independent'), which marginalises over nuisance parameters as we do here, they predict a 0.5% measurement of the AP parameter, somewhat weaker than our 0.3%, and a 4% measurement of the growth rate, slightly better than our value of 5%–8%. They further report that if the nuisance parameters are instead fixed to their best-fit values (referred to as the 'calibrated' case), these results could improve to 0.4% and 1% precision on the AP parameter and growth rate, respectively. The additional improvement in the AP measurement that is possible from our method can have a large effect on the cosmological constraints obtained, as shown in Fig. 10, where uncertainties in w and Ω_m are significantly reduced. Taken together, the results of these two papers showcase both the promise and the robustness of the void-galaxy cross-correlation method applied to *Euclid* data.

Acknowledgements. SR acknowledges support from the Research Council of Norway through project 325113. SN acknowledges support from an STFC Ernest Rutherford Fellowship, grant reference ST/T005009/2. AW acknowledges the support of the Natural Sciences and Engineering Research Council of Canada

(NSERC) [funding reference number 547744]. Research at Perimeter Institute is supported in part by the Government of Canada through the Department of Innovation, Science and Economic Development Canada and by the Province of Ontario through the Ministry of Colleges and Universities. This work has made use of CosmoHub (Carretero et al. 2017; Tallada et al. 2020). CosmoHub has been developed by the Port d'Informació Científica (PIC), maintained through a collaboration of the Institut de Física d'Altes Energies (IFAE) and the Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT) and the Institute of Space Sciences (CSIC and IEEC), and was partially funded by the "Plan Estatal de Investigación Científica y Técnica y de Innovación" programme of the Spanish government. The Euclid Consortium acknowledges the European Space Agency and a number of agencies and institutes that have supported the development of *Euclid*, in particular the Academy of Finland, the Agenzia Spaziale Italiana, the Belgian Science Policy, the Canadian Euclid Consortium, the French Centre National d'Études Spatiales, the Deutsches Zentrum für Luft- und Raumfahrt, the Danish Space Research Institute, the Fundação para a Ciência e a Tecnologia, the Ministerio de Ciencia e Innovación, the National Aeronautics and Space Administration, the National Astronomical Observatory of Japan, the Netherlandse Onderzoekschool Voor Astronomie, the Norwegian Space Agency, the Romanian Space Agency, the State Secretariat for Education, Research and Innovation (SERI) at the Swiss Space Office (SSO), and the United Kingdom Space Agency. A complete and detailed list is available on the *Euclid* web site (http://www.euclid-ec.org).

