

Fermilab

Characterization of DESI fiber assignment incompleteness effect on
2-point clustering and mitigation methods for DR1 analysis

FERMILAB-PUB-24-0849-PPD

arXiv:2411.12025

This manuscript has been authored by Fermi Research Alliance, LLC
under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy,
Office of Science, Office of High Energy Physics.

Characterization of DESI fiber assignment incompleteness effect on 2-point clustering and mitigation methods for DR1 analysis

D. Bianchi¹, M. M. S Hanif², A. Carnero Rosell^{3,4},
J. Lasker^{5,6}, A. J. Ross^{7,8,9}, M. Pinon¹⁰, A. de Mattia¹⁰,
M. White^{11,12}, S. Ahlen¹³, S. Bailey¹⁴, D. Brooks¹⁵,
E. Burtin¹⁰, E. Chaussidon¹⁴, T. Claybaugh¹⁴, S. Cole¹⁶, A. de la
Macorra¹⁷, S. Ferraro^{14,12}, A. Font-Ribera^{15,18},
J. E. Forero-Romero^{19,20}, E. Gaztañaga^{21,22,23}, S. Gontcho A
Gontcho¹⁴, G. Gutierrez²⁴, J. Guy¹⁴, C. Hahn²⁵,
K. Honscheid^{7,26,9}, C. Howlett²⁷, S. Juneau²⁸, D. Kirkby²⁹,
T. Kisner¹⁴, A. Kremin¹⁴, M. Landriau¹⁴, L. Le Guillou³⁰,
M. E. Levi¹⁴, P. McDonald¹⁴, A. Meisner²⁸, R. Miquel^{31,18},
J. Moustakas³², N. Palanque-Delabrouille^{10,14},
W. J. Percival^{33,34,35}, F. Prada³⁶, I. Pérez-Ràfols³⁷,
A. Raichoor¹⁴, G. Rossi³⁸, E. Sanchez³⁹, D. Schlegel¹⁴,
M. Schubnell^{40,2}, R. Sharples^{41,16}, J. Silber¹⁴, D. Sprayberry²⁸,
G. Tarlé², M. Vargas-Magaña¹⁷, B. A. Weaver²⁸, P. Zarrouk³⁰,
R. Zhou¹⁴, H. Zou⁴²

¹Dipartimento di Fisica “Aldo Pontremoli”, Università degli Studi di Milano, Via Celoria 16, I-20133 Milano, Italy

²University of Michigan, Ann Arbor, MI 48109, USA

³Departamento de Astrofísica, Universidad de La Laguna (ULL), E-38206, La Laguna, Tenerife, Spain

⁴Instituto de Astrofísica de Canarias, C/ Vía Láctea, s/n, E-38205 La Laguna, Tenerife, Spain

⁵Department of Physics, Southern Methodist University, 3215 Daniel Avenue, Dallas, TX 75275, USA

⁶Schmidt Sciences, 155 W 23rd St, New York, NY 10011, USA

⁷Center for Cosmology and AstroParticle Physics, The Ohio State University, 191 West Woodruff Avenue, Columbus, OH 43210, USA

- ⁸Department of Astronomy, The Ohio State University, 4055 McPherson Laboratory, 140 W 18th Avenue, Columbus, OH 43210, USA
- ⁹The Ohio State University, Columbus, 43210 OH, USA
- ¹⁰IRFU, CEA, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France
- ¹¹Department of Physics, University of California, Berkeley, 366 LeConte Hall MC 7300, Berkeley, CA 94720-7300, USA
- ¹²University of California, Berkeley, 110 Sproul Hall #5800 Berkeley, CA 94720, USA
- ¹³Physics Dept., Boston University, 590 Commonwealth Avenue, Boston, MA 02215, USA
- ¹⁴Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA
- ¹⁵Department of Physics & Astronomy, University College London, Gower Street, London, WC1E 6BT, UK
- ¹⁶Institute for Computational Cosmology, Department of Physics, Durham University, South Road, Durham DH1 3LE, UK
- ¹⁷Instituto de Física, Universidad Nacional Autónoma de México, Cd. de México C.P. 04510, México
- ¹⁸Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, 08193 Bellaterra Barcelona, Spain
- ¹⁹Departamento de Física, Universidad de los Andes, Cra. 1 No. 18A-10, Edificio Ip, CP 111711, Bogotá, Colombia
- ²⁰Observatorio Astronómico, Universidad de los Andes, Cra. 1 No. 18A-10, Edificio H, CP 111711 Bogotá, Colombia
- ²¹Institut d'Estudis Espacials de Catalunya (IEEC), 08034 Barcelona, Spain
- ²²Institute of Cosmology and Gravitation, University of Portsmouth, Dennis Sciama Building, Portsmouth, PO1 3FX, UK
- ²³Institute of Space Sciences, ICE-CSIC, Campus UAB, Carrer de Can Magrans s/n, 08913 Bellaterra, Barcelona, Spain
- ²⁴Fermi National Accelerator Laboratory, PO Box 500, Batavia, IL 60510, USA
- ²⁵Department of Astrophysical Sciences, Princeton University, Princeton NJ 08544, USA
- ²⁶Department of Physics, The Ohio State University, 191 West Woodruff Avenue, Columbus, OH 43210, USA
- ²⁷School of Mathematics and Physics, University of Queensland, 4072, Australia
- ²⁸NSF NOIRLab, 950 N. Cherry Ave., Tucson, AZ 85719, USA
- ²⁹Department of Physics and Astronomy, University of California, Irvine, 92697, USA
- ³⁰Sorbonne Université, CNRS/IN2P3, Laboratoire de Physique Nucléaire et de Hautes Energies (LPNHE), FR-75005 Paris, France
- ³¹Institució Catalana de Recerca i Estudis Avançats, Passeig de Lluís Companys, 23, 08010 Barcelona, Spain
- ³²Department of Physics and Astronomy, Siena College, 515 Loudon Road, Loudonville, NY 12211, USA
- ³³Department of Physics and Astronomy, University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1, Canada
- ³⁴Perimeter Institute for Theoretical Physics, 31 Caroline St. North, Waterloo, ON N2L 2Y5, Canada
- ³⁵Waterloo Centre for Astrophysics, University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1, Canada

³⁶Instituto de Astrofísica de Andalucía (CSIC), Glorieta de la Astronomía, s/n, E-18008 Granada, Spain

³⁷Departament de Física, EEBE, Universitat Politècnica de Catalunya, c/Eduard Maristany 10, 08930 Barcelona, Spain

³⁸Department of Physics and Astronomy, Sejong University, Seoul, 143-747, Korea

³⁹CIEMAT, Avenida Complutense 40, E-28040 Madrid, Spain

⁴⁰Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA

⁴¹Centre for Advanced Instrumentation, Department of Physics, Durham University, South Road, Durham DH1 3LE, UK

⁴²National Astronomical Observatories, Chinese Academy of Sciences, A20 Datun Rd., Chaoyang District, Beijing, 100012, P.R. China

E-mail: davide.bianchi1@unimi.it

Abstract. We present an in-depth analysis of the fiber assignment incompleteness in the Dark Energy Spectroscopic Instrument (DESI) Data Release 1 (DR1). This incompleteness is caused by the restricted mobility of the robotic fiber positioner in the DESI focal plane, which limits the number of galaxies that can be observed at the same time, especially at small angular separations. As a result, the observed clustering amplitude is suppressed in a scale-dependent manner, which, if not addressed, can severely impact the inference of cosmological parameters. We discuss the methods adopted for simulating fiber assignment on mocks and data. In particular, we introduce the fast fiber assignment (FFA) emulator, which was employed to obtain the power spectrum covariance adopted for the DR1 full-shape analysis. We present the mitigation techniques, organised in two classes: measurement stage and model stage. We then use high fidelity mocks as a reference to quantify both the accuracy of the FFA emulator and the effectiveness of the different measurement-stage mitigation techniques. This complements the studies conducted in a parallel paper for the model-stage techniques, namely the θ -cut approach. We find that pairwise inverse probability (PIP) weights with angular upweighting recover the “true” clustering in all the cases considered, in both Fourier and configuration space. Notably, we present the first ever power spectrum measurement with PIP weights from real data.

ArXiv ePrint: [xxxx.xxxxx](https://arxiv.org/abs/xxxx.xxxxx)

Contents

1	Introduction	1
2	Data	4
3	Simulating fiber assignment	7
3.1	Alternative merged target ledger (altMTL)	7
3.2	Fast fiber assignment (FFA)	8
4	Simulated catalogues	10
4.1	Simulated DESI DR1 target samples	11
4.2	Comparing altMTL and FFA mock catalogues	11
5	Estimator-level mitigation techniques	15
5.1	Completeness weights	16
5.2	Inverse probability weights	17
5.2.1	Inverse probability with angular upweighting	18
5.2.2	Counteracting the effect of zero-probability galaxies: w_{NTMP}	20
6	Model-level mitigation techniques	20
6.1	Removing small angular separations (θ -cut)	21
7	Comparison of results	22
8	Conclusions	31
A	Adding small-scale anticorrelation to the FFA catalogues	33
B	Collision kernel from the mocks as a mitigation strategy	35
C	Clustering measurements averaged over multiple FFA mocks	36

1 Introduction

The three-dimensional distribution of galaxies encodes a wealth of cosmological information about the evolution of our Universe and the nature of its constituents. Over the past twenty-five years, large spectroscopic surveys have been essential in extracting this information by gathering millions of galaxy redshifts across increasingly larger volumes [1–3].

The Dark Energy Spectroscopic Instrument (DESI) survey [4] relies on multiplexed spectroscopic observations achieved through a multi-object spectrograph [5]. 5020 robotic positioners move optical fibers into place on the focal plane [6–8] and allow 5000 spectra¹ to be measured at once [9]. DESI thus achieves a multiplexing that is approximately five times greater than its predecessor, the Sloan Digital Sky Survey (SDSS). The improved multiplexing, combined with the ability to dynamically adapt exposure times and tiling strategy to the observing conditions [10] and an approximately $2.5\times$ increase in mirror area,

¹20 of the fibers send their light to a camera that monitors sky conditions.

enabled DESI to obtain more than four times as many extragalactic redshifts in its first year of main survey operation compared to what SDSS collected during its entire twenty-year history.

However, to fully benefit of such an increase in statistical power, it is essential to have precise control over all the potential sources of systematic error. In other words, is it crucial to identify and remove or model all the spurious fluctuations in the galaxy density that stem from non-cosmological factors. Like for eBOSS and SDSS, in DESI one significant source of these clustering artifacts comes from missing observations, particularly those due to fiber assignment (FA) issues.

There are two main instrumental limitations at play. First, each of the 5000 positioners is limited to a “patrol radius” which covers only a small fraction of the focal plane. Although each patrol region overlaps with those of adjacent fibers (approximately 15% of the total focal plane area is reachable by two fibers), in densely populated regions of the sky the total number of galaxies may surpass the number of available fibers. Second, due the physical size of the positioner, it is impossible to place two optical fibers closer than a minimum angular separation, on the focal plane, which prevents the observation of closely spaced galaxy pairs with a single pass of the instrument. For historical reasons, in the following we will often refer to the combination of these effects as *fiber collisions*, even though, as just explained, it would be more accurate to describe it as a competition (between galaxies) for a fiber.

The overall effect of fiber collisions is a complex pattern of missing observations unevenly distributed across the survey area, with a primary concentration in high-density regions. This results in a non trivial, scale dependent reduction in the amplitude of the observed galaxy clustering, which, if not properly corrected, can significantly affect the cosmological constraints.

Through the years different countermeasures were put forward, which can be roughly grouped in two classes depending on whether the proposed correction is limited to the clustering measurements or rather it requires additional manipulation of the theory model:

1. Estimator level, e.g. [11–17]
2. Model level, e.g. [15, 18–20]

A well known example of an estimator-level approach is angular upweighting [11]. This technique involves weighting each pair, when estimating the three-dimensional correlation function, by the ratio of total pairs to observed pairs at that angular separation. This correction alone is not sufficient in general because it assumes that the radial properties of the unobserved pairs are equivalent to those of the observed pairs. Typically, an imperfect matching between observed and missing galaxies is the reason why many popular missing observations countermeasures fail to achieve the desired level of accuracy. For example, this is why the method proposed by [13] and the first of the two methods proposed by [15], do not deliver unbiased results. Both approaches utilize the regions of overlapping observations to gain insight into the missing data. The first method uses this information to adjust the pair counts, while the second probabilistically assigns redshifts to the missing galaxies.

The standard approach used by the BOSS team [21] is commonly known as the nearest neighbor method. It works by assigning the weight of the missing galaxies to nearest galaxy in angular separation that obtained a redshift. This upweighting approximately corrects the total pair counts on large separations but fails at smaller scales, as the pair between the missed and nearest galaxy is lost. More recently, the eBOSS collaboration improved on this

method by introducing completeness weights [22], which more accurately capture the nature of fiber incompleteness. However, they still struggle to recover the “true” clustering on small scales. A similar approach has been implemented for DESI and is discussed in detail below.

A formally unbiased estimator-level approach was introduced by [16, 23] and further refined in [17]. These studies demonstrated that the “true” clustering can consistently be recovered using Pairwise Inverse Probability (PIP) weights. As the name implies, this method involves weighting each galaxy pair by its inverse probability of being observed. This ensures that the expectation value of both pair and individual galaxy counts aligns with their true values on all scales, as long as there is some area of the survey footprint which is observed multiple times. In addition to real data applications, e.g. [24, 25], the inverse probability approach was tested in preliminary DESI studies with approximate synthetic catalogues [26, 27]. Its application to DESI data is explored in detail throughout this paper.

Some model-level approaches focus on eliminating the (angular) modes impacted by fiber assignment [19, 20] and adjusting the model accordingly. However, this results in a significant loss of large-scale information. Another popular model-level approach is the second method proposed in [15]. The small-scale effect of fiber collisions is described as a top-hat window in transverse separation, which is convolved with the theoretical power spectrum when performing parameter inference. The main limitation of this approach is the approximate description of the window.

A more traditional model-level approach consist of simply removing the small transverse separations or, similarly, the small angles between line of sight and separation, and adjusting the theory model accordingly. Typically, this method have been used in configuration space, with relatively minor variations, e.g. [18, 28, 29]. As extensively discussed in [30] and summarized here in a dedicated section, this approach can be applied directly to the separation angle between galaxies, which more accurately capture the nature of the DESI fiber assignment, and can also be extended to Fourier space. This is the strategy adopted for the DESI DR1 full-shape analysis [31, 32].

In this paper, we conduct a comprehensive study of the impact of fiber-assignment incompleteness on 2-point statistics for DESI data release 1 (DR1) and we present the countermeasures put in place to address this issue, in both configuration and Fourier space.

To effectively study the impact of fiber collisions, it is essential to create realistic synthetic galaxy catalogs (“mocks”) that include this effect. We discuss the strategies we employed to simulate the fiber-assignment process for the DESI-DR1 mock catalogs and to derive PIP weights. In particular, we present the fast fiber assignment (FFA) emulator used to obtain the covariance matrix for the DR1 full shape analysis of the power spectrum [31, 32].

The paper is structured as follows. In Section 2, we present the DESI data, along with an overview of the fundamental constituents of the fiber assignment algorithm and the effects of fiber collisions for DR1. In section 3 we illustrate the two methodologies used to simulate the fiber assignment process on both real data and mock catalogs. In section 4 we describe the mock catalogues and use them to validate the above assignment strategies. In section 5 and 6 we introduce the different estimator-level and model-level techniques we used for fiber mitigation, respectively. In section 7 we present configuration and Fourier space measurements of the 2-point statistics, for both mocks and data, obtained with the different mitigation techniques. We conclude in section 8.

2 Data

DESI is a multi-object spectroscopic instrument [5]: robotic positioners guide 5000 optical fibers [33, 34] to the celestial coordinates of the designated targets [35], with the fibers transmitting the collected light to the spectrographs. The full coverage of the survey footprint is achieved through multiple, overlapping, pointings of the instrument, over different regions of the sky. We refer to such pointings or, equivalently, to the corresponding sets of 5000 targets, as tiles. In this work we study the first, main survey, DESI data release, DR1 [36], which includes the observations completed between May 14, 2021 and June 14, 2022. Specifically, we focus on the redshift catalogs assembled using the version named “iron” of the spectroscopic processing [37].

DESI large-scale observations are organized in two separate programs: “bright” and “dark” time, corresponding, for DR1, to 2275 and 2744 observed tiles, respectively. In dark time DESI observes three classes of targets: Emission Line Galaxies (ELG) [38], Luminous Red Galaxies (LRG) [39] and Quasars (QSO) [40]. In addition, a low redshift Bright Galaxy Sample (BGS) [41] is observed in bright time. All the targets were selected based on photometry from the Legacy Survey Data Release 9 [42, 43]. For cosmological inference, the different target types are organized in redshift bins as follows,

- ELG: $0.8 < z < 1.1$; $1.1 < z < 1.6$
- LRG: $0.4 < z < 0.6$; $0.6 < z < 0.8$; $0.8 < z < 1.1$
- QSO: $0.8 < z < 2.1$; $2.1 < z < 3.5$
- BGS: $0.1 < z < 0.4$

where the highest QSO bin is used for Lyman alpha forest analysis [44] but not for galaxy clustering studies, except for those on primordial non Gaussianity, which adopt $0.8 < z < 3.1$ [45]. In this work, for the sake of compactness, we instead collapse each tracer into a unique redshift bin (we only consider QSOs with $z < 2.1$) and we focus on the dark-program tracers: ELG, LRG and QSO. In principle, all the studies and methods presented here can be applied to BGS tracers as well, as they are equipped with the same tools, such as weights and simulated catalogues, and exhibit a similar response to FA. The raw, unweighted, redshift distributions of the observed galaxies are shown in figure 1.

We refer the reader to [46], and references therein, for an exhaustive description of all the properties of the different catalogues, the methodologies adopted to construct them and the corresponding nomenclature/conventions. Here we focus on those aspects of the catalogues that are particularly relevant for fiber-assignment related phenomena.