References

Achitouv, I. 2017, Phys. Rev. D, 96, 083506
Achitouv, I. 2019, Phys. Rev. D, 100, 123513
Alam, S., Ata, M., Bailey, S., et al. 2017, MNRAS, 470, 2617
Alam, S., Aubert, M., Avila, S., et al. 2021, Phys. Rev. D, 103, 083533
Alcock, C., & Paczynski, B. 1979, Nature, 281, 358
Alonso, D., Hill, J. C., Hložek, R., & Spergel, D. N. 2018, Phys. Rev. D, 97, 063514
Aubert, M., Cousinou, M.-C., Escoffier, S., et al. 2022, MNRAS, 513, 186
Bautista, J. E., Paviot, R., Vargas Magaña, M., et al. 2021, MNRAS, 500, 736
Bayer, A. E., Villaescusa-Navarro, F., Massara, E., et al. 2021, ApJ, 919, 24
Behroozi, P. S., Wechsler, R. H., & Wu, H.-Y. 2013, ApJ, 762, 109
Beutler, F., Seo, H.-J., Saito, S., et al. 2017, MNRAS, 466, 2242
Blake, C., & Glazebrook, K. 2003, ApJ, 594, 665
Bonici, M., Carbone, C., Vielzeuf, P., et al. 2023, A&A, 670, A47
Bonnaire, T., Aghanim, N., Kuruvilla, J., & Decelle, A. 2022, A&A, 661, A146
Burden, A., Percival, W. J., & Howlett, C. 2015, MNRAS, 453, 456
Cai, Y.-C., Taylor, A., Peacock, J. A., & Padilla, N. 2016, MNRAS, 462, 2465
Carretero, J., Tallada, P., Casals, J., et al. 2017, in Proceedings of the European Physical Society Conference on High Energy Physics. 5-12 July, 488
Carron, J., & Szapudi, I. 2014, MNRAS, 439, L11
Chan, K. C., Hamaus, N., & Desjacques, V. 2014, Phys. Rev. D, 90, 103521
Chiang, C.-T., Wagner, C., Schmidt, F., & Komatsu, E. 2014, JCAP, 2014, 048
Chuang, C.-H., Kitaura, F.-S., Liang, Y., et al. 2017, Phys. Rev., D, 95, 063528
Contarini, S., Verza, G., Pisani, A., et al. 2022, A&A, 667, A162
Correa, C. M., Paz, D. J., Padilla, N. D., et al. 2019, MNRAS, 485, 5761
Correa, C. M., Paz, D. J., Padilla, N. D., et al. 2022, MNRAS, 509, 1871
Costille, A., Caillat, A., Rossin, C., et al. 2018, Proc. SPIE, 10698, 106982B
Cousinou, M. C., Pisani, A., Tilquin, A., et al. 2019, Astron. Comput., 27, 53
Cuceu, A., Font-Ribera, A., Nadathur, S., Joachimi, B., & Martini, P. 2023, Phys. Rev. Lett., 130, 191003
Eisenstein, D. J., Seo, H.-J., Sirko, E., & Spergel, D. 2007, ApJ, 664, 675
Euclid Collaboration (Blanchard, A., et al.) 2020, A&A, 642, A191
Fiorini, B., Koyama, K., & Izard, A. 2022, JCAP, 12, 028
Gil-Marín, H., Percival, W. J., Verde, L., et al. 2016, MNRAS, 465, 1757
Gil-Marín, H., Bautista, J. E., Paviot, R., et al. 2020, MNRAS, 498, 2492
Granett, B. R., Neyrinck, M. C., & Szapudi, I. 2008, ApJ, 683, L99
Guo, H., Zheng, Z., Jing, Y. P., et al. 2015, MNRAS, 449, L95
Hamaus, N., Sutter, P. M., & Wandelt, B. D. 2014, Phys. Rev. Lett., 112, 251302
Hamaus, N., Sutter, P. M., Lavaux, G., & Wandelt, B. D. 2015, JCAP, 2015, 036
Hamaus, N., Pisani, A., Sutter, P. M., et al. 2016, Phys. Rev. Lett., 117, 091302
Hamaus, N., Pisani, A., Choi, J.-A., et al. 2020, JCAP, 2020, 023
Hamaus, N., Aubert, M., Pisani, A., et al. 2022, A&A, 658, A20
Hawken, A. J., Granett, B. R., Iovino, A., et al. 2017, A&A, 607, A54
Hawken, A. J., Aubert, M., Pisani, A., et al. 2020, JCAP, 2020, 012
Kaiser, N. 1987, MNRAS, 227, 1
Kazin, E. A., Sánchez, A. G., & Blanton, M. R. 2012, MNRAS, 419, 3223
Kitaura, F.-S., Chuang, C.-H., Liang, Y., et al. 2016, Phys. Rev. Lett., 116, 171301
Kovács, A., Sánchez, C., García-Bellido, J., et al. 2019, MNRAS, 484, 5267
Landy, S. D., & Szalay, A. S. 1993, ApJ, 412, 64
Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, ArXiv e-prints [arXiv:1110.3193]
Lavaux, G., & Wandelt, B. D. 2012, ApJ, 754, 109
Lewis, A., Challinor, A., & Lasenby, A. 2000, ApJ, 538, 473
Linder, E. V. 2005, Phys. Rev. D, 72, 043529
Marín, F. A., Blake, C., Poole, G. B., et al. 2013, MNRAS, 432, 2654
Massara, E., Percival, W. J., Dalal, N., et al. 2022, MNRAS, 517, 4458
Massara, E., Villaescusa-Navarro, F., Hahn, C., et al. 2023, ApJ, 951, 70
Mohammad, F. G., & Percival, W. J. 2022, MNRAS, 514, 1289
Nadathur, S. 2016, MNRAS, 461, 358
Nadathur, S., & Crittenden, R. 2016, ApJ, 830, L19
Nadathur, S., & Percival, W. J. 2019, MNRAS, 483, 3472
Nadathur, S., Carter, P. M., Percival, W. J., Winther, H. A., & Bautista, J. E. 2019a, Phys. Rev. D, 100, 023504
Nadathur, S., Carter, P., & Percival, W. J. 2019b, MNRAS, 482, 2459
Nadathur, S., Carter, P. M., Percival, W. J., Winther, H. A., & Bautista, J. E. 2019c, Astrophysics Source Code Library [record ascl:1907.023]
Nadathur, S., Woodfinden, A., Percival, W. J., et al. 2020a, MNRAS, 499, 4140
Nadathur, S., Percival, W. J., Beutler, F., & Winther, H. A. 2020b, Phys. Rev. Lett., 124, 221301
Neyrinck, M. C. 2008, MNRAS, 386, 2101
Neyrinck, M. C., Szapudi, I., & Szalay, A. S. 2009, ApJ, 698, L90
Nichol, R. C., Sheth, R. K., Suto, Y., et al. 2006, MNRAS, 368, 1507
Padmanabhan, N., Xu, X., Eisenstein, D. J., et al. 2012, MNRAS, 427, 2132
Paillas, E., Cai, Y.-C., Padilla, N., & Sánchez, A. G. 2021, MNRAS, 505, 5731
Paillas, E., Cuesta-Lazaro, C., Zarrouk, P., et al. 2023, MNRAS, 522, 606
Paz, D. J., Sgro, M. A., Merchan, M., & Padilla, N. 2011, MNRAS, 414, 2029
Paz, D., Lares, M., Ceccarelli, L., Padilla, N., & Lambas, D. G. 2013, MNRAS, 436, 3480
Peebles, P. J. E. 1979, AJ, 84, 730
Peebles, P. J. E. 1980, The Large-scale Structure of the Universe (Princeton University Press)
Philcox, O. H. E., & Ivanov, M. M. 2022, Phys. Rev. D, 105, 043517
Pisani, A., Lavaux, G., Sutter, P. M., & Wandelt, B. D. 2014, MNRAS, 443, 3238
Pisani, A., Sutter, P. M., Hamaus, N., et al. 2015, Phys. Rev. D, 92, 083531
Pisani, A., Massara, E., Spergel, D. N., et al. 2019, BAAS, 51, 40
Planck Collaboration VI. 2020, A&A, 641, A6
Potter, D., Stadel, J., & Teyssier, R. 2017, Comput. Astrophys. Cosmol., 4, 2
Raghunathan, S., Nadathur, S., Sherwin, B. D., & Whitehorn, N. 2020, ApJ, 890, 168
Ryden, B. S. 1995, ApJ, 452, 25
Sánchez, C., Clampitt, J., Kovacs, A., et al. 2017, MNRAS, 465, 746
Satpathy, S., Alam, S., Ho, S., et al. 2017, MNRAS, 469, 1369
Satpathy, S., Croft, R. A. C., Ho, S., & Li, B. 2019, MNRAS, 484, 2148
Schuster, N., Hamaus, N., Dolag, K., & Weller, J. 2023, JCAP, 05, 031
Scoccimarro, R., Feldman, H. A., Fry, J. N., & Frieman, J. A. 2001, ApJ, 546, 652
Seo, H.-J., & Eisenstein, D. J. 2003, ApJ, 598, 720
Sinha, M., & Garrison, L. 2019, in Software Challenges to Exascale Computing, eds. A. Majumdar, & R. Arora (Singapore: Springer Singapore), 3
Sinha, M., & Garrison, L. H. 2020, MNRAS, 491, 3022
Slepian, Z., Eisenstein, D. J., Beutler, F., et al. 2017, MNRAS, 468, 1070
Tallada, P., Carretero, J., Casals, J., et al. 2020, Astron. Comput., 32, 100391
Tinker, J. L. 2007, MNRAS, 374, 477
Torrado, J., & Lewis, A. 2021, JCAP, 2021, 057
Uhlemann, C., Pajer, E., Pichon, C., et al. 2017, MNRAS, 474, 2853
Valogiannis, G., & Dvorkin, C. 2022a, Phys. Rev. D, 106, 103509
Valogiannis, G., & Dvorkin, C. 2022b, Phys. Rev. D, 105, 103534
Verde, L., Heavens, A. F., Percival, W. J., et al. 2002, MNRAS, 335, 432
Wang, Y., Yang, X., Mo, H. J., van den Bosch, F. C., & Chu, Y. 2004, MNRAS, 353, 287
Wang, Y., Zhao, G. B., Koyama, K., et al. 2022, ArXiv e-prints [arXiv:2202.05248]
White, M. 2016, JCAP, 2016, 057
Wolk, M., Carron, J., & Szapudi, I. 2015, MNRAS, 454, 560
Woodfinden, A., Nadathur, S., Percival, W. J., et al. 2022, MNRAS, 516, 4307
Yang, A., & Saslaw, W. C. 2011, ApJ, 729, 123
Zhao, C., Variu, A., He, M., et al. 2022, MNRAS, 511, 5492