The data can be organized into classes of catalogues, each playing a specific role in the FA process. The mitigation methods we will present aim to recover the clustering of all galaxies, of a given tracer type, that had the opportunity to be targeted. This is a binary true-or-false attribute that primarily depends on whether a galaxy entered the patrol radius of at least one fiber, but also takes into account additional effects, e.g. hardware failures, as described in [46]. We collectively refer to those galaxies as the *complete* sample, not to be confused with the parent sample from which they are extracted. For DESI data the parent sample is simply given by the entire photometric catalogue. Crucially, any difference between the two samples is not driven by the clustering itself, since, unlike the assignment probability, the chance of being targeted is independent of the local density, at least in

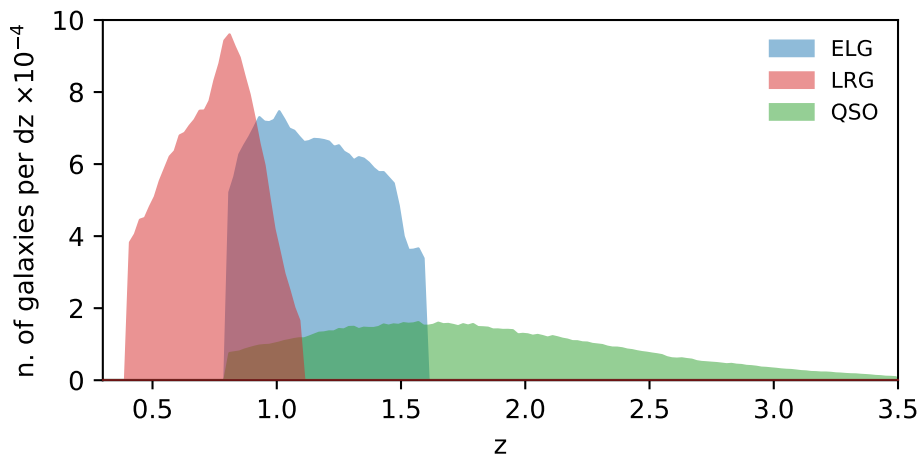


Figure 1. Redshift distribution of the different DESI dark-time tracers, ELGs (blue), LRGs (red) and QSOs (green), with $dz = 0.02$. For illustrative purposes, the figure includes the QSO high-redshift tail up to $z = 3.5$. However, note that objects with $z > 2.1$ are not considered for our clustering measurements.

principle. It thus can be handled with a proper random sample, tracing the effective volume of the complete sample. We define the set of galaxies that were actually targeted and had their redshifts successfully measured as the *clustering* sample. Lastly, we define the *full* sample, primarily used for angular upweighting, as the set of all galaxies of a given tracer type in the parent sample that meet the selection criteria (i.e. the veto mask, see [46]) of the clustering sample, regardless of whether they were observed or not. For consistency, the complete and clustering samples must be strictly defined within the same redshift bins, whereas the full sample does not include redshifts at all (it has by construction the same z -range of the parent sample). Clearly, for real data, the complete sample is not available; it can only be accessed in simulated catalogs. Each of these samples is paired with the correspondent random catalogue of unclustered tracers covering the same volume.

The allocation of fibers to designated targets is governed by the DESI targeting algorithm, `fiberassign`. This code requires the merged-target-ledger (MTL) file as one of its inputs. MTLs are lists of potential DESI targets, each with its own unique identifier (TARGETID), sky coordinates (RA and Dec), PRIORITY, SUBPRIORITY and a survey/observing conditions dependent bitmask (see [47] for a detailed description).

When two or more targets are competing for the same fiber the conflict is settled by first looking at the PRIORITY value, which is deterministic and solely depends on target type. Specifically, for the dark time survey, $\text{PRIORITY}_{\text{QSO}} > \text{PRIORITY}_{\text{LRG}} > \text{PRIORITY}_{\text{ELG}}$, except for a 10% subsample of ELGs (ELG_HIP class), whose PRIORITY have been raised to match that of LRGs as a way to facilitate the extraction of the LRG-ELG cross-correlation signal. If the conflicting targets share the same PRIORITY, then the assignment is determined by comparing their SUBPRIORITY entry, which is simply a list of predetermined random numbers.

In figure 2 we show the angular assignment completeness for the three dark-time tracers. Unless otherwise stated, in this work we examine the North Galactic Cap (NGC) and the South Galactic Cap (SGC) separately. These two regions are defined by their Galactic coordinates, not to be confused with the equatorial coordinates used for the figure. More

explicitly, NGC and SGC correspond to the two disconnected areas, from left to right, separated by the galactic plane (not shown in the figure). One clear artefact created by fiber

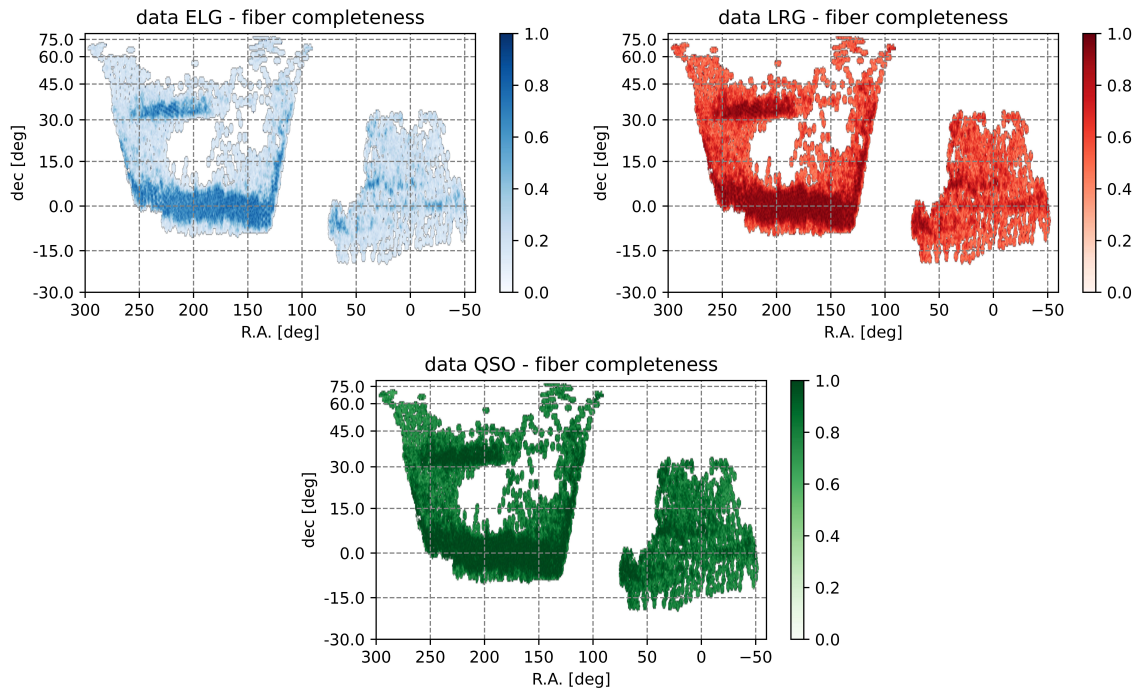


Figure 2. Angular assignment completeness of the galaxies observed by DESI in its first year (DR1), defined as the number of tracers assigned to a fiber divided by the total number of tracers in the corresponding full sample. All three dark-time tracers, ELGs, LRGs and QSOs are shown, as labeled in the figure.

assignment immediately stands out: uneven large-scale distribution of galaxies inside the survey area. These fluctuations are not of cosmological origin but rather they are artificially produced by the observing strategy. In particular, as shown in [46], they correlate almost perfectly with the number of tile overlaps at each point in the survey footprint. We refer to this number, which plays a crucial role for both fiber mitigation and emulation, as n_{tile} .

A second, even more important, artefact is not directly visible in the figure, as it manifests itself on much smaller scales and goes beyond 1-point statistics: fiber collisions. As discussed in section 1, for DESI this effect extends beyond mere physical collisions. In essence, in any area small enough to have a uniquely defined n_{tile} value, we cannot observe more targets than n_{tile} times the number of fibers in that area.

Figure 3 illustrates the impact of fiber collisions on 2-point measurements, as a function of the separation angle θ . The solid pink curve represents the ratio of the pair counts from the sample of successfully observed galaxies in DR1 to the counts from the full sample of potential targets². In order to counteract the effect of fiber assignment, the observed galaxies are weighted with completeness weights³, as detailed in [46] and further discussed below in section 5, which explains why the ratio nicely converges to 1 at large angles. However, for separations $\lesssim 0.02$ deg there is a significant loss of power, up to 60% for the ELGs, that clearly cannot be handled by the weights. Due to its lower completeness (figure 2), the effect

²The pair counts are divided by the total number of pairs in the respective sample.

³Here by completeness weights we mean $w_{\text{comp}}/f_{\text{tile}}$, as described in section 5

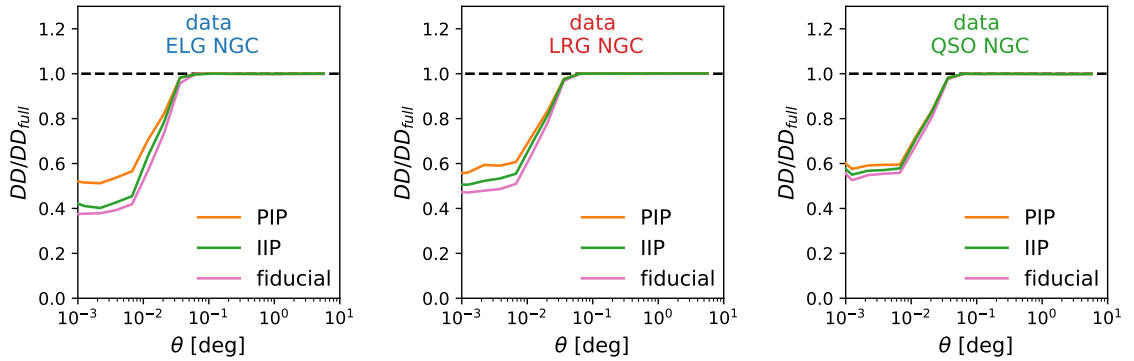


Figure 3. Normalised pair counts measured from DESI DR1, as a function of the separation angle θ , with fiducial weights (solid pink), IIP weights (solid green) and PIP weights (solid orange), divided by the normalised pair counts from the full sample of potential targets. All three dark-time tracers are shown, as labeled in the figure.

gets even stronger for the south galactic cap, not shown here for compactness, reaching about 80%. In the following we will often refer to this small-scale suppression as *collision window*.

The angular upweighting method introduced above and detailed in section 5 restores the “true” angular clustering by assigning to each galaxy pair a weight equal to the inverse of the curve shown in the figure. However, this alone does not guarantee that the corresponding 3-dim clustering is accurately recovered.

3 Simulating fiber assignment

Being able to understand and accurately reproduce the process of fiber assignment is of fundamental importance for, at least, two reasons. First, it allows us to build realistic mock catalogues for covariance matrix estimations and validation of all the different inference models and estimators applied to the data, including the fiber mitigation techniques discussed in this work. Second, by building multiple realisations of the targeting, for both data and mocks, we can directly study the statistical properties of the fiber assignment process itself and use them to construct tailored countermeasures, as the inverse-probability weights. In particular, we can associate to each galaxy a bitwise array, with as many bits as the number of realisations, where ones and zeros indicate whether the galaxy has been assigned or not in a given realisation. These *bitweights* [16], usually compressed into integer numbers⁴, are the building blocks that enable us to efficiently evaluate the assignment probability of any t -plet of galaxies in our samples (see section 5).

We developed two strategies for simulating fiber assignment: one follows the same FA processing as the data to create each replica and therefore it is computationally intensive, while the other relies on a simplified statistical representation of the process, enabling the generation of a much larger number of approximate replicas.

3.1 Alternative merged target ledger (altMTL)

As described in section 2, the allocation of fibers relies on MTLs. Any object in the MTL file gets a SUBPRIORITY entry, i.e. a random value that controls the assignment when the object

⁴The standard format adopted for DESI is 64 bit signed integers.

is competing against other targets with the same `PRIORITY`. The actual fiber assignment process can therefore be replicated, on data or mocks, by generating realistic alternative merged target ledgers (altMTLs). Since the only random component is given by the `SUBPRIORITY` values, the assignment probability of the targets (and of the t -plets they form) can be estimated by rerunning `fiberassign` multiple times, each with a different collection of `SUPRIORITY` values.

A detailed description of the process we have adopted to construct the altMTLs is provided in [47]. For what concerns the clustering-analysis tools, altMTLs allowed us to obtain bitweights for the DR1 samples and to build a set of realistic fiber-assigned mocks (see section 4), with the corresponding bitweights. Since these mocks have been processed through the actual targeting algorithm, the synthetic galaxies are delivered with the same entries as the real ones, including, e.g., completeness weights, w_{comp} , that we will explicitly define in section 5. The downside is that the procedure is time consuming, limiting the number of samples we can create.

3.2 Fast fiber assignment (FFA)

As a way around the cpu-time limitations, we developed an emulator, which enables us to create thousands of fiber assigned samples, with an arbitrarily large number of targeting realisations each. We generically denote as emulator any surrogate algorithm that learns from a set of input-output pairs obtained with a reference algorithm. Using machine-learning jargon, the inputs are the training sample and the outputs the corresponding target labels. Specifically, our reference algorithm is the `fiberassign` code. The training sample can be either a set of accurate mocks or, notably, actual data. The target label is simply the property of having been assigned or not, which each galaxy obtain through `fiberassign`. As regards the architecture of the emulator, since we have a good understanding of the FA-algorithm average behaviour and a clear idea of which key features we need to emulate, we decided to adopt a shallow learning approach. The advantage of our architecture, compared to approaches like deep learning with neural networks, is that it involves far fewer learnable parameters, all of which are easily interpretable, and is less sensitive to potential limitations of the training set, such as the number of samples available. In brief, the emulator works as follows.

- As an input we need a catalogue of all the targets that had a chance of being assigned and a value of n_{tile} for each of them. For the data this informations are obtained by concatenating the `fiberassign` output files as the survey progresses and the code get repeatedly applied to the parent sample of all the DESI targets. For the synthetic data a mock parent sample is constructed with the same volume and number density as the real data and the mock potential assignments are obtained similarly as for the data. This concatenation of potential assignments provides all of the tiles on which a given target could have been assigned and n_{tile} can thus be determined for every target.
- We run a friend-of-friend (FOF) algorithm, based on angular separation, on this sample of potential targets. The linking length, which is actually a linking angle, θ_{LL} , is one of our free parameters. θ_{LL} can be roughly interpreted as the scale below which the assignment probabilities are not independent. In other words, below such scale, the assignment probability of a pair of galaxies is not the product of the two individual probabilities. See sections 5 and 6 for further discussions on the topic. In essence, this step serves to isolate groups of assignment-entangled galaxies.

- Simultaneously, for each galaxy we count the number of companions within angular separation $\theta < \theta_{LL}$. We name this process close-friend count (CFC) and indicate the corresponding variable with n_{CFC} . Repetitions are allowed, i.e. a given galaxy can contribute to the CFC of more than one galaxy. In essence, each galaxy’s n_{CFC} roughly tells us how many other galaxies are competing with it for the same fiber.
- By comparing the pre- and post-assignment samples we obtain a two dimensional function, \mathcal{F} , which maps n_{CFC} and n_{tile} onto the fraction of assigned galaxies. Specifically, we count the number of galaxies with a given combination of n_{CFC} and n_{tile} , in the sample of potential targets, and how many of them got assigned by the reference algorithm. We define $\mathcal{F}(n_{CFC}, n_{tile})$ as the ratio of such counts. By construction, \mathcal{F} is an estimate of the (individual) assignment probability of a galaxy with a given n_{CFC} and n_{tile} .

Once we learned the \mathcal{F} kernel, we can use it to emulate fiber assignment on any test sample of interest.

- We process the test sample through the same n_{tile} finder algorithm used for the training sample: each galaxy in the test sample gets its own n_{tile} .
- We also apply the same FOF-CFC finder algorithm: each galaxy in the test sample gets its own n_{CFC} and a link to its entangled partners.
- We decide how many realisations, N_{bit} , of the targetting we want to generate. Realistically, N_{bit} should be at least $\mathcal{O}(100)$ in order to have enough resolution for the anticorrelation post-processing stage, see below. We randomly sample N_{bit} times each galaxy in the test sample, with its own individual probability, given by $\mathcal{F}(n_{CFC}, n_{tile})$.
- We use the FOF information to add the desired anticorrelation on small scale, as described below.

The relation between anticorrelation and FOFs can be understood as follows. Two galaxies compete for the same fiber only if they are both inside the patrol radius of that fiber, thereby establishing a characteristic exclusion scale (see also sections 5 and 6 for further considerations). However, since the patrol radii overlap, it is formally possible to create chains of entangled galaxies that extend over scales larger than the exclusion one. By construction, these larger entangled structures correspond to, or, more precisely, are contained in the (angular) FOF catalogue obtained using the exclusion scale as a linking length. In other words, the anticorrelation induced by fiber assignment cannot travel distances larger than the size of the corresponding FOF group and, as a consequence, each group can be processed individually.

The anticorrelation algorithm works as follows. For each galaxy inside the group we have N_{bits} realisations of the assignment, conveniently encoded as individual bitwise arrays, with the obvious correspondence between zeros/ones and discarded/assigned. By construction, any operation that keeps the number of zeros and ones in each single array conserved does not affect the individual assignment properties, but it can change the collective properties, specifically the assignment correlation. With this in mind, we developed an algorithm that permutes ones and zeros, in each single bitwise array, in such a way that the anticorrelation of the pairs inside a FOF group gets maximized. Details on how we build this maximum-anticorrelation state can be found in appendix A, in brief, the key requirement

is for the number of assigned galaxies in a FOF group to vary the least possible from one targeting realisations to one another. We can then tune the amount of anticorrelation by randomly swapping ones with zeros inside the bitwise arrays, effectively undoing part of the anticorrelation that we just imposed, while keeping all the individual properties untouched. The fraction of randomly swapped bits, f_{anti} , is one of our free parameters.

There is a third free parameter, β_{tilt} , that enables us to adjust the \mathcal{F} kernel by linearly changing its steepness along the n_{CFC} direction. While we have not utilized this parameter with any of the dark-time tracers examined here, we did employ it for the bright galaxy sample to address the issue of the pre-assignment mocks being less dense compared to the actual data. Consequently, we needed to suppress the high n_{CFC} tail of the kernel, which would have assigned too many galaxies otherwise.