[1] Institute of Theoretical Astrophysics, University of Oslo, PO Box 1029 Blindern, 0315 Oslo, Norway
e-mail: sladana.radinovic@astro.uio.no

[2] Institute of Cosmology and Gravitation, University of Portsmouth, Portsmouth PO1 3FX, UK
[3] Centre for Astrophysics, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada
[4] Department of Physics and Astronomy, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada
[5] Perimeter Institute for Theoretical Physics, Waterloo, Ontario N2L 2Y5, Canada
[6] Dipartimento di Fisica e Astronomia "Augusto Righi" – Alma Mater Studiorum Università di Bologna, via Piero Gobetti 93/2, 40129 Bologna, Italy
[7] INFN-Sezione di Bologna, Viale Berti Pichat 6/2, 40127 Bologna, Italy
[8] INAF-Osservatorio di Astrofisica e Scienza dello Spazio di Bologna, Via Piero Gobetti 93/3, 40129 Bologna, Italy
[9] Universitäts-Sternwarte München, Fakultät für Physik, Ludwig-Maximilians-Universität München, Scheinerstrasse 1, 81679 München, Germany
[10] Instituto de Astrofísica de Canarias, Calle Vía Láctea s/n, 38204 San Cristóbal de La Laguna, Tenerife, Spain
[11] Departamento de Astrofísica, Universidad de La Laguna, 38206 La Laguna, Tenerife, Spain
[12] MTA-CSFK Lendület Large-Scale Structure Research Group, Konkoly-Thege Miklós út 15-17, 1121 Budapest, Hungary
[13] Center for Computational Astrophysics, Flatiron Institute, 162 5th Avenue, 10010 New York, NY, USA
[14] The Cooper Union for the Advancement of Science and Art, 41 Cooper Square, New York, NY 10003, USA
[15] Department of Astrophysical Sciences, Peyton Hall, Princeton University, Princeton, NJ 08544, USA
[16] INFN-Padova, Via Marzolo 8, 35131 Padova, Italy
[17] Dipartimento di Fisica e Astronomia "G.Galilei", Universitá di Padova, Via Marzolo 8, 35131 Padova, Italy
[18] Univ. Lyon, Univ. Claude Bernard Lyon 1, CNRS/IN2P3, IP2I Lyon, UMR 5822, 69622 Villeurbanne, France
[19] INAF-Osservatorio Astrofisico di Torino, Via Osservatorio 20, 10025 Pino Torinese (TO), Italy
[20] Dipartimento di Fisica, Università di Genova, Via Dodecaneso 33, 16146 Genova, Italy
[21] INFN-Sezione di Roma Tre, Via della Vasca Navale 84, 00146 Roma, Italy
[22] Department of Physics "E. Pancini", University Federico II, Via Cinthia 6, 80126 Napoli, Italy
[23] Dipartimento di Fisica, Universitá degli Studi di Torino, Via P. Giuria 1, 10125 Torino, Italy
[24] INFN-Sezione di Torino, Via P. Giuria 1, 10125 Torino, Italy
[25] INAF-IASF Milano, Via Alfonso Corti 12, 20133 Milano, Italy
[26] INAF-Osservatorio Astronomico di Roma, Via Frascati 33, 00078 Monteporzio Catone, Italy
[27] INFN-Sezione di Roma, Piazzale Aldo Moro, 2 – c/o Dipartimento di Fisica, Edificio G. Marconi, 00185 Roma, Italy
[28] Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, 08193 Bellaterra (Barcelona), Spain
[29] Port d'Informació Científica, Campus UAB, C. Albareda s/n, 08193 Bellaterra (Barcelona), Spain
[30] INAF-Osservatorio Astronomico di Capodimonte, Via Moiariello 16, 80131 Napoli, Italy
[31] INFN section of Naples, Via Cinthia 6, 80126 Napoli, Italy
[32] Dipartimento di Fisica e Astronomia "Augusto Righi" – Alma Mater Studiorum Universitá di Bologna, Viale Berti Pichat 6/2, 40127 Bologna, Italy
[33] Centre National d'Études Spatiales, 31400 Toulouse, France
[34] Institut national de physique nucléaire et de physique des particules, 3 rue Michel-Ange, 75794 Paris Cedex 16, France
[35] Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK
[36] ESAC/ESA, Camino Bajo del Castillo, s/n., Urb. Villafranca del Castillo, 28692 Villanueva de la Cañada, Madrid, Spain
[37] European Space Agency/ESRIN, Largo Galileo Galilei 1, 00044 Frascati, Roma, Italy
[38] Institute of Physics, Laboratory of Astrophysics, École Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, 1290 Versoix, Switzerland
[39] Departamento de Física, Faculdade de Ciências, Universidade de Lisboa, Edifício C8, Campo Grande, 1749-016 Lisboa, Portugal
[40] Instituto de Astrofísica e Ciências do Espaço, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal
[41] Université Paris-Saclay, CNRS, Institut d'astrophysique spatiale, 91405 Orsay, France
[42] Department of Astronomy, University of Geneva, ch. d'Ecogia 16 , 1290 Versoix, Switzerland
[43] Université Paris-Saclay, Université Paris Cité, CEA, CNRS, Astrophysique, Instrumentation et Modélisation Paris-Saclay, 91191 Gif-sur-Yvette, France
[44] Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, 08193 Barcelona, Spain
[45] Institut d'Estudis Espacials de Catalunya (IEEC), Carrer Gran Capitá 2-4, 08034 Barcelona, Spain
[46] INAF-Osservatorio Astronomico di Trieste, Via G. B. Tiepolo 11, 34143 Trieste, Italy
[47] Aix-Marseille Université, CNRS/IN2P3, CPPM, Marseille, France
[48] INAF-Osservatorio Astronomico di Padova, Via dell'Osservatorio 5, 35122 Padova, Italy
[49] Max Planck Institute for Extraterrestrial Physics, Giessenbachstr. 1, 85748 Garching, Germany
[50] Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA
[51] Technical University of Denmark, Elektrovej 327, 2800 Kgs. Lyngby, Denmark
[52] Cosmic Dawn Center (DAWN), Copenhagen, Denmark
[53] Max-Planck-Institut für Astronomie, Königstuhl 17, 69117 Heidelberg, Germany
[54] Mullard Space Science Laboratory, University College London, Holmbury St Mary, Dorking, Surrey RH5 6NT, UK
[55] Department of Physics, PO Box 64, 00014 University of Helsinki, Finland
[56] Helsinki Institute of Physics, Gustaf Hällströmin katu 2, University of Helsinki, Helsinki, Finland
[57] NOVA optical infrared instrumentation group at ASTRON, Oude Hoogeveensedijk 4, 7991PD Dwingeloo, The Netherlands
[58] Argelander-Institut für Astronomie, Universität Bonn, Auf dem Hügel 71, 53121 Bonn, Germany
[59] Department of Physics, Institute for Computational Cosmology, Durham University, South Road, Durham, DH1 3LE, UK
[60] Université Paris Cité, CNRS, Astroparticule et Cosmologie, 75013 Paris, France
[61] University of Applied Sciences and Arts of Northwestern Switzerland, School of Engineering, 5210 Windisch, Switzerland
[62] Institut d'Astrophysique de Paris, 98bis Boulevard Arago, 75014 Paris, France
[63] Institut d'Astrophysique de Paris, UMR 7095, CNRS, and Sorbonne Université, 98 bis boulevard Arago, 75014 Paris, France
[64] CEA Saclay, DFR/IRFU, Service d'Astrophysique, Bat. 709, 91191 Gif-sur-Yvette, France
[65] European Space Agency/ESTEC, Keplerlaan 1, 2201 AZ Noordwijk, The Netherlands
[66] Leiden Observatory, Leiden University, Niels Bohrweg 2, 2333 CA Leiden, The Netherlands
[67] Kapteyn Astronomical Institute, University of Groningen, PO Box 800, 9700 AV Groningen, The Netherlands
[68] Department of Physics and Astronomy, University of Aarhus, Ny Munkegade 120, 8000 Aarhus C, Denmark
[69] AIM, CEA, CNRS, Université Paris-Saclay, Université de Paris, 91191 Gif-sur-Yvette, France

[70] Space Science Data Center, Italian Space Agency, Via del Politecnico snc, 00133 Roma, Italy
[71] Institute of Space Science, Bucharest 077125, Romania
[72] Departamento de Física, FCFM, Universidad de Chile, Blanco Encalada 2008, Santiago, Chile
[73] Institut de Ciencies de l'Espai (IEEC-CSIC), Campus UAB, Carrer de Can Magrans, s/n Cerdanyola del Vallés, 08193 Barcelona, Spain
[74] Aix-Marseille Université, CNRS, CNES, LAM, 13000 Marseille, France
[75] Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Avenida Complutense 40, 28040 Madrid, Spain
[76] Instituto de Astrofísica e Ciências do Espaço, Faculdade de Ciências, Universidade de Lisboa, Tapada da Ajuda, 1349-018 Lisboa, Portugal
[77] Universidad Politécnica de Cartagena, Departamento de Electrónica y Tecnología de Computadoras, 30202 Cartagena, Spain
[78] Institut de Recherche en Astrophysique et Planétologie (IRAP), Université de Toulouse, CNRS, UPS, CNES, 14 Av. Edouard Belin, 31400 Toulouse, France
[79] Infrared Processing and Analysis Center, California Institute of Technology, Pasadena, CA 91125, USA
[80] Junia, EPA department, 59000 Lille, France