The FFA approach immediately provides us with proper bitweights that we can use to implement the fiber-mitigation strategies described in section 5. In particular, we can obtain individual weights for each galaxy as $w_{IIP} = (c + 1)/(K + 1)$, where c is the number of times the galaxy have been assigned and $K = N_{bits}$ is the number of realisations. When dealing with FFA mocks we will sometimes refer to them as fiducial or weights, despite the fact that their definition formally differs from that of the completeness weights of data and altMTL mocks.

For all the FFA catalogues discussed in this work we trained the \mathcal{F} kernel on actual data and set $\theta_{LL} = 0.02\text{deg}$, $f_{anti} = 0.1$. Finally, it is important to mention that in its current implementation, the emulator operates with one tracer at a time. Extending it to a multi-tracer strategy is straightforward, but we leave that for future studies.

4 Simulated catalogues

The simulation of DESI DR1 LSS samples (“mocks”) is described fully in [48] and summarized in [46]. Here, we provide some additional details on the simulations of DESI fiber assignment on these mocks. We constructed three classes of fiber-assigned mocks, each with a different point of balance between level of realism and computational cost. The three classes are obtained by coupling two sets of cosmological simulations, described below, with the two different strategies for simulating fiber assignment introduced in section 3,

- AbacusSummit + altMTLs; 25 mocks, one of which⁵ with 128 independent targeting realisations
- AbacusSummit + FFA; 25 mocks with 256 independent targeting realisations each
- EZmocks + FFA; 1000 mocks with 256 independent targeting realisations each

This large collection of mocks enables us to perform all sort of studies that require realistic survey geometry and completeness.

As a foundation for creating fiber-assigned mocks we need lightcones with realistic distribution of the DESI targets, covering the same volume of DR1 data and having the same pre-fiber-assignment number density. A set of 25 independent realisations of such catalogues has been extracted from the **AbacusSummit** N-body suite [49, 50]. It is worth noting, however, that not all features present in real observations, such as star contamination, have been simulated. This high fidelity set is complemented by a larger set of 1000 realisations, obtained

⁵Abacus mock n.11

via approximate, faster methods (EZmocks, [51]). Broadly speaking, the high fidelity set allows us to isolate systematic effects whereas the approximate set can be used to estimate covariance matrices. In this work, we focus on the Abacus set because it provides a more accurate description of the small-scale clustering, leading to a more realistic FA output. Furthermore, for the Abacus set we have samples processed with both the altMTL and FFA methods, facilitating a direct comparison of the two. Unless specified otherwise, the results in this work are based on Abacus mock n.11. All other mocks give consistent outcomes.

4.1 Simulated DESI DR1 target samples

`AbacusSummit` is a suite of large, high-accuracy cosmological N-body simulations designed to meet (and exceed) the cosmological simulation requirements of the DESI survey [49, 50]. There are over 150 simulations with different cosmologies, containing 6912^3 particles in a cubic volume of side $2 h^{-1}\text{Gpc}$. For the mocks on which we performed fiber assignment, the following different snapshots were employed: for LRGs, the snapshots used are `z0.500` for $z < 0.6$ and `z0.800` for $z > 0.6$; for ELGs, the snapshots used are `z0.950` for $z < 1.1$ and `z.1325` for $z > 1.1$; For QSOs, the snapshot used is `z1.400` for all redshifts. Simulated galaxies are placed within dark matter halos based on the results of analysis of early DESI clustering data by [52, 53]. These snapshots were then concatenated to produce mocks with all DESI dark-time tracers at the correct respective target densities.

The DR1 EZmocks were produced at the same redshift snapshots as the `AbacusSummit` mocks, with their clustering trained on the same `AbacusSummit` mocks. They were produced within boxes of side $6 h^{-1}\text{Gpc}$, which is large enough to simulate the entire DESI DR1 sample within each Galactic cap (GC). 2000 such simulations were produced at and concatenated across each of the same snapshots used for the `AbacusSummit` mocks, which provided the North and South GCs. Full details are provided in [48].

4.2 Comparing altMTL and FFA mock catalogues

In the left column of figure 4 we show the angular assignment completeness for the three dark-time tracers, ELGs, LRGs and QSOs, obtained by applying the altMTL approach. The angular patterns, mostly driven by the tile coverage, are fully compatible with those of the data, shown in figure 2. In the right column of figure 4 we present the relative angular density, defined as $\rho_{\text{FFA}}/\rho_{\text{altMTL}} - 1$, of the FFA and altMTL samples of assigned galaxies, where ρ_{FFA} and ρ_{altMTL} are the respective densities. This kind of comparisons can be interpreted as validation tests for FFA against the, more realistic, altMTL method, but, of course, finding good agreement between the two also reinforces our confidence in altMTLs. As shown in the figure, the angular distributions are compatible among themselves, with the less complete regions revealing larger statistical fluctuations but no obvious signs of systematic trends. For reference, the explicit values of the total number of objects in each catalogue pre and post fiber assignment is reported in table 1.

The redshift distribution of the different tracers is shown in figure 5. The altMTL and FFA mocks produce consistent results, represented by the pink shaded area and the dashed blue curve, respectively. As expected, due to fiber assignment, both distributions are significantly suppressed compared to the complete sample (green shaded area). However, when rescaled by a factor $N_{\text{complete}}/N_{\text{assigned}}$, where N_{complete} and N_{assigned} are the total number of objects in the complete and fiber-assigned samples, respectively, the correct amplitude is consistently recovered across all tracer types (solid pink and dotted blue). This confirms that the FA process does not significantly alter the shape of redshift distribution of the targets.

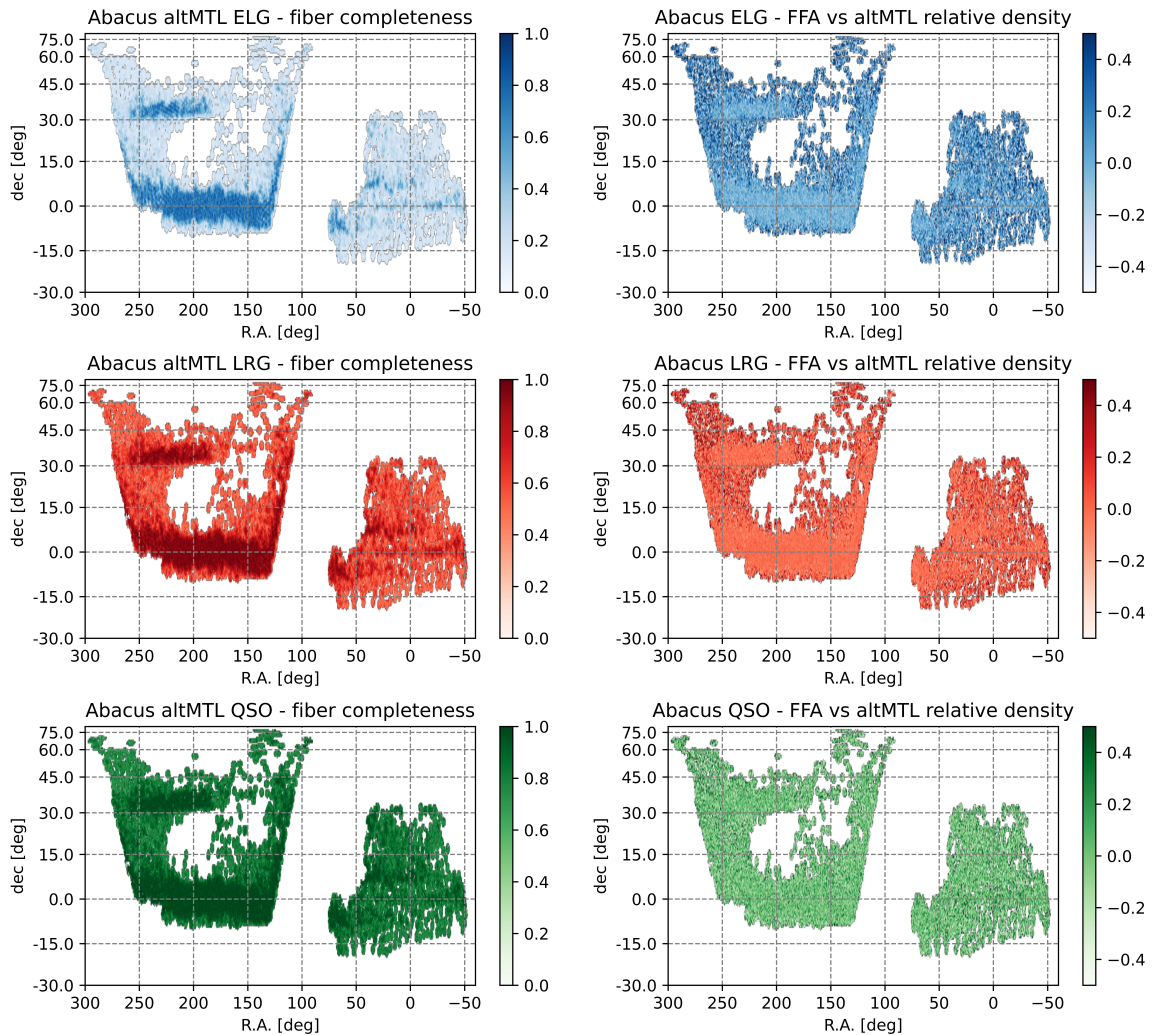


Figure 4. Left column: angular assignment completeness of the three dark-time tracers, ELGs, LRGs and QSOs, as labeled in the figure, measured from altMTL mocks. Right column: relative behaviour of FFA and altMTLs, $\rho_{\text{FFA}}/\rho_{\text{altMTL}} - 1$, where ρ is the angular density of the assigned galaxies.

		altMTL	FFA	Complete
ELG	n. of galaxies	2.41×10^6	2.44×10^6	12.09×10^6
	fraction assigned	0.20	0.20	1
LRG	n. of galaxies	2.16×10^6	2.15×10^6	3.88×10^6
	fraction assigned	0.56	0.56	1
QSO	n. of galaxies	0.86×10^6	0.86×10^6	1.49×10^6
	fraction assigned	0.58	0.58	1

Table 1. Total number of galaxies measured from the Abacus mocks pre and post fiber assignment with both the altLTM and FFA methods. We also report the fraction of assigned galaxies, which slightly differs from the definition of completeness used, e.g., for figure 2, as here we show the ratio between clustering and complete sample. Values for all the three dark-time tracers, ELGs, LRGs and QSOs, are reported.

The only noticeable exception is given by the ELGs at $z > 1.5$. This discrepancy is not a result of fiber assignment; rather, it arises from redshift failures, which we intentionally incorporate into the clustering mocks, after the assignment process. Specifically, the drop in the ELG distribution reflects the fact that, in real data, our ability to accurately measure redshifts above 1.5 is limited by sky-line contamination (see figure 1). This also explains the slight excess of power at lower redshifts in the rescaled ELG distributions, as it reflects the “conservation” of the total number of galaxies.

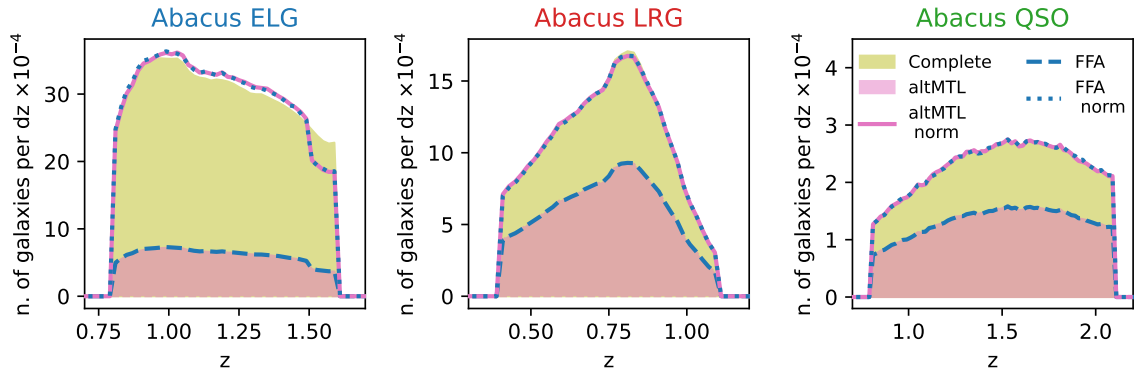


Figure 5. Comparison of the different redshift distribution ($dz = 0.02$) measured from the mocks without fiber assignment (complete sample, green shaded area) and with the two simulated assignment methods considered in this work: altMTLs (pink shaded area), FFA (dashed blue curve). The solid pink and dotted blue curves show the same altMTLs and FFA distributions rescaled by factor $N_{\text{complete}}/N_{\text{assigned}}$ to match the amplitude of the complete sample, where N_{complete} and N_{assigned} are the total number of objects in the complete and fiber-assigned samples, respectively. All three dark tracers are considered, as labeled in the figure.

Figure 6 displays the histograms of the fiducial weights obtained through the altMTL and FFA approaches, represented by the pink shaded and blue shaded rectangles, respectively. As explained in detail in section 5, the two weight definitions are similar but formally distinct: the fiducial altMTL weights are derived from completeness considerations, while the fiducial FFA weights correspond to inverse probabilities, w_{IIP} . In this section, we are not evaluating the performance of these weights as mitigation strategies; rather, we are examining how well the approximate FFA mocks replicate the key features of the more realistic altMTL mocks, including the behaviour of the corresponding default weights. For this reason, we refer to both of them as “fiducial”.

Unlike altMTLs, the FFA weights are not inherently constrained to take discrete integer values. Therefore, to facilitate comparison between the two, we include in the figure an illustrative histogram that represents the FFA weights rounded to the nearest integer (blue empty rectangles). Overall, we observe a satisfactory agreement, which aligns with what can be reasonably expected considering the two different definitions. More specifically, there is a general trend, more pronounced in the ELG case, for altMTLs to exhibit a shorter high-weight tail and a significantly larger proportion of weights equal to one. The high- w tail might influence the noise properties of the clustering statistics. However, as the weight increases, the number of galaxies drops rapidly, making it difficult to anticipate whether the net effect will be significant.

These discrepancies between altMTL and FFA weights may also arise from variations in how fiber assignment is implemented. To separate the two potential contributions, we can

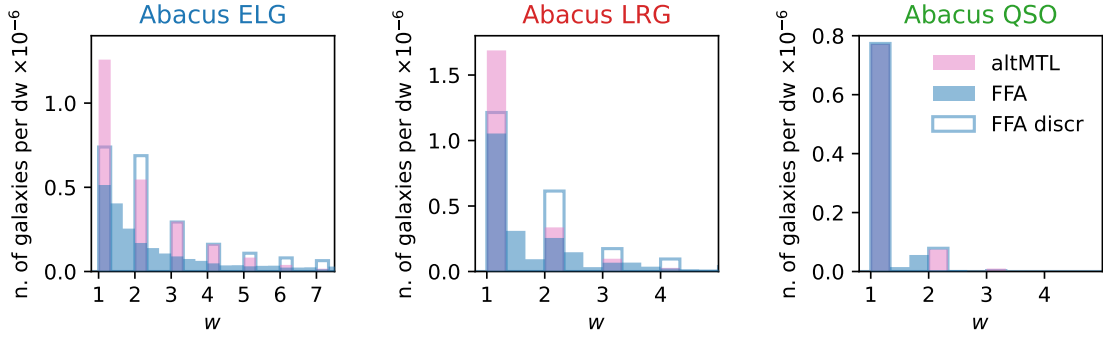


Figure 6. Histograms of the fiducial weights from the altMTL (blue shaded) and FFA (pink shaded) mocks ($dw = 0.33$). The blue empty rectangles represent the same FFA weights rounded to the nearest integer. All the three dark-time tracers are show, as labeled in the figure.

leverage the fact that, for the data (and altMTL mocks), both weight definitions are available simultaneously. Figure 7 shows the two histograms (shaded pink and green rectangles), together with the rounded version of w_{IIP} (empty green rectangles). The altMTL mocks yield almost indistinguishable results, not shown here for brevity, further reinforcing the reliability of the method. The comparison is not entirely consistent, since, for practical reasons, these plots focus on the full sample of potential targets instead of that of the observed ones used for figure 6. Nevertheless, as illustrated in the figure, the two sets of weights show closer agreement when the fiber assignment strategy is fixed, suggesting that the main source of the discrepancies in figure 6 is the assignment procedure itself, FFA versus altMTLs, at least for the $w = 1$ peak.

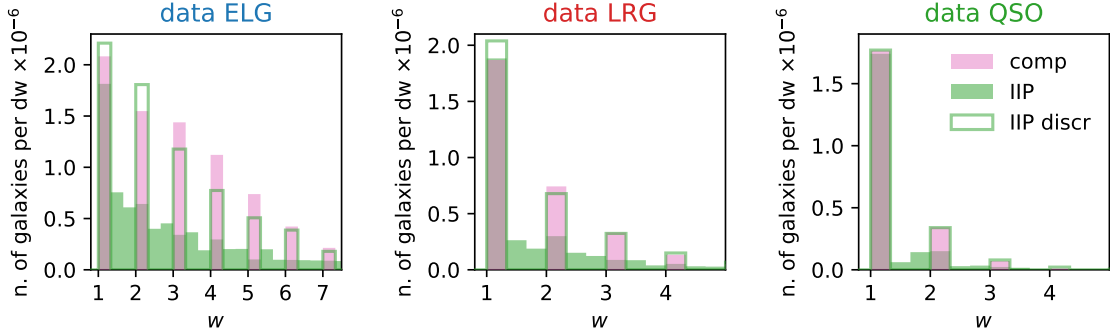


Figure 7. Histograms of fiducial (pink shaded rectangles) and IIP (green shaded rectangles) weights from the data ($dw = 0.33$). The green empty rectangles represent the same IIP weights rounded to the nearest integer. All the three dark-time tracers are show, as labeled in the figure.

Figures 8 and 9 show the collision windows extracted from the altMTL and FFA mocks, respectively. By focusing on the fiducial weight curves, solid pink and solid green, respectively, we observe that while altMTLs yield results fully compatible with those from actual data (figure 3), FFA mocks display a less pronounced collision window, characterized by a slightly steeper transition and reduced overall depth, particularly for the QSOs. It is nonetheless noteworthy the overall qualitative agreement, in particular for the angular size of the window,

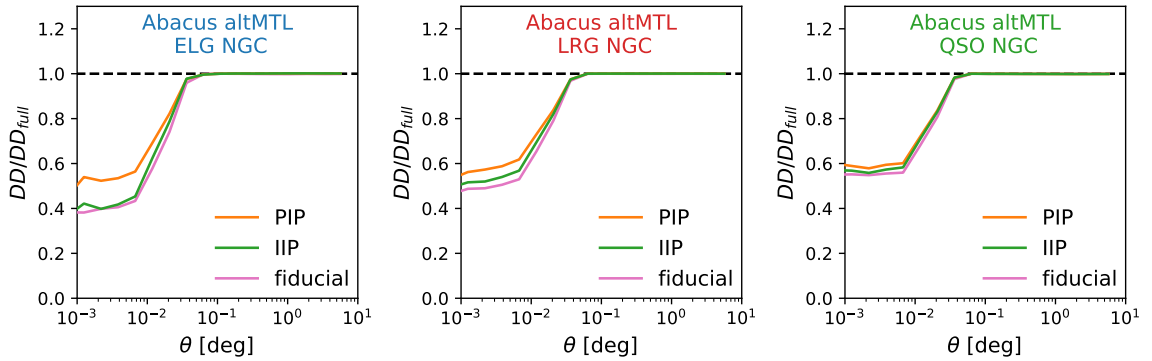


Figure 8. Normalised pair counts measured from altMTL mocks, as a function of the separation angle θ , with fiducial weights (solid pink), IIP weights (solid green) and PIP weights (solid orange), divided by the normalised pair counts from the full sample of potential targets. All three dark tracers are shown, as labeled in the figure.

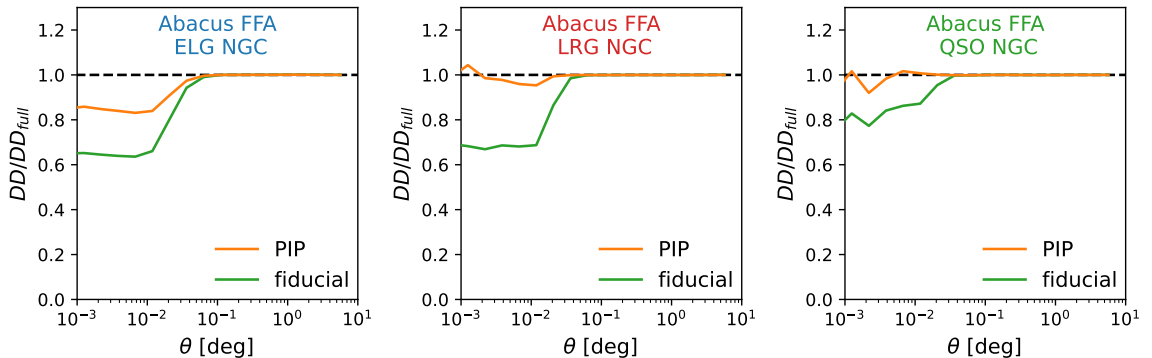


Figure 9. Same as figure 8 but for FFA mocks. For these mocks, by design, fiducial weights coincide with IIP weights (green solid). In contrast to altMTLs, for FFA we do not have access to a proper full sample, so we use the complete sample instead.

given the radical approximations involved.

5 Estimator-level mitigation techniques

In this section we focus on how to counteract the effect of fiber assignment at the estimator level, but it is worth mentioning that some of the techniques described here, namely w_{comp} (or w_{IIP}), are required even when addressing the issue at the model level (see section 6). By *estimator-level* corrections we mean any weighting scheme applied to either individual galaxies or pairs of galaxies (or randoms) at the stage of measuring the clustering statistics of interest. Such schemes are designed to statistically recover the “true” clustering power of all the galaxies in the DESI volume, thus allowing us to ignore the effect of fiber assignment when modelling the statistic of interest for cosmological inference, except, of course, for the covariance part.

We consider four different estimator-level corrections,

1. Completeness

2. Individual inverse probability
3. Pairwise inverse probability
4. Pairwise inverse probability with angular upweighting

detailed in sections 5.1, 5.2. In appendix B we also discuss a fifth approach based on learning the collision window from the mocks and apply it to the data.

All configuration-space measurements in this work are obtained through the [54] estimator,

$$\xi(\mathbf{s}) = \frac{DD(\mathbf{s})}{RR(\mathbf{s})} - 2\frac{DR(\mathbf{s})}{RR(\mathbf{s})} + 1, \quad (5.1)$$

where DD , RR and DR stand for data-data, random-random and data-random (normalised) pairs counts, respectively, and \mathbf{s} is the separation. Fourier space measurements are obtained through the fast-Fourier-transform (FFT) implementation [55–57] of the [58] estimator,

$$P_\ell(k) = \frac{(2\ell + 1)}{I} \int \frac{d\Omega_k}{4\pi} \left[\int d^3r_1 \int d^3r_2 F(\mathbf{r}_1)F(\mathbf{r}_2)e^{i\mathbf{k}\cdot(\mathbf{r}_1-\mathbf{r}_2)} \mathcal{L}_\ell(\hat{\mathbf{k}} \cdot \hat{\eta}) - S_\ell(\mathbf{k}) \right], \quad (5.2)$$

where $\hat{\eta} = \hat{\eta}(\mathbf{r}_1, \mathbf{r}_2)$ is the line of sight (LOS) of the pair formed by galaxies 1 and 2, $d\Omega_k$ the solid-angle element in k -space, \mathcal{L}_ℓ the ℓ -th order Legendre polynomial, S_ℓ the shot noise term (which is non-zero only for $\ell = 0$) and I the normalisation factor. The density fluctuation is defined as $F(\mathbf{r}) = n(\mathbf{r}) - \alpha n_s(\mathbf{r})$, where n and n_s are, respectively, the weighted number density of galaxies and randoms, the latter being $1/\alpha$ times more abundant. The normalization is obtained as $I = \alpha \sum nn_s/dV$, with the sum performed over grid cells of fixed size $dV^{1/3} = 10h^{-1}\text{Mpc}$, as described in [46]. The shot noise term is given by $S_0 = (\sum n^2 + \alpha^2 \sum n_s^2)/I$, where the sums are performed over galaxies and randoms.

5.1 Completeness weights

This approach, originally developed for the SDSS survey [22], relies on two separate sets of weights, described in detail in [46, 59]. In brief, we assign a completeness value to each target, f_{TLID} , evaluated as the inverse of the total number of targets of the same class at the given tile and fiber, and we define the corresponding completeness weight as:

$$w_{\text{comp}} = 1/f_{\text{TLID}}. \quad (5.3)$$

f_{TLID} tells us how many targets were competing for the fiber and, consequently, represent a direct estimate of the assignment probability, in the case of a uniformly randomized process. f_{TLID} alone does not capture all incompleteness effects, such as, e.g., fibers assigned to standard stars or sky in order to meet a minimum threshold, or low-priority ELGs having no chance of being assigned in the proximity of an LRG (exclusion regions induced by QSO are already included in the veto mask). These processes ultimately depend on random choices, made by the targeting algorithm, whose impact on the completeness is expected to be fairly uniform within any given group of overlapping tiles. For each of such groups we therefore introduce an additional completeness fraction, defined as

$$f_{\text{tile}} = \frac{N_{\text{assigned}} + N_{\text{associated}}}{N_{\text{tot}}}, \quad (5.4)$$

where N_{tot} is the total number of targets of a given type and $N_{\text{associated}}$ the number of not assigned targets at combinations of tile and fiber that got assigned to that target type. The

numerator corresponds to the sum of the w_{comp} weights over the assigned galaxies, thus if we divide each weight by f_{tile} their sum becomes N_{tot} , by construction. In practice, rather than introducing an additional weight $1/f_{tile}$ for the data, we use f_{tile} as a weight for the randoms. The procedure is well defined, as the same tile groupings exist in both the data and random samples, and it serves as an effective way to mitigate noise by utilizing a large number of tracers, since the number of randoms can be made much larger. Throughout the manuscript we frequently refer to this weighting scheme as “fiducial weights”.

5.2 Inverse probability weights

Introduced by [16] and further developed by [17], inverse-probability (IP) weights are designed to counteract the effect of missing observations in an exact way. Specifically, it was shown that unbiased n -point clustering measurements can be obtained by upweighting each galaxy n -plet by the inverse of its probability of being observed, provided there are no zero-probability n -plets.

In the case of 2-point statistics the appropriate inverse-probability correction is therefore pairwise in nature: each galaxy-pair requires its own specific weight, not necessarily expressible as the product of two individual weights. We refer to this weighting scheme as pairwise inverse probability (PIP) and denote the corresponding pair weights as w_{PIP} .

However, if we can find some characteristic angular separation θ_{ind} above which the probability of a galaxy of being observed is (approximately) independent on whether its neighbours have been observed or not, for $\theta > \theta_{ind}$ we can safely use individual weights. We refer to this weighting scheme as individual inverse probability (IIP) and denote the corresponding galaxy weights as w_{IIP} .

In summary, for both PIP and IIP, the correlation function is obtained by substituting in eq. (5.1),

$$DD(\mathbf{s}) = \sum_{\mathbf{x}_m - \mathbf{x}_n \approx \mathbf{s}} w_{mn} , \quad (5.5)$$

but the weights can be factorised as $w_{mn} = w_m^{IIP} w_n^{IIP}$ only under the IIP assumption. For the, more general, PIP scheme each pair weight is obtained via the bitwise weights introduced by [16] and implemented for the DESI survey by [47]. The symbol $\sum_{\mathbf{x}_m - \mathbf{x}_n \approx \mathbf{s}}$ is shorthand notation for a binned sum. DR and RR are always obtained using individual weights.

The two approaches can be combined. As a way to reduce the computational complexity of the pure PIP approach in Fourier space, [17] proposed an hybrid IIP-PIP procedure. If θ_{ind} exists, this strategy yields results identical to those obtained by solely using PIP weights. Therefore, we still categorize it as a PIP approach, for simplicity. In this scenario, the power-spectrum (multipole) estimator becomes

$$\begin{aligned} P_\ell^{PIP}(k) &= P_\ell^{IIP}(k) + \frac{(2\ell + 1)}{I} \int \frac{d\Omega_k}{4\pi} \sum_{ij} A_{ij} e^{i\mathbf{k} \cdot (\mathbf{r}_i - \mathbf{r}_j)} \mathcal{L}_\ell(\hat{\mathbf{k}} \cdot \hat{\eta}_{ij}) \\ &= P_\ell^{IIP}(k) + (-i)^\ell \frac{(2\ell + 1)}{I} \sum_{ij} A_{ij} j_\ell(k s_{ij}) \mathcal{L}_\ell(\hat{\mathbf{s}}_{ij} \cdot \hat{\eta}_{ij}) , \end{aligned} \quad (5.6)$$

where P_ℓ^{IIP} is obtained by supplementing eq. (5.2) with IIP weights, j_ℓ are spherical Bessel functions of the first kind and

$$A_{ij} = w_{ij}^{PIP} - w_i^{IIP} w_j^{IIP} . \quad (5.7)$$

The selection probability of each object (galaxy or pair) and, most importantly, its inverse $1/p$, can be evaluated by running the targeting algorithm K times and counting how many times, c , it gets selected (we follow the [17] notation and refer to c as *recurrence*). For a detailed description of how the procedure is implemented for DESI see [47]. There is more than one, formally correct, way of combining recurrence and number of realisations into an inverse-probability estimator, which returns the true value of $1/p$ for $K \rightarrow \infty$. Given the limited amount of computational resources available, it is of crucial importance to use an estimator that converges the fastest possible with K . The issue was studied by [17], who recommended adopting the form

$$w^{eff} = \frac{K + 1}{c + 1}, \quad (5.8)$$

and named it *efficient estimator*⁶. The authors also derived the exact distribution of the recurrence of a pair given the recurrences of the two individual galaxies, i and j , when the latter are selected independently, $\mathcal{P}_K(c_{ij}|c_i, c_j)$. This allowed them to compute the expectation value, $g_K(c_i, c_j)$, of w_{ij}^{eff} , which they then used to obtain an accurate pair-count normalisation, free from unwanted sampling artefacts.

For DESI we essentially follow the same strategy, but with an important refinement: rather than using g_K for the normalisation, we include it in the pair-counting core. Specifically, our definition of PIP weights becomes

$$w_{ij}^{PIP} = w_i^{eff} w_j^{eff} \frac{w_{ij}^{eff}}{g_K(c_i, c_j)}, \quad (5.9)$$

unless otherwise stated⁷. In other words, we factorise the weight into three terms: two purely individual, w_i^{eff} , and one purely pairwise, $w_{ij}^{eff}/g_K(c_i, c_j)$, which incorporates the selection correlation. The advantage over doing it at the normalisation level is that, with this pair-by-pair strategy, any residual scale dependent effect induced by, e.g., the clustering of low probability objects⁸, is removed. Individual and pairwise recurrences c are obtained, as usual, from the bitweights as `popcnt(w_i^{bit})` and `popcnt($w_i^{\text{bit}} \& w_j^{\text{bit}}$)`, respectively, where `popcnt` is the standard bit-counting function and `&` the bitwise “and” operator.

As illustrated in figure 10, sampling artefacts become particularly significant for highly incomplete samples, such as the ELG catalogs. The right panel of the figure compares normalized angular pair counts obtained with (solid) and without (dotted) the new definition introduced in eq. 5.9. The combined impact of low assignment probability and a limited number of realizations ($K = 128$) causes the dotted curve to diverge significantly from the correct value, even on large angles. If not corrected, the 5% systematic effect in the figure (note the different ordinate scale compared to the previous pair count plots) would result in an arbitrarily large error on 2-point measurements when $DD/RR \rightarrow 1$, i.e. on large scale. In contrast, the solid curve converges nicely to 1 for separations larger than the collision window (but see section 5.2.2 below). Higher probability samples like LRGs and QSOs (not shown in the figure) are significantly less affected by these sampling artefacts.

5.2.1 Inverse probability with angular upweighting

By construction PIP weights return (on average) the true clustering of the pairs with non-zero probability. The presence of zero-probability pairs manifests itself as power suppression

⁶In this scenario K represents the number of replicas, the true realisation is not included in the count.

⁷PIP weights are simply defined as $w_i^{PIP} = w_i^{eff}$, unless otherwise stated.

⁸By low probability we mean probability roughly of the order of ten divided by the number of realisations.

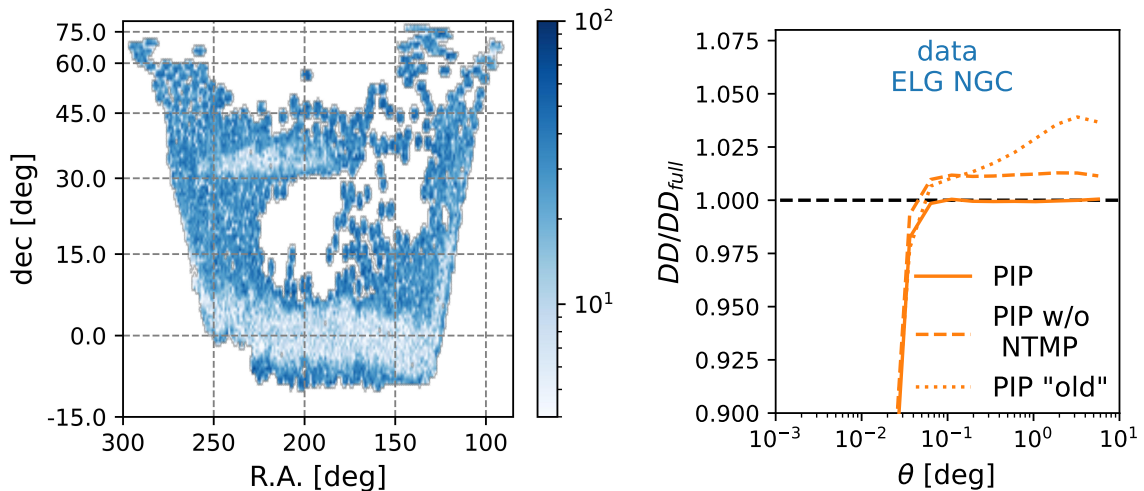


Figure 10. Left panel: angular number density of zero-probability ELGs in the north galactic cap (real data). Right panel: normalised pair counts as a function of the separation angle θ obtained using PIP weights with and without the sampling correction described in eq. 5.9, solid and dotted orange curves, respectively, and without w_{NTMP} , orange dashed, all divided by the normalised pair counts from the full sample of potential targets. Please note the different scales on the y axes compared to the other pair count plots in the paper.

on the scales where such pairs are more abundant, typically those corresponding to small separations. When it occurs, it is always possible to compensate through angular upweighting [16, 23]. This comes with the implicit assumption that the clustering of the zero-probability pairs can be represented by the clustering, on the same angular scale, of the observed pairs. Despite this being not necessarily true in general, it is in practice a valid assumption in most realistic scenarios. Specifically, for DESI the main sources of zero-probability pairs are 1-pass regions and priority conflicts. The fact itself that PIP weights alone do not fully correct for the collision window is direct proof of the presence of these pairs in real DESI data (figure 3), altMTLs (figure 8) and, to a lesser extent, FFA mocks (figure 9).

At variance with what originally proposed by [16] and implemented in the various works that followed (see [47]), here we use normalised pair counts for the angular upweighting factors, $(DD_{full}^{raw}(\theta)/p_{full}) \times (DD^{raw}(\theta)/p)^{-1}$, and similarly for the cross count DR , where the subscript *full* stands for full sample, as defined above, and p is the total number of weighted pairs. The advantage of using normalised pair counts over raw pair counts is mostly practical, as it allows us not to touch the normalisation factors⁹ of ξ and, most importantly, of P .