Appendix A: Accounting for correlation between data and theory

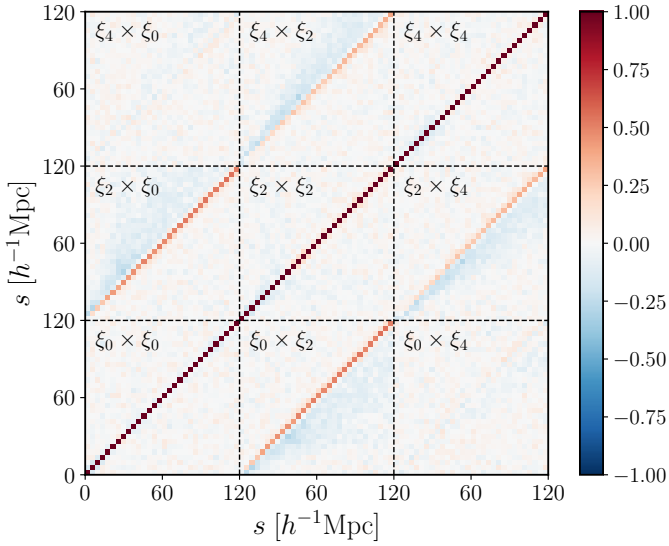

Fig. A.1. Normalised full covariance matrix $\mathsf{C}^{\rm tot}$ from the Flagship $z_{\rm eff} = 1.0$ redshift bin, estimated using jackknife resampling. Dashed lines differentiate between monopole, quadrupole and hexadecapole blocks of the data vector.

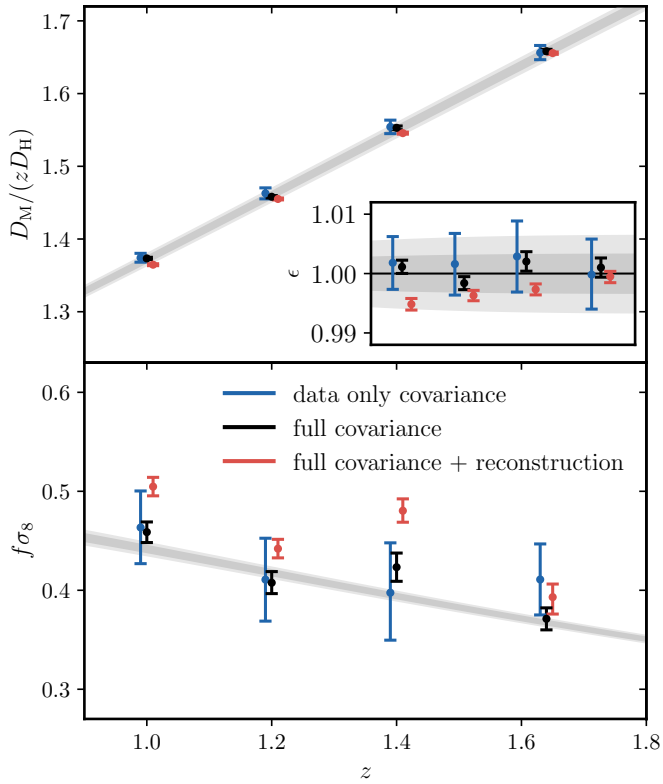

Fig. A.2. Redshift dependence of the measured cosmological parameters obtained from Flagship using different covariance prescriptions. The blue points show results for fits to the data vector for the perfect reconstruction case using $\mathsf{C}^{\rm data}$ and are the same as in Fig. 8. Black points are for fits to the same data vector but using the full covariance $\mathsf{C}^{\rm tot}$. Light red points are for the fit to the realistic reconstruction case when using $\mathsf{C}^{\rm tot}$. The use of the full covariance $\mathsf{C}^{\rm tot}$ greatly reduces the statistical uncertainty but can introduce systematic offsets when combined with the use of reconstruction.

Equation (25) defines the jackknife estimate of the covariance $\mathsf{C}^{\rm data}$ of the redshift-space data vector $\boldsymbol{\xi}^{\rm s}_{\rm data}$, which we have used for computing the likelihood in the main analysis. However, this is not the only source of uncertainty: the computation of the theory model $\boldsymbol{\xi}^{\rm s}_{\rm theory}$ depends on the multipoles of the real-space CCF $\xi^{\rm rr}_\ell$, which are themselves determined from measurements performed on the galaxy catalogue using the estimators described in Sect. 4.2. These estimators have an associated uncertainty, which is negligible when $\xi^{\rm rr}_\ell$ is estimated from the mean of a large number ($N \approx 1000$) of mock realisations as done, for example, by Nadathur et al. (2019b) and Woodfinden et al. (2022), but is significant in our case as the estimate is derived from a single Flagship realisation. Equally, since $\xi^{\rm rr}_\ell$ is estimated from the same Flagship void and galaxy catalogues which are used to measure the data vector, the uncertainties in $\boldsymbol{\xi}^{\rm s}_{\rm theory}$ and $\boldsymbol{\xi}^{\rm s}_{\rm data}$ will in principle be correlated with each other.