We do in fact apply angular upweighting also to the power spectrum, for the first time. Specifically, we modify eq. (5.7) as follows,

$$A_{ij} = \frac{DD^{full}}{DD^{PIP}} w_{ij}^{PIP} - w_i^{IIP} w_j^{IIP}, \quad (5.10)$$

with DD^{full} and DD^{PIP} respectively indicating the galaxy-galaxy count of the full and of the (PIP-weighted) clustering sample, normalised as described above, and w_{ij}^{PIP} defined in

⁹This is desirable, e.g., if one wants to restrict angular upweighting to a limited range of separations, as a way to save computational resources.

eq. (5.9).

5.2.2 Counteracting the effect of zero-probability galaxies: w_{NTMP}

The left panel of figure 10 shows the distribution across the north galactic cap of the ELG targets that, based on the information encoded into the bitweights, have zero probability of assignment (about 2.5×10^5 objects). Different tracers and regions exhibit similar behavior, albeit to a lesser degree. It is evident that these zero-probability galaxies are not randomly distributed; instead, they follow a pattern that, not surprisingly, resembles the completeness pattern (see, for example, figure 2) and hence correlates with n_{tile} . As already said, by definition inverse probability weights cannot counteract the effect of zero-probability objects or, more precisely, clustered zero-probability objects. The impact of these “unobservable” galaxies on the angular pair counts is shown in the right panel of figure 10 by the dashed orange curve. We observe a $\sim 1\%$ systematic effect, which, as mentioned earlier, would lead to a severe large-scale error in 2-point measurements if not appropriately countered.

For this reason we define f_{NTMP} as the sum of the IIP weights, w_{IIP} , of all the successfully observed galaxies with a given n_{tile} value divided by the total number of galaxies in the full sample with the same n_{tile} value, where NTMP stands for “ n_{tile} missing power”¹⁰. Since the expectation value of $\sum w_{IIP}$ is, by construction, the number of galaxies with a non-zero probability of assignment, $1 - f_{NTMP}$ estimates the fraction of zero-probability galaxies with a given n_{tile} . We can therefore use it to compensate for the pattern shown in figure 10 by either defining an additional weight for the galaxies, $w_{NTMP} = 1/f_{NTMP}$, or for the randoms, $w_{NTMP}^{\text{randoms}} = f_{NTMP}$. Unless otherwise stated, for the measurements presented in this work we opted for the latter option, simply because the larger number density of the randoms helps to reduce the noise.

Please note that we have deliberately formalized the problem in a way that resembles the definition of f_{tile} in section 5.1 to emphasize the similarities between the two quantities: the pairs (w_{comp}, f_{tile}) and (w_{IIP}, f_{NTMP}) play analog roles, with the first element representing an estimate of the inverse probability of assignment and the second element compensating for what the first does not capture. Just as with the f_{tile} definition, f_{NTMP} could also be constructed from tile groups, rather than n_{tile} . We defer this investigation to future work.

As discussed in the previous section, angular upweighting is another effective countermeasure for zero-probability galaxies. For example, with angular upweighting the ratio of the angular pair counts, right panel of figure 10, would be identically equal to one across the entire separation range. The primary role of w_{NTMP} is to bypass the angular upweighting step on large scale, where the assignment probabilities are independent. This is essential to enable the use of PIP weights in Fourier space, and desirable, though not critical, in configuration space.

6 Model-level mitigation techniques

We classify as *model level* any fiber mitigation technique applied at the stage of cosmological inference, i.e. when evaluating the theoretical value of the clustering statistics, as either a function of compressed variables or actual cosmological parameters. Note, however, that, despite their name, neither the method presented here nor those proposed in the literature

¹⁰We could have simply counted the number of zero probability objects instead of using $\sum w_{IIP}$. The primary benefit of the latter approach is that it would also account for unidentified systematic effects, should they exist.

are entirely model level, but rather hybrids, as they require some auxiliary estimator-level action, such as weighting and applying scale cuts.

For DR1 different model level corrections were explored, including the method proposed by [15] and its natural extension to achieve the accuracy required by DESI: measuring the shape of the collision window from mocks, rather than assuming a top-hat shape. This approach is essentially the model-level counterpart to the estimator-level technique presented in appendix B and shares with it good performances but also the limitation of being inherently mock dependent. For the full-shape analysis we therefore opted for the more the agnostic θ -cut approach, presented in a dedicated paper [30], and summarized below.

6.1 Removing small angular separations (θ -cut)

Once established that θ_{ind} exists, it becomes natural to simply exclude the scales with $\theta < \theta_{ind}$ from the analysis. Specifically, we can pass to the clustering estimators only the pairs with angular separation larger than some θ_{cut} , with $\theta_{cut} \gtrsim \theta_{ind}$, and model the resulting correlations and power spectra accordingly. This is a valid fiber-mitigation approach as long as the following three conditions hold. First, θ_{cut} must be small enough such that only a small fraction of the information that we want to extract gets lost. Second, we need a robust, unbiased correction for fiber incompleteness on scales larger than θ_{cut} , with the advantage that, on such scales, we can rely on individual weights, by definition. Third, the cut must be computationally tractable when applied to both the clustering estimator and the corresponding model.

A thorough analysis of θ -cut approach can be found in [30], where the above requirements are discussed in detail. In brief, figure 3 shows that by adopting $\theta_{cut} = 0.05\text{deg}$ we can safely exclude the scales where the assignment is intrinsically pairwise in nature. The standard 2-point clustering analysis [31] extracts information from linear and quasi-linear modes, corresponding to $k \lesssim 0.2h\text{Mpc}^{-1}$ or, equivalently, separations around $30h^{-1}\text{Mpc}$ and above. As a reference, at $z = 1$, adopting a Planck-like cosmology for the redshift-to-distance relation, a 0.05deg angular separation translates into a transverse comoving separation of approximately $3h^{-1}\text{Mpc}$, which suggests that most of the quasi-linear information is indeed preserved by the θ -cut. This is confirmed by direct tests on mocks, through the comparison of the posterior of cosmological parameters with and without θ -cut, as shown in the companion paper.

The effectiveness of the of the individual weights, either w_{comp} and w_{IIP} , on scales larger than θ_{cut} is demonstrated, at the level of angular clustering, by figure 3 and confirmed for three-dimensional clustering by the results presented in section 7. The theoretical motivation of why w_{IIP} weights yield unbiased clustering is presented in [17] and can be extended to w_{comp} if we adopt the inverse-probability interpretation discussed in section 5.

Making the clustering estimators insensitive to small angular separations is straightforward in configuration space but not so trivial for the power spectrum, due to the non-local nature of the Fourier transform. We address the problem by adopting the pair-count based solution introduced by [17] and summarised above, eq. (5.6), with the difference that, in the case of pure pair removal, eq. (5.7) simply becomes $A_{ij} = -w_i w_j$. The procedure is applied to the randoms, as well. Lastly, since the impact of θ -cut on the measured clustering is purely geometrical, it can be modelled along the line of what is usually done for survey footprint and selection effects [60, 61], i.e. as an additional contribution to the window matrix. Specifically,

the expectation value of the ℓ -th Legendre multipoles of the the correlation function reads

$$\langle \widehat{\xi}_\ell^{cut}(s) \rangle = \sum_{\ell'} W_{\ell\ell'}^{cut}(s) \xi_{\ell'}(s) , \quad (6.1)$$

whereas for the power spectrum multipoles, including a wide-angle expansion [62] encoded by the index n , we get

$$\langle \widehat{P}_\ell(k) \rangle = 4\pi \int k'^{2-n} dk' \sum_{n,\ell'} W_{\ell\ell'}^{cut,(n)}(k, k') P_{\ell'}(k') . \quad (6.2)$$

The explicit expressions for the window matrices, $W_{\ell\ell'}^{cut}(s)$ and $W_{\ell\ell'}^{cut,(n)}(k, k')$, are provided in [30], together with the description of the numerical algorithm used to compute them efficiently.

One potential issue with θ -cut is given by the extent of the window tails in Fourier space, which couples long and short modes. The resulting window matrix is very non-diagonal. This means that in order to model the θ -cut power spectrum on the scales of interest for standard cosmological analyses $k \lesssim 0.2h\text{Mpc}^{-1}$, one should formally know the correct amplitude of the theory power spectrum on smaller scales, where the theory is not valid anymore. This effect was found to be negligible for DESI, as shown in [30]. Nevertheless, a general strategy to compactify the window so that the impact of high theory modes gets suppressed was developed [63] and successfully implemented by [30].

7 Comparison of results

In this section we present the clustering measurements from the Abacus mocks processed through the altMTL and FFA pipelines and we compare them to actual data measurements from DESI DR1. The goal is twofold: assessing the relative effectiveness of the different fiber incompleteness countermeasures and testing the soundness of the two strategies we have implemented to simulate fiber assignment. We do not show clustering measurements obtained with the θ -cut approach as they are already presented and extensively tested against the corresponding analytical model in [30].

Given the high computational demands of the `fiberassign` code, we could only afford altMTL bitweights for one Abacus mock. We arbitrarily chose mock 11 and ran $K = 128$ alternative realisations of the targeting in addition to the original one and use them to construct the corresponding bitweights, with $N_{bits} = 128$. The same number of alternative realisations/bits was used for the data sample. With FFA we do not have such limitations, we therefore built $K = 256$ realisations for each of the 25 Abacus mocks. To ensure a fair comparison among the different approaches, in this section we present clustering measurements for mock 11 alone, including all three dark-time tracers, ELGs, LRGs and QSOs. An example of measurements averaged over a larger set of FFA mocks can be found in appendix C. For compactness' sake we only show clustering measurements for the North Galactic Cap (NGC). The results for the South Galactic Cap (SGC) are fully compatible but exhibit a lower signal-to-noise ratio due to a combination of higher incompleteness and a smaller area, see figures 2 and 4.

We measure the correlation functions using both linear bins of $4h^{-1}\text{Mpc}$ and logarithmic bins with centers given by $s_n = 10^{x_0+n\Delta x}$, where $x_0 = -0.894$ and $\Delta x = 0.211$. In all the figures we show a combination of the two, with logarithm bins (and abscissa scale) for

$s < 20h^{-1}\text{Mpc}$ and linear bins (and abscissa scale) on larger separations, as a way to compress all the relevant information into a single plot. The numbers of galaxies in the NGC footprint are $N_{\text{ELG}} = 1.8 \times 10^6$, $N_{\text{LRG}} = 1.5 \times 10^6$, $N_{\text{QSO}} = 5.6 \times 10^5$. For each tracer we have multiple random samples spanning the same volume, with $N_{\text{ran}} = 1.2 \times 10^7$ particles each, which we can combine in different ways to reach a convenient balance between computational speed and mitigation of the noise. For all the ξ measurements presented here the pair counts supplied to eq. (5.1) are obtained by merging 4 random samples. More precisely, we use the full merged sample only for $s < 20^{-1}\text{Mpc}$, whereas for the larger separations we split it into 6 randomly selected subsets and perform individual pair counts, which we then combine with the appropriate normalisation factors in order to obtain consistent all-scale estimates [64]. For the ξ estimates that rely on angular upweighting we compute angular pair counts in logarithmic bins with centers given by $\theta_n = 10^{x_0+n\Delta x}$, where $x_0 = -3.7$ and $\Delta x = 0.66$. The upweighting is performed for angular separations smaller than $\theta_{\text{max}} = 0.2\text{deg}$. Thanks to w_{NTMP} weights, the choice of θ_{max} does not influence our results, as far as it is larger than the angular size of the collision window¹¹.

The power spectra are estimated as described in section 5, using a grid of 1024^3 nodes, painted with cloud-in-cell assignment (with the appropriate deconvolution in Fourier space, [65]). For all tracers, we obtain the density contrast by merging 2 random samples and, when computing FFTs, we mitigate the aliasing by performing one deinterlacing step, [66]. As usual, the full 3-dim power spectrum is eventually compressed into Legendre multipoles, with linear k bins of size $0.005h\text{Mpc}^{-1}$. We always subtract the shot noise contribution, as described in section 5. For the P estimates that rely on angular upweighting we compute angular pair counts in logarithmic bins with centers given by $\theta_n = 10^{x_0+n\Delta x}$, where $x_0 = -5.7$ and $\Delta x = 0.64$. Both PIP correction and angular upweighting (eqs. 5.7, 5.10) are limited to angular separations smaller than $\theta_{\text{max}} = 0.25\text{deg}$. Similarly to the correlation function case, θ_{max} does not influence our results, as long as it exceeds the angular size of the collision window.

As explained in [46], for purely technical reasons, we created two separate versions of the clustering catalogs containing the same galaxies: one for use with fiducial weights and the other for use with IP weights. For all the measurements presented below it is intended that the appropriate catalog was used, except for the FFA mocks, where separate catalogs were unnecessary.

Below we present results for the four different estimator-level corrections discussed in section 5: fiducial, IIP, PIP and PIP weights plus angular upweighting. The clustering catalogs that we used to obtain them feature a **WEIGHT** column that specifies the total weight to be applied to galaxies and randoms. This weight incorporates factors that are not related to fiber assignment, as imaging systematics and redshift failures [46]. All the clustering measurements in this section are obtained by reading w_{galaxies} and w_{randoms} directly from these **WEIGHT** columns. For the fiducial weights, no further action is needed. Similarly, the IP dedicated catalogues already include IIP weights in the **WEIGHT** column so that no further operation would be required to use them. However, this catalogues were created before of the introduction of w_{NTMP} weights, which are uniquely used in this work. We therefore had to, first, compute them and, second, redefine the random weights as $w'_{\text{randoms}} = w_{\text{randoms}} \times f_{\text{NTMP}}$ (see section 5). For PIP weights we follow the same procedure, except that we also need

¹¹Formally, a larger θ_{max} would have beneficial effects on the statistical error, as show in [23]. However, here we focus on systematics effects, using angular upweighting to address the presence of zero-probability pairs, and a lower value of θ_{max} allows us to reduce the overall computational time.

to remove the IIP contribution by defining $w'_{\text{galaxies}} = w_{\text{galaxies}}/w_{\text{IIP}}$. The power that is removed through this operation is ultimately restored (and enhanced) by the bitweights, w_{bit} , which we obtain directly from BITWEIGHTS column. When we use PIP weights with angular upweighting we also need the full catalogue and the correspondent random sample. The weights for galaxies, observed galaxies and randoms, are all obtained from the WEIGHT_NTILE columns, which account for variation of $w_{\text{tot}}/w_{\text{IIP}}$ with n_{tile} [46]. To ensure consistency, the angular pair counts of the observed galaxies must be computed using bitweights. In addition, at variance with the strategy adopted for the 3-dim pair counts from the clustering sample, for the angular counts we incorporate the NTMP correction into the (observed) galaxy weights, so that $w_{\text{observed}} = w_{\text{ntile}}/f_{\text{NTMP}}$ plus bitweights, as we do not perform random-random pair counts. We intentionally chose not to use FKP weights [67] in order to keep our studies centered on fiber assignment issues; however, all the estimators employed in this work are fully compatible with FKP weights.

In Figure 11 we show the correlation function multipoles measured from altMTL mocks. The shaded areas represent the standard deviation extracted from the fiducial DESI covariance matrices, which, for the correlation function, are obtained through a semi-analytical approach, implemented as the `RascalC` code [68]. Since on small scale this approach is not reliable, we complete the semi-analytic covariance with a purely empirical one. Specifically, for $s < 20h^{-1}\text{Mpc}$, we use as error bars the standard deviation of the correlation functions measured with PIP weights plus angular upweighting from 7 FFA mocks. To reduce noise caused by the small number of mocks, we interpolate it with a power law. This approach clearly limits the reliability of the small-scale error and should be seen more as a convenient method for normalizing and visualizing the systematic trends of the different mitigation techniques, rather than providing an exact quantification of their confidence intervals.

For the altMTL mocks we can test all the four different estimator-level corrections considered in this work: fiducial weights (pink solid), IIP weights (green solid), PIP weights (orange solid), PIP weights plus angular upweighting (blue solid). Their effectiveness can be assessed by comparing them with the “true” correlation, directly measured from the complete sample (grey dashed), i.e. the set of all potential targets of a given tracer class, ELG, LRG or QSO. For the complete samples we also show the error bars obtained from the variance of 7 mocks, as a rough measure of the intrinsic scatter of the clustering itself, regardless of fiber assignment.

Figure 11 clearly shows that all the various weighting schemes produce fully consistent estimates at scales greater than $20h^{-1}\text{Mpc}$, and, most importantly, they all successfully recover the expected clustering, i.e. that of the complete sample. At smaller scales, the impact of the collision window is distinctly visible, especially in the insets at the bottom of each plot, as a suppression of the power. The effect becomes more pronounced and extends to larger scales for higher order multipoles. The behavior of fiducial and IIP weights is similar, which is expected, since the former can be seen as a limiting case of the latter, as discussed in section 5. More precisely, the overall trend across different tracers and multipoles seems to indicate that IIP weights are slightly more effective in countering the collision window, but not in a significant way. PIP weights provides a further boost against the small-scale suppression, but the adjustment is still not sufficient to restore the “true” clustering. This is not surprising, given the substantial number of zero-probability pairs, as illustrated in figure 8. To compensate for these pairs we supplement PIP weights with angular upweighting. As demonstrated in the figure, this combination produces estimates that are closer to the “truth” than the others, with no discernible trend in the residuals.

For what concerns the statistical error, the figure clearly shows that, for the ELG sample, the fiducial weights produce clustering estimates with less noise on large scale. A quantitative assessment of this effect, based on the scatter between 60 jackknife samples extracted from the ELG data sample, is presented in [46]. The test reveals that the variance obtained with IIP weights is approximately 60% greater than that obtained with fiducial weights. This result cannot be straightforwardly extrapolated to the measurements presented in this work, as they are obtained with different settings, the most significant being the inclusion of the w_{NTMP} weights discussed above, but the trend seems confirmed, at least qualitatively (top row of figure 11).

It is important to note that the excess variance is significant only for the ELG sample, which is highly incomplete (see table 1) and characterized by the lowest assignment probabilities, with a large portion of the survey footprint covered by just one pass of the instrument. In the next data release, dramatic improvements in all these areas are expected, and any variance-related issues should be highly suppressed.

We can identify two primary underlying causes for the excess variance. Firstly, low probabilities (more precisely, high inverse probabilities) are inherently more difficult to sample, where by low probability we roughly mean $p \lesssim 10/N_{bits}$. With only 128 realisations, a relatively small number of bits compared to previous studies in the literature, the variance of the weights is directly impacted by sampling noise, with the most unlikely targets being more exposed to this effect.