In order to self-consistently account for these model uncertainties and correlations with the data vector, we can define a combined effective data vector $\boldsymbol{\xi}^{\rm tot} \equiv \boldsymbol{\xi}^{\rm s}_{\rm theory} - \boldsymbol{\xi}^{\rm s}_{\rm data}$ and estimate the covariance $\mathsf{C}^{\rm tot}$ using the same jackknife procedure described in Sect. 4.3. Then Eq. (26) for the log-likelihood remains unchanged, provided that C^{-1} is understood to refer to the inverse covariance matrix of $\boldsymbol{\xi}^{\rm tot}$ and not $\boldsymbol{\xi}^{\rm s}_{\rm data}$ alone. We can write this schematically as:

$$\mathsf{C}\,(\text{data} - \text{theory}) = \mathsf{C}\,(\text{data}) + \mathsf{C}\,(\text{theory}) - 2\mathsf{C}\,(\text{data}, \text{theory})\ , \tag{A.1}$$

where the last term represents the cross-covariance between the data vector and the theory model. The familiar scenario corresponds to a situation where the second term on the right is negligible and the last term completely vanishes. This is the case by construction in previous studies in the literature (e.g. Nadathur et al. 2020b; Woodfinden et al. 2022), since $\xi^{\rm rr}_\ell$ was taken as the mean from many mocks and thus data were theory are not correlated. For the method used in this work, given the limitations of only a single Flagship mock sample, this is not true. For future *Euclid* data analyses, whether this term is important or not will depend on the choice of how to determine $\xi^{\rm rr}_\ell$.

For Flagship, we found a high degree of correlation between the model and the data vectors, which changes the correlation structure of the full covariance $\mathsf{C}^{\rm tot}$ with respect to $\mathsf{C}^{\rm data}$. This is shown in Fig. A.1, and can be compared to Fig. 5. Two features are immediately noticeable: off-diagonal correlations between the monopole and quadrupole components in particular are enhanced, and within each diagonal block of the covariance matrix, correlations between different radial bins are relatively suppressed. This behaviour arises because uncertainties in the measurement of the real and redshift-space monopoles $\xi^{\rm rr}_0$ and $\xi^{\rm rs}_0$ in each bin are naturally strongly correlated, as the real and redshift-space galaxy positions are closely correlated. In each radial bin, the monopole component $\xi^{\rm rr}_0$ strongly influences the corresponding modelled multipoles of $\boldsymbol{\xi}^{\rm s}_{\rm theory}$, thus creating the correlation between the elements of $\boldsymbol{\xi}^{\rm s}_{\rm theory}$ and the monopole component of $\boldsymbol{\xi}^{\rm s}_{\rm data}$.

Even more importantly, the strong correlation between real- and redshift-space multipole moments causes the amplitude of the diagonal elements of $\mathsf{C}^{\rm tot}$ to be significantly smaller than those of $\mathsf{C}^{\rm data}$. This is shown in Fig. A.3. At first glance, it may

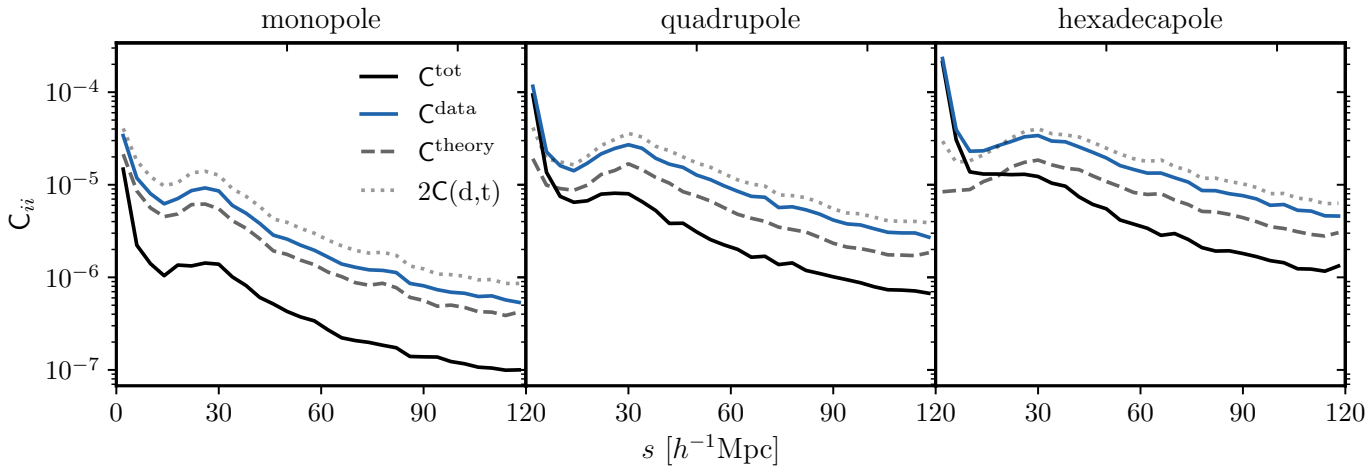

Fig. A.3. Diagonal elements of the full covariance matrix $\mathsf{C}^{\rm tot}$ (black lines), as well as the three different contributions from Eq. (A.1), showing the reduced variance that results from properly accounting for correlation between data and theory. The three different panels represent the monopole, quadrupole and hexadecapole part, respectively.

seem unusual that accounting for additional uncertainty in the model effectively reduces the total uncertainty $\mathsf{C}^{\rm tot}$, but this is simply a reflection of the correlation that means that variations in $\boldsymbol{\xi}^{\rm s}_{\rm theory} - \boldsymbol{\xi}^{\rm s}_{\rm data}$ are reduced since uncertainties move both components in the same direction. It is intuitive that when $\boldsymbol{\xi}^{\rm s}_{\rm theory}$ and $\boldsymbol{\xi}^{\rm s}_{\rm data}$ are highly correlated with each other, models that deviate from observation should be more severely penalised in the log-likelihood, and this is achieved by the smaller values of $\mathsf{C}^{\rm tot}$. An equivalent way to view this is that when $\boldsymbol{\xi}^{\rm s}_{\rm theory}$ and $\boldsymbol{\xi}^{\rm s}_{\rm data}$ are both obtained from the same realisation of the initial conditions, the cosmic variance in them cancels out.

Changing the covariance matrix used in the likelihood evaluation in this way naturally propagates through to the recovered statistical uncertainty on the cosmological parameters. To estimate this, we performed all the model fits again using the appropriate full covariance $\mathsf{C}^{\rm tot}$ instead of $\mathsf{C}^{\rm data}$. The results obtained for $D_{\rm M}/D_{\rm H}$ and $f\sigma_8$ are shown in Fig. A.2, where the black points are for the perfect reconstruction scenario fit using $\mathsf{C}^{\rm tot}$, and the blue points are for fits to the same data using $\mathsf{C}^{\rm data}$. It is clear from this that correctly accounting for the correlation between model and data leads to a very significant increase in the statistical precision of the fit, roughly by a factor of 3 in both $D_{\rm M}/D_{\rm H}$ and $f\sigma_8$. The main results of our paper, which do not account for this effect, are therefore conservative overestimates of the statistical error that can be achieved with *Euclid* if the real-space cross-correlation function is measured from the data instead of being taken from a simulation.

However, the reduction in statistical uncertainty makes systematic errors more important. While in the perfect reconstruction scenario we are able to recover unbiased estimates of $D_{\rm M}/D_{\rm H}$ and $f\sigma_8$ at the higher precision, this is not so for the realistic reconstruction scenario (red points in Fig. A.2), where systematic offsets from the fiducial values are apparent. The source of these offsets is likely to be the residuals in the modelling of the monopole moment at small separations, an example of which can be seen in Fig. 6. These are caused by the imperfections of the practical reconstruction technique. Figure A.3 shows that the relative reduction in covariance terms from using $\mathsf{C}^{\rm tot}$ instead of $\mathsf{C}^{\rm data}$ is largest for the monopole, so these residuals have a larger effect on the recovered fit. For the realistic reconstruction scenario, the full covariance matrix $\mathsf{C}^{\rm tot}$ therefore does not give a good estimate of the total error budget, and systematic uncertainties would have to be folded in.

As we only have one Flagship simulation available, we cannot quantitatively assess the true size of these systematic uncertainties. However, it appears that the apparent large improvement in statistical precision would not be practically realised for *Euclid* in the reconstruction scenarios, even if $\xi^{\rm rr}_{\ell}$ were estimated from the data and the full covariance $\mathsf{C}^{\rm tot}$ were used. Therefore the pessimistic forecasts we have obtained for the existing methodology using $\mathsf{C}^{\rm data}$ alone are more realistic.