Secondly, even if the probabilities were perfectly measured ($N_{bits} \rightarrow \infty$) the resulting shot noise properties would not necessarily align with those of the fiducial weights. IP weights are designed to be agnostic, meaning they do not rely on specific assumptions about the targeting algorithm or the underlying clustering. This is accomplished by compensating for each individual assignment choice made by the algorithm, such that the contribution of each object or pair is exactly 1, when averaged over all possible realisations. As a consequence, in extreme scenarios, some pairs might be given an inflated weight or, more precisely, some pairs might obtain different weights despite being statistically equivalent.

One possible cause of this effect is chance alignments with higher-priority targets, though it is not immediately clear why the fiducial weights would be less susceptible to this issue. It is nonetheless true that w_{comp} effectively acts as a local smoothing of the probabilities over areas the size of the fiber patrol radius, suggesting that averaging over (inverse) probabilities could indeed have an impact.

Additionally, the fiducial weights include the f_{tile} term, which compensates for missing power not captured by w_{comp} on the tile-group angular scale. In contrast, the equivalent correction for the IP weights, w_{NTMP} , follows a simpler n_{tile} angular pattern. The f_{tile} correction is more accurate, in the sense that it adjusts for fluctuations on a more localized scale, potentially helping to reduce variance. There are straightforward countermeasures to address all the aforementioned potential causes of variance-related issues, such as increasing the number of realisations to mitigate sampling noise and redefining w_{NTMP} based on tile groups, which we leave for future studies.

Figure 12 shows the multipoles of the power spectrum obtained from the same altMTL samples. We adopt the same color coding conventions as in the previous figure. To our knowledge, this is the first time PIP weights have been tested in Fourier space with realistic survey geometry and targeting. The shaded areas represent the standard deviation extracted from the fiducial DESI covariance matrices. For the power spectrum, these are measured from the scatter of 1000 FFA EZmocks and subsequently scaled in amplitude using the semi-

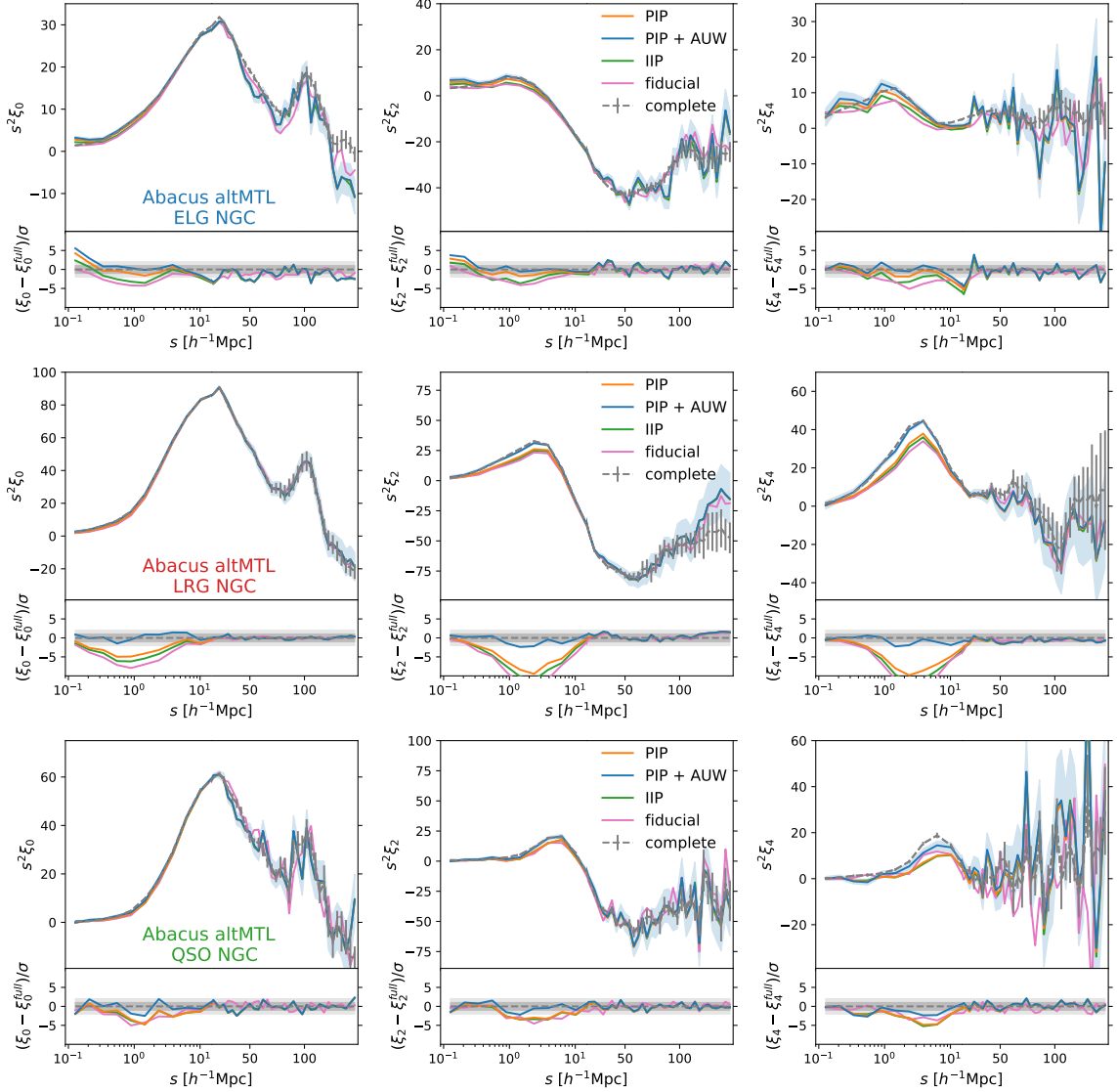


Figure 11. Multipoles of the correlation function ξ_ℓ as a function of the separation s , measured from the altMTL mocks (mock 11, NGC). For visualisation purpose, we follow the common practice of multiplying the amplitude of the multipoles by s^2 . The columns correspond to $\ell = 0, 2, 4$, from left to right, whereas the rows correspond to different dark-time tracers, ELGs, LRGs, QSOs, from top to bottom. The abscissa scale and the binning of correlation multipoles is logarithmic for $s < 20h^{-1}\text{Mpc}$ and linear for larger separations in order to capture all the relevant fiber-assignment-related features at once. The grey dashed curves correspond to measurements obtained from the complete sample with no additional weights and represent a reference target for all the other measurements. The corresponding error bars represent the scatter between different mocks. The remaining curves are obtained from the fiber-assigned samples using: fiducial weights (solid pink); IIP weights (solid green); PIP weights (solid orange); PIP weights plus angular upweighting (solid blue), with the blue shaded area indicating the corresponding 1σ error. In the insets at the bottom of each panel we show the difference between fiber-assigned and complete samples in units of standard deviation.

analytic ξ covariance mentioned earlier, as detailed in [46].

In essence, the same considerations made for the correlation function also hold for the power spectrum, with the key difference being that in Fourier space, the influence of fiber assignment affects significantly larger scales. This is expected, due to the non local nature of the Fourier transform, which produces extended tails in the collision window. The relative behaviour of the different weighting schemes is confirmed, with PIP weights plus angular upweighting recovering the clustering of the complete sample with no indication of systematic error. The only significant exception is the monopole of the QSO sample, which shows an excess of power at large k , consistent across all curves. Given that this sample has the lowest Nyquist frequency due to its large volume, and no similar trends are observed in the corresponding configuration space measurements, we argue this is likely a result of aliasing.

Using the same settings adopted for the altMTL mocks, we measured both the correlation function and power spectrum from the FFA mocks. The results are shown in figures 13 and 14, respectively. With this set of mocks, we can compare three fiber mitigation techniques, fiducial weights (green solid), PIP weights (orange solid) and PIP weights with angular upweighting (blue solid), to the reference provided by the complete sample (gray dashed). It is important to note that, by design (see section 3.2), the fiducial weights in the FFA case are actually IIP weights. Furthermore, unlike the altMTL catalogs, the FFA mocks were not processed to generate a proper full sample for angular upweighting. To work around this limitation, we use the complete sample instead, which is, by construction, cut to the same redshift range of the tracer under investigation. This scenario is not entirely realistic because, in a real situation, we would not have access to all the (spectroscopic) redshifts required to make such a cut, specifically the redshifts of galaxies that were not observed. As a result, in these tests, angular upweighting may be more effective at mitigating systematic effects than it would in a real-world application, at least in principle. In addition, the complete sample does not include w_{ntile} weights, so for the clustering sample we cannot apply the factorization $w'_{galaxies} = w_{galaxies}/w_{IIP}$ described above (or use FKP weights), as that would interfere with angular upweighting. Consequently, the PIP plus angular upweighting measurements show a greater statistical error than they ideally would. Finally, since the complete sample does not incorporate the veto mask of the clustering sample, angular upweighting would technically require normalizing the angular pair counts with the corresponding random samples. For all these reasons, the comparison with the other two weighting schemes is not entirely fair.

In figure 13 the enhanced statistical noise is clearly noticeable, e.g. when comparing the ξ multipoles measured with PIP to those measured with PIP plus angular upweighting, especially for the ELG sample in the top row. Nevertheless, PIP weights with angular upweighting provide unbiased estimates, as expected. Interestingly, PIP weights alone also successfully recover the correct clustering across all scales. This is not surprising, as FFA mocks contain significantly fewer zero-probability pairs compared to altMTL mocks and real data (see figures 3, 8, 9), making angular upweighting unnecessary, at least for addressing systematic effects.

Similar considerations apply to the power spectrum multipoles, shown in figure 14. The most noticeable difference, compared to the correlation function, is arguably the behavior of the PIP measurements from the LRG sample. Specifically, the PIP power spectrum exhibits a slight systematic suppression at high k , which is not readily noticeable in its configuration space counterpart (as seen in the bottom insets of the central row in both figures). This is likely due to the combined effects of the collision window's impact being more pronounced in Fourier space and the fact that we use two different prescriptions for the

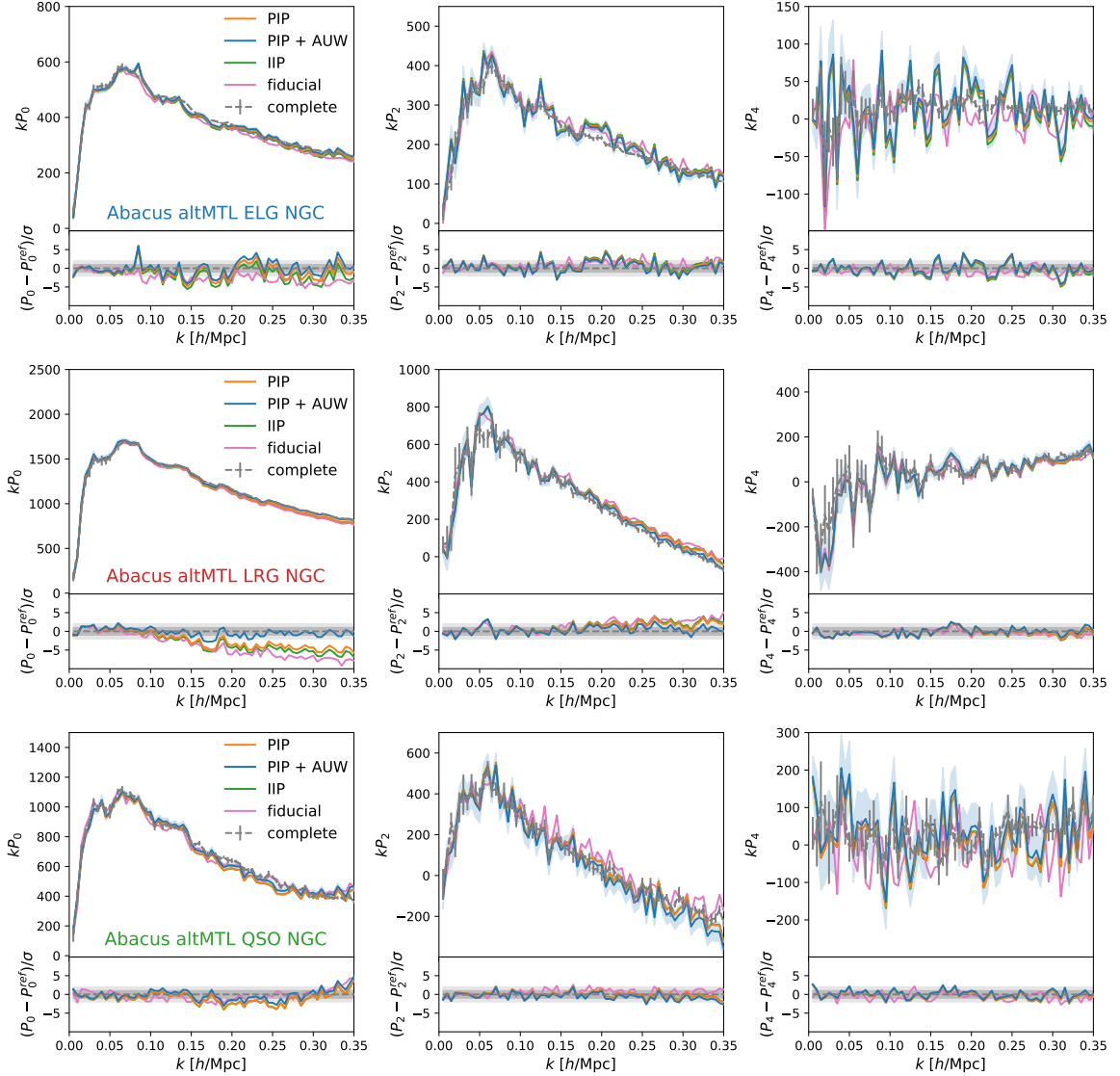


Figure 12. Multipoles of the power spectrum P_ℓ as a function of the wave number k , measured from the altMTL mocks (mock 11, NGC). For visualisation purpose, we follow the common practice of multiplying the amplitude of the multipoles by k . The columns correspond to $\ell = 0, 2, 4$, from left to right, whereas the rows correspond to the different dark-time tracers, ELGs, LRGs, QSOs, from top to bottom. The grey dashed curves correspond to measurements obtained from the complete sample with no additional weights and represent a reference target for all the other measurements. The corresponding error bars represent the scatter between different mocks. The remaining curves are obtained from the fiber-assigned samples using: fiducial weights (solid pink); IIP weights (solid green); PIP weights (solid orange); PIP weights plus angular upweighting (solid blue), with the blue shaded area indicating the corresponding 1σ error. In the insets at the bottom of each panel we show the difference between fiber-assigned and complete samples in units of standard deviation.

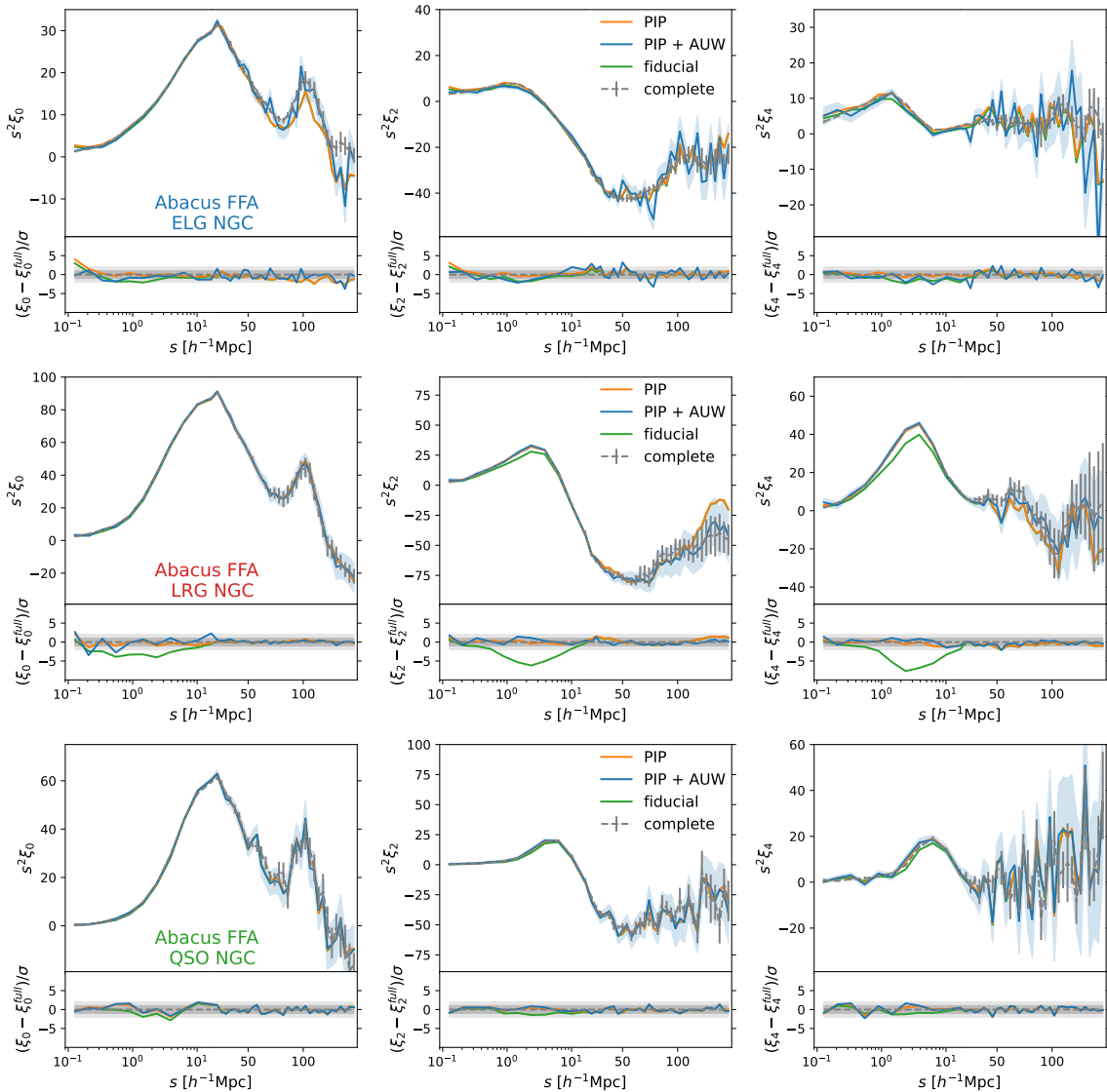


Figure 13. Same as figure 11 but for the FFA mocks (mock 11, NGC). Unlike altMTL mocks, the fiducial weights here (solid green) are, by design, IIP weights.

standard deviations. As mentioned earlier, the small-scale variance of the correlation is not sufficiently reliable for a quantitative comparison of confidence intervals. In fact, a closer look at the LRG’s PIP correlation monopole in figure 13 reveals indications of power suppression at this scale, roughly of the order of σ . Setting aside the definitional issue mentioned earlier, the combination of PIP and angular upweighting consistently produces unbiased estimates for all the FFA samples.

The direct comparisons between figures 11 and 13 and between figures 12 and 14 provide valuable insights into the reliability of the two assignment techniques we developed. Overall, we observe good agreement between the more realistic altMTL method and the computationally simpler FFA emulator. The most significant difference is the reduced impact of the collision window in the FFA case, though it remains evident for the fiducial weights. Additionally, due to a significantly lower number of zero-probability pairs, PIP weights alone are

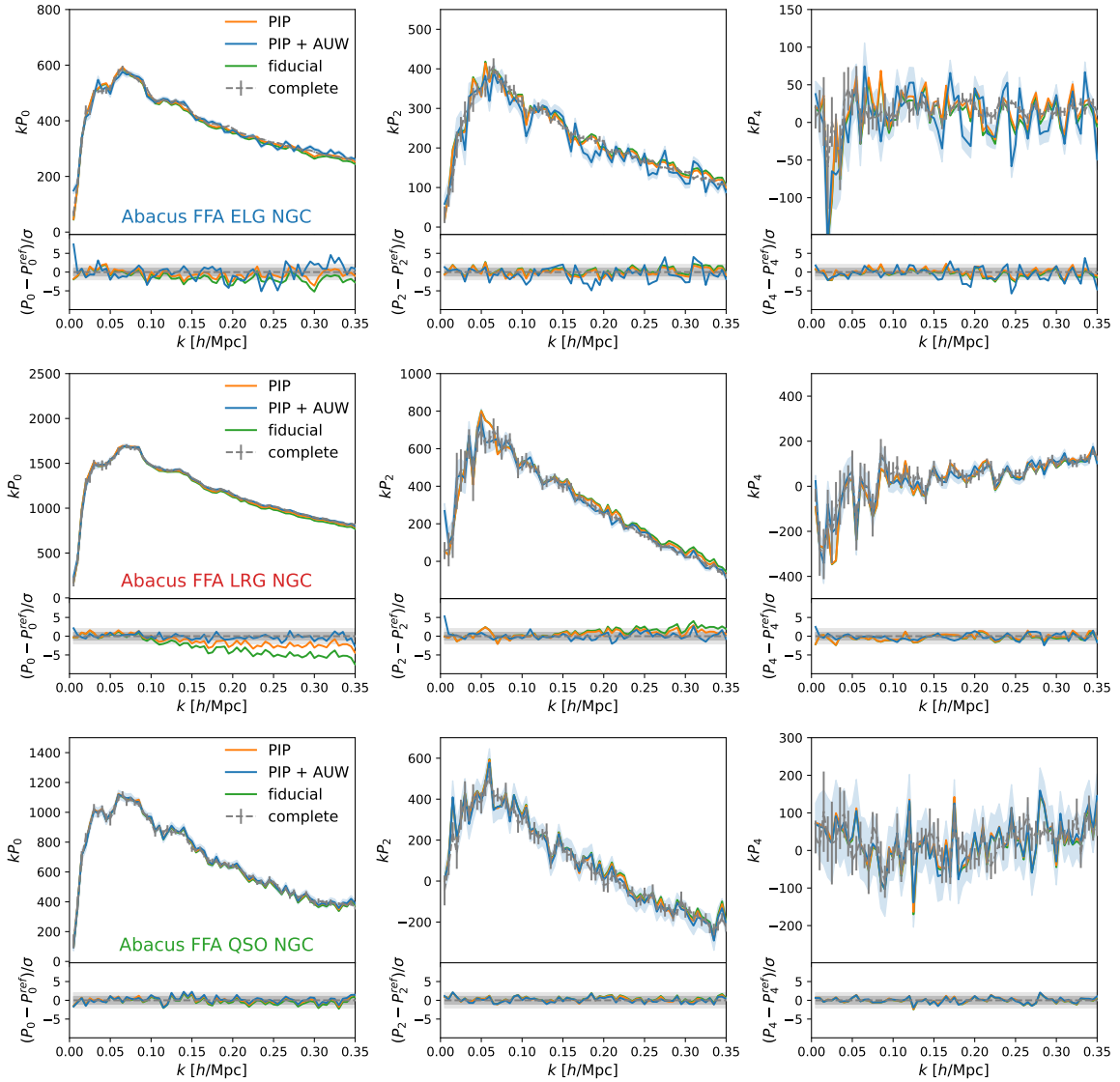


Figure 14. Same as figure 12 but for the FFA mocks (mock 11, NGC). Unlike altMTL mocks, the fiducial weights here (solid green) are, by design, IIP weights.

(almost) sufficient to restore the “true” clustering without the need for angular upweighting. Both the effects are not surprising, given how the collision window is handled by the FFA algorithm (section 3.2), and they are perfectly consistent with the behaviour of the angular pair counts shown in figure 9. Possible improvements to enhance the realism of the FFA algorithm involve optimizing its hyperparameters, adapting it for multi-tracer emulation, and developing strategies to actively manipulate the shape of the collision window. We leave these investigations to future work.

Finally, figures 15 and 16 display the multipoles of the correlation function and power spectrum, respectively, as measured directly from the DESI DR1 samples. These calculations were performed using the same settings as for the altMTL mocks, and the same color coding conventions are followed. To the best of our knowledge, this is the first time PIP weights have been applied to measure the power spectrum from real data. Although we lack a “true”

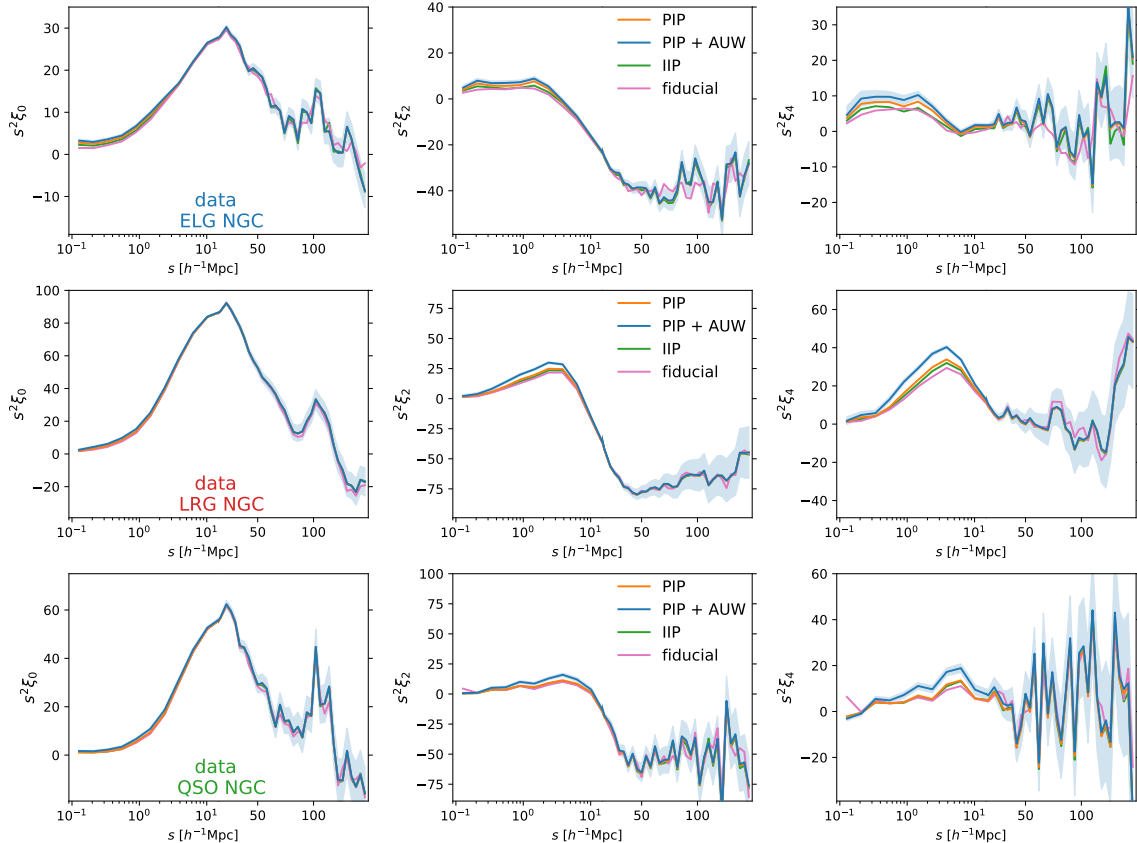


Figure 15. Same as figure 11 but applied to the DESI DR1 catalogues, with the key difference that no reference sample is available for comparison in this case.

sample for comparison, it is evident that the behavior of the various mitigation schemes aligns well with the altMTL results and, more broadly, with expectations. On one hand, this further reinforces the validity of altMTLs, and to a lesser extent, FFA, as realistic fiber assignment methods. On the other hand, it confirms the effectiveness of the mitigation techniques, or at least their consistency across different datasets, with no unexpected trends emerging.

8 Conclusions

We have presented a detailed analysis of the impact of fiber assignment incompleteness on 2-point statistics for DESI DR1, discussed the countermeasures implemented to address it, and described the methods developed to simulate the fiber assignment process on synthetic galaxy catalogs.

This work is part of a larger effort to characterize the DESI fiber incompleteness, with one of its main goals being to provide a coherent overview of this effort. For a more detailed description of the altMTL assignment method and the θ -cut fiber mitigation, we refer the reader to [47] and [30], respectively, whereas for a comprehensive explanation of the process used to build the catalogues and obtain the 2-point statistics for DR1 please see [46].

We have conducted systematic studies on the impact of fiber assignment on the clustering of all the three DESI dark-time tracers, ELGs, LRGs, QSOs, across the entire DR1 volume, using both real and simulated data. As expected, we have found clear evidence

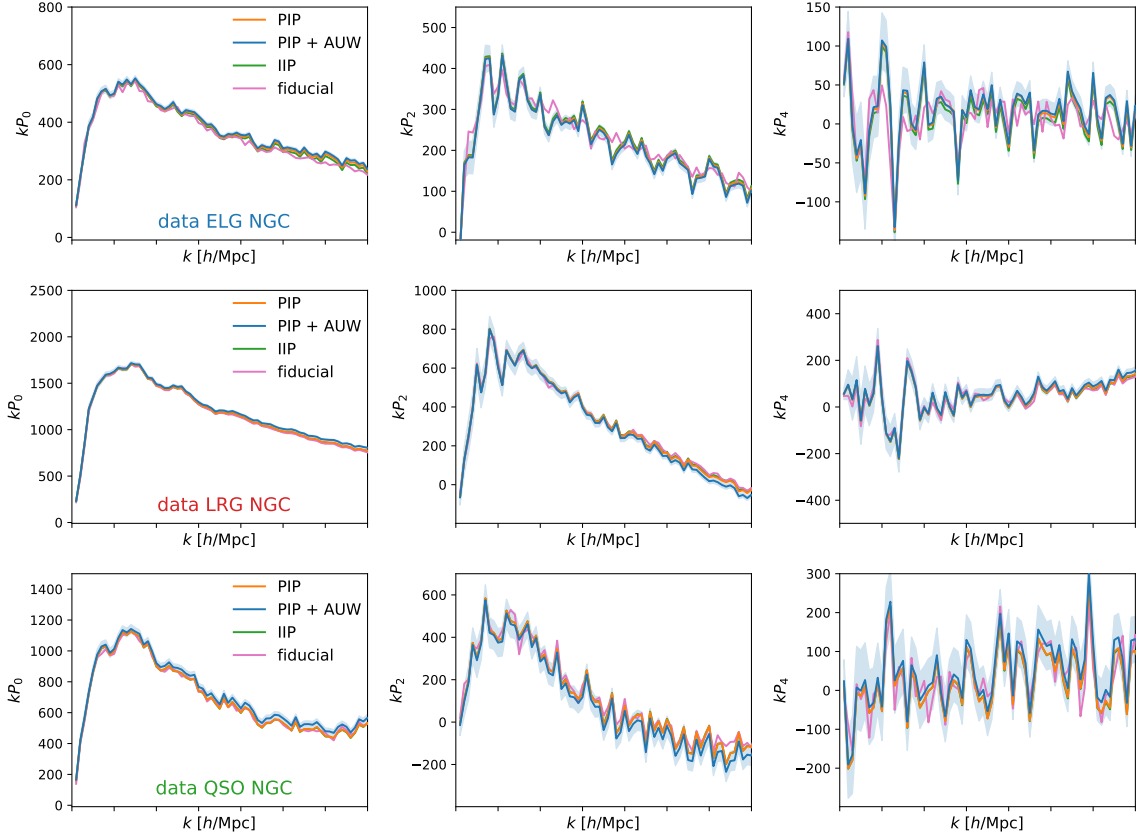


Figure 16. Same as figure 12 but applied to the DESI DR1 catalogues, with the key difference that no reference sample is available for comparison in this case.

of fiber incompleteness patterns in the DESI DR1 footprint and, most importantly, in the 2-point clustering. These effects are successfully removed on large and intermediate scale with the use of proper weights, applied to the galaxy and random samples. However, these weights alone do not address those aspects of fiber assignment that are intrinsically pairwise in nature. As a consequence, the angular pair counts show a large power suppression on scales $\lesssim 0.02\text{deg}$ for all the tracer types and sky areas considered, reaching $\sim 80\%$ for the ELGs in south Galactic cap. Properly addressing the impact of this “collision window” on power spectrum and correlation function estimates is the primary goal of the fiber mitigation techniques presented in this work.

We have explained what concepts motivate the fast-fiber-assignment procedure (FFA) and how the algorithm is practically implemented. We have tested its performance against altMTL mocks and real data, finding reasonably good agreement considering the approximations involved. The most notable differences we observed are the reduced number of zero-probability pairs and a less pronounced collision window, both of which are fully explained by the FFA algorithm’s procedures for assigning probabilities and imposing small-scale anticorrelation. The primary reason for developing FFA was the necessity to run fiber assignment on thousands EZmocks to extract realistic covariance matrices for cosmological inference. This is how we obtained the power-spectrum covariance used for the DR1 full-shape analysis, [31, 32]. However, as discussed in [46] and [69], we have indications that this covariance

might have roughly 10% less power than it should. We therefore applied a final rescaling in amplitude, as described in [46]. FFA is a potential reason for this missing variance, alongside with pre-existing discrepancies in the 3- and 4-point functions. These aspects will be further investigated in future data releases.

We have assessed the reliability of the altMTL mocks, which are more realistic but computationally demanding. They have successfully passed all tests against the data, with no apparent signs of any issues. For future data releases, we are looking into ways to accelerate the altMTL process and enhancing the accuracy of FFA by incorporating more realism into the procedure, such as integrating a multi-tracer kernel [70].

We have conducted a comprehensive analysis of the different fiber-mitigation techniques, including fiducial weights derived by matching the local completeness of different tracer types, individual inverse probability (IIP) weights, pairwise inverse probability (PIP) weights, and their integration with angular upweighting. To achieve the high level of accuracy required by DESI and address limitations like the limited number of targeting realizations and the significant incompleteness in DR1, we have used inverse probability weights in a novel manner. In particular, we have introduced a new definition for the PIP weights that enhances their robustness against scale-dependent variations in assignment probability. Additionally, we have defined a new set of weights, w_{NTMP} , to correct for the irregular distribution of the numerous zero-probability galaxies found in DR1.

These new developments allowed us not only to obtain reliable estimates of the correlation function down to the smallest scale for DR1 but also to successfully apply, for the first time ever, PIP weights (plus angular upweighting) to power spectrum measurements from real data. Since these innovations, especially the w_{NTMP} weights, were finalized only recently, this is the only DR1 paper where they have been utilized.

In summary, we found that PIP weights with angular upweighting deliver the most accurate estimates of the “true” clustering, in all scenarios considered, both in configuration and Fourier space, and for all types of tracers. These results complement those in a parallel paper, [30], which shows that similar performances can be achieved with completeness weights, provided that small scales are removed from both the estimator and the model of the statistics of interest (θ -cut method).

So far, PIP weights have been employed in various analyses of DESI data, from both SV3 [71, 72] and DR1, to recover the correct small-scale clustering (e.g., [73–77]). For the full-shape analysis, [31, 32], the completeness weights with θ -cut method was chosen, whereas the BAO analysis [78, 79] relied solely on completeness weight, as it was shown that no further correction was necessary. However, for those studies that, like BAO, depend on reconstruction, the θ -cut approach is less effective, due to the transfer of small-scale incompleteness to larger scale. This work opens the opportunity to use PIP weights with angular upweighting for all these analyses in future data releases, thus allowing for the preservation of small-scale information and providing a more clearly defined framework for reconstruction.

A Adding small-scale anticorrelation to the FFA catalogues

Once a group of entangled galaxies has been identified through the FOF algorithm, we set it into a maximum anticorrelation state, by manipulating the corresponding bitweights. Specifically, we allow displacements of bits from one realisation to one another, inside each bitweight, but not from one bitweight to one another. As a consequence, the procedure conserves individual probabilities while arbitrarily changing the pairwise ones. By maximum

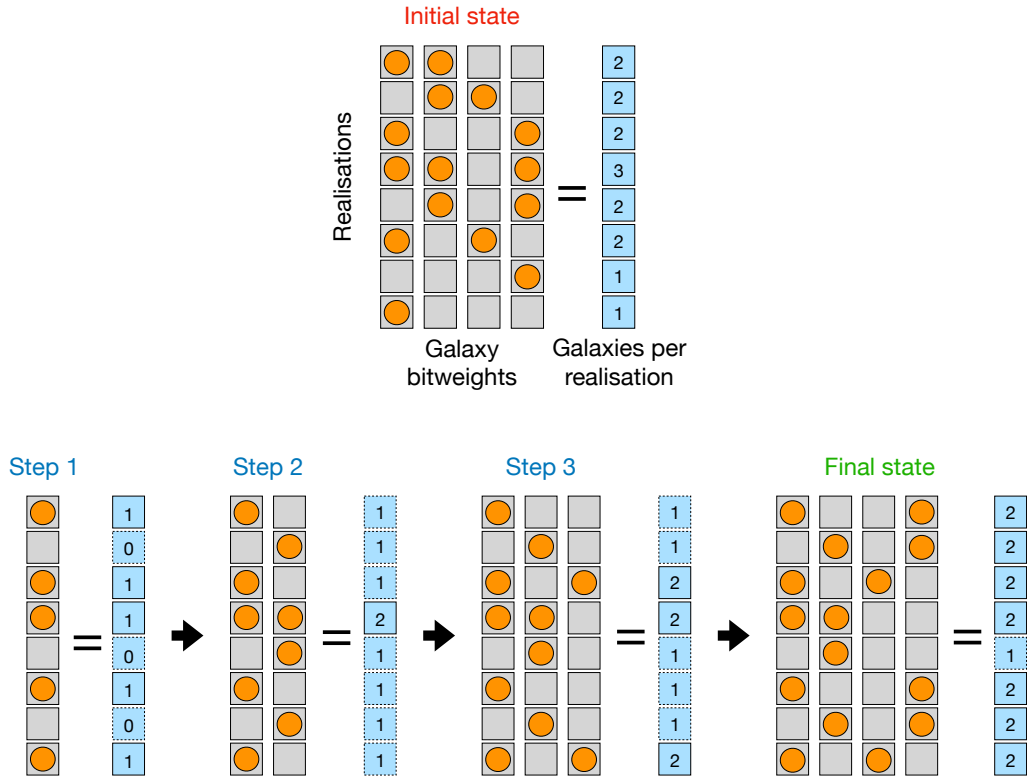


Figure 17. Scheme of the procedure used to find a maximum anticorrelation state for the bitweights, as described in the text.

anticorrelation state we mean bitweights such that the corresponding pair count (and any t -plet count with $t \geq 2$), averaged over all the realisations, is the lowest compatible with the given numbers of bits per bitweight or, in other words, compatible with the given individual probabilities.

Two preliminary considerations: (i) for a given group there are many of such states; (ii) all galaxies in a FOF group are treated equally, i.e. we ignore their spatial locations and relative separations. Regarding (i), our procedure is such that the maximum anticorrelation state is fully determined by the order with which the galaxies are processed, which can be easily randomised to obtain alternative states. Regarding (ii), it would be technically possible to add scale dependency to the procedure but, realistically, that would come at the cost of more computational complexity/time, which is precisely what we want to avoid with FFA.

The actual algorithm works as sketched in figure 17. In brief, given a group of N_{FOF} ordered galaxies, each with its own bitweight, we initialise the process with galaxy 1, for which we count the total number of bits (i.e. galaxies) in each realisation. Clearly, such number can only be 1 or 0 at this stage. We then move to galaxy 2 and permute its bits randomly, but giving priority to the “holes” left by galaxy 1, i.e. those realisations with no galaxies yet. We then iterate the process to all the remaining $N_{FOF} - 2$ galaxies. With this approach the sum of galaxies per realisation at each step will be either a_n or $a_n - 1$, where a_n is an integer which varies with the iteration number n . What we call “holes” at the n -th

step are those realisations with $a_n - 1$ total bits.

Since the number of pairs in each realisation grows with the square of the number of objects and since the number of objects is non negative, it follows that no further bit permutation of the so obtained final state can decrease the total number of pairs (and t -plets with $t \geq 2$), i.e. it is a maximum anticorrelation state.

B Collision kernel from the mocks as a mitigation strategy

The method described here builds on the individual weighting schemes discussed in section 5, either w_{comp} or w_{IIP} , for a first-stage correction. When these weights are applied, the residual bias is of pure pairwise nature and confined to the $\theta < \theta_{ind}$ separations, inside the collision window. As described throughout the paper, PIP weights statistically extract this window from the data, whereas θ -cut removes it and adjusts the theory model accordingly.

A third option consists of assuming a given shape for the window, e.g., a top hat function [15], or the actual shape measured from a set of high-fidelity mock catalogues. This shape can be used inside either a model-stage or an estimator-stage mitigation. Here we discuss the latter option, with shape measured from the mocks.

We are not interested in applying this method to configuration-space statistics, as the net effect would be to leave the large-scale correlation of the data virtually unchanged, while making mock-dependent the small-scale one, with little control on the correspondent error. Instead, we want to use it as way to compensate for the large-scale ($k \lesssim 0.2$) tails of the collision window in Fourier space.

With the expression “measuring the window” we mean that we can measure from mocks the following collision kernel,

$$\mathcal{K}(s_{\parallel}, s_{\perp}) = \frac{DD_{complete}(s_{\parallel}, s_{\perp})}{DD_{IIP}(s_{\parallel}, s_{\perp})} - 1 , \quad (\text{B.1})$$

where $DD_{complete}$ represents the pair counts from the complete sample and DD_{IIP} the pair counts computed with IIP weights from the fiber-assigned sample, as function of the parallel and perpendicular separations, s_{\parallel} and s_{\perp} . An equivalent expression can be written for the fiducial weights. We can then measure the data power spectrum using the estimator defined in eq. (5.6), but with

$$A_{ij} = w_i^{IIP} w_j^{IIP} \mathcal{K}(s_{\parallel}^{ij}, s_{\perp}^{ij}) . \quad (\text{B.2})$$

In the idealised case in which the mocks are a perfect representation of the (unknown) clustering of the data, this approach returns the exact pairwise correction. In a more realistic scenario, the unavoidable discrepancies between mocks and data are mitigated by the differential nature of the \mathcal{K} kernel. In other words, \mathcal{K} can be seen as an impulse-response device, designed to capture the effect of fiber assignment alone, regardless of the underlying clustering. More precisely, this is a first order description of such a device, as the true targeting mechanism cannot be fully captured by a multiplicative factor: the fraction of observed pairs must drop for increasing number density and eventually saturate. Lastly, despite the fact that formally the kernel depends on the s_{\parallel} variable, its variability along the parallel direction turned out to be negligible. We therefore decided to integrate s_{\parallel} and keep only the s_{\perp} dependence, as a way to compress the information and minimize the noise.

Figure 18 shows the LRG power spectrum multipoles measured from the altMTL mocks using eq. B.2. We considered both fiducial and IIP weights as the individual weights required

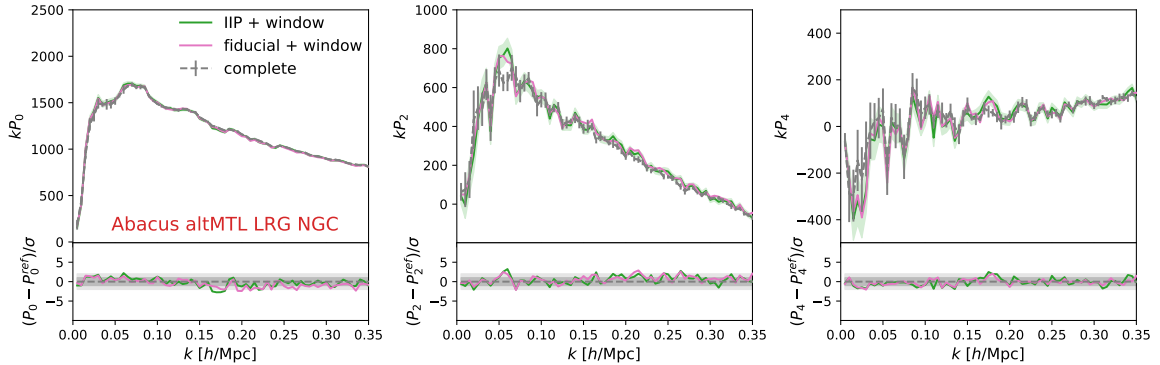


Figure 18. Multipoles of the power spectrum P_ℓ as a function of the wave number k , measured from the altMTL mocks (mock 11, LRGs NGC). For visualisation purpose, we follow the common practice of multiplying the amplitude of the multipoles by k . The gray dashed curves correspond to measurements obtained from the complete sample with no additional weights and error bars indicating the scatter between different mocks. They represent a reference target for the other two measurements, obtained from the fiber-assigned samples using the transverse collision window coupled to fiducial weights (solid pink) and IIP weights (solid green with shaded green area for the corresponding 1σ error), as described in the text. In the insets at the bottom of each panel we show the difference between fiber-assigned and complete samples in units of standard deviation.

by the equation, represented by the pink and green solid curves, respectively. Since the complete and clustering samples do not share the same mask we have normalised the pair counts of eq. B.1 with those of the corresponding random samples. For this test the \mathcal{K} function is derived from the same mock used to measure the power spectrum. By doing so we are unrealistically providing the estimator with nearly all the missing information, except for the integration along s_{\parallel} . The results displayed in the figure should therefore be viewed as proof of concept rather than a definitive test. In a more realistic scenario one should take the average of \mathcal{K} measured over the 25 altMTL mocks and apply it to the data. That said, the figure shows the effectiveness of this mitigation strategy in recovering the “true” clustering, with a minor preference for IIP weights. As discussed in the text, we eventually opted for more agnostic approaches, but this is a valid way to restore the correct amplitude of P on large and intermediate scale, as far as the mock catalogues provide a good representation of fiber assignment process and underlying clustering.

C Clustering measurements averaged over multiple FFA mocks

Figure 19 shows the monopole of the power spectrum averaged over 7 FFA mocks, with the variance rescaled accordingly. The trends discussed in section 7 are confirmed and, to some extent, enhanced. For instance, the systematic suppression of the power spectrum measured from the ELG sample with fiducial weights emerges more clearly. As previously noted, with FFA mocks we could not apply angular upweighting in a fully consistent way because the catalogs lack some of the necessary components. This is at the origin of the low k peak observed in the bottom insets of both the ELG and, to a lesser extent, the LRG samples. It also explains why the estimates obtained with PIP plus angular upweighting are noisier than the others.

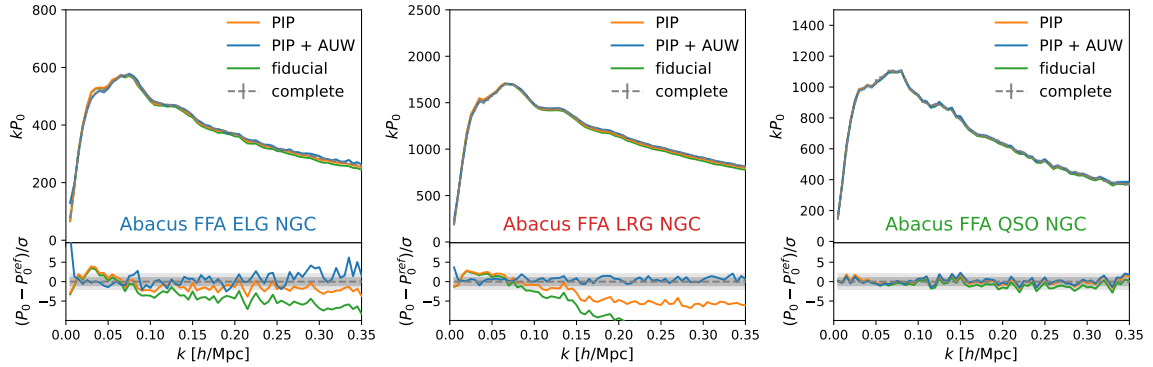


Figure 19. Monopole of the ELG, LRG, and QSO power spectrum averaged over 7 FFA mocks (mocks 11-17, NGC). The grey dashed curves indicate the measurements obtained from the complete sample with no additional weights and error bars indicating the scatter between different mocks. The remaining curves are obtained from the fiber-assigned samples using: fiducial (IIP) weights (solid green); PIP weights (solid orange); PIP weights plus angular upweighting (solid blue), with the blue shaded area indicating the corresponding 1σ error. The bottom insets show the difference between fiber-assigned and complete samples in units of the standard deviation of the mean.

Data availability

Data from the plots in this paper will be available on Zenodo as part of DESI’s Data Management Plan. The data used in this analysis will be made public along the Data Release (details in <https://data.desi.lbl.gov/doc/releases/>).

Acknowledgments

This material is based upon work supported by the U.S. Department of Energy (DOE), Office of Science, Office of High-Energy Physics, under Contract No. DE-AC02-05CH11231, and by the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility under the same contract. Additional support for DESI was provided by the U.S. National Science Foundation (NSF), Division of Astronomical Sciences under Contract No. AST-0950945 to the NSF’s National Optical-Infrared Astronomy Research Laboratory; the Science and Technology Facilities Council of the United Kingdom; the Gordon and Betty Moore Foundation; the Heising-Simons Foundation; the French Alternative Energies and Atomic Energy Commission (CEA); the National Council of Humanities, Science and Technology of Mexico (CONAHCYT); the Ministry of Science, Innovation and Universities of Spain (MICIU/AEI/10.13039/501100011033), and by the DESI Member Institutions: <https://www.desi.lbl.gov/collaborating-institutions>. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the U. S. National Science Foundation, the U. S. Department of Energy, or any of the listed funding agencies.

The authors are honored to be permitted to conduct scientific research on Iolkam Du’ag (Kitt Peak), a mountain with particular significance to the Tohono O’odham Nation.

References

- [1] SDSS collaboration, *The Sloan Digital Sky Survey: Technical Summary*, *Astron. J.* **120** (2000) 1579 [[astro-ph/0006396](#)].
- [2] K.S. Dawson, D.J. Schlegel, C.P. Ahn, S.F. Anderson, É. Aubourg, S. Bailey et al., *The Baryon Oscillation Spectroscopic Survey of SDSS-III*, *AJ* **145** (2013) 10 [[1208.0022](#)].
- [3] K.S. Dawson, J.-P. Kneib, W.J. Percival, S. Alam, F.D. Albareti, S.F. Anderson et al., *The SDSS-IV Extended Baryon Oscillation Spectroscopic Survey: Overview and Early Data*, *AJ* **151** (2016) 44 [[1508.04473](#)].
- [4] M. Levi, C. Bebek, T. Beers, R. Blum, R. Cahn, D. Eisenstein et al., *The DESI Experiment, a whitepaper for Snowmass 2013*, *arXiv e-prints* (2013) arXiv:1308.0847 [[1308.0847](#)].
- [5] DESI Collaboration, B. Abareshi, J. Aguilar, S. Ahlen, S. Alam, D.M. Alexander et al., *Overview of the Instrumentation for the Dark Energy Spectroscopic Instrument*, *AJ* **164** (2022) 207 [[2205.10939](#)].
- [6] DESI Collaboration, A. Aghamousa, J. Aguilar, S. Ahlen, S. Alam, L.E. Allen et al., *The DESI Experiment Part I: Science, Targeting, and Survey Design*, *arXiv e-prints* (2016) arXiv:1611.00036 [[1611.00036](#)].
- [7] J.H. Silber, P. Fagrelus, K. Fanning, M. Schubnell, J.N. Aguilar, S. Ahlen et al., *The Robotic Multiobject Focal Plane System of the Dark Energy Spectroscopic Instrument (DESI)*, *AJ* **165** (2023) 9 [[2205.09014](#)].
- [8] T.N. Miller, P. Doel, G. Gutierrez, R. Besuner, D. Brooks, G. Gallo et al., *The Optical Corrector for the Dark Energy Spectroscopic Instrument*, *AJ* **168** (2024) 95 [[2306.06310](#)].
- [9] J. Guy, S. Bailey, A. Kremin, S. Alam, D.M. Alexander, C. Allende Prieto et al., *The Spectroscopic Data Processing Pipeline for the Dark Energy Spectroscopic Instrument*, *AJ* **165** (2023) 144 [[2209.14482](#)].
- [10] E.F. Schlafly, D. Kirkby, D.J. Schlegel, A.D. Myers, A. Raichoor, K. Dawson et al., *Survey Operations for the Dark Energy Spectroscopic Instrument*, *AJ* **166** (2023) 259 [[2306.06309](#)].
- [11] E. Hawkins, S. Maddox, S. Cole, O. Lahav, D.S. Madgwick, P. Norberg et al., *The 2dF Galaxy Redshift Survey: correlation functions, peculiar velocities and the matter density of the Universe*, *MNRAS* **346** (2003) 78 [[astro-ph/0212375](#)].
- [12] H.K. Herrera-Alcantar, A. Muñoz-Gutiérrez, T. Tan, A.X. González-Morales, A. Font-Ribera, J. Guy et al., *Synthetic spectra for Lyman- α forest analysis in the Dark Energy Spectroscopic Instrument*, *arXiv e-prints* (2023) arXiv:2401.00303 [[2401.00303](#)].
- [13] H. Guo, I. Zehavi and Z. Zheng, *A New Method to Correct for Fiber Collisions in Galaxy Two-point Statistics*, *ApJ* **756** (2012) 127 [[1111.6598](#)].
- [14] L. Pinol, R.N. Cahn, N. Hand, U. Seljak and M. White, *Imprint of DESI fiber assignment on the anisotropic power spectrum of emission line galaxies*, *J. Cosmology Astropart. Phys.* **4** (2017) 008 [[1611.05007](#)].
- [15] C. Hahn, R. Scoccimarro, M.R. Blanton, J.L. Tinker and S.A. Rodríguez-Torres, *The Effect of Fiber Collisions on the Galaxy Power Spectrum Multipoles*, *MNRAS* **467** (2017) 1940 [[1609.01714](#)].
- [16] D. Bianchi and W.J. Percival, *Unbiased clustering estimation in the presence of missing observations*, *MNRAS* **472** (2017) 1106 [[1703.02070](#)].
- [17] D. Bianchi and L. Verde, *Confronting missing observations with probability weights: Fourier space and generalized formalism*, *MNRAS* **495** (2020) 1511 [[1912.08803](#)].

- [18] B.A. Reid, H.-J. Seo, A. Leauthaud, J.L. Tinker and M. White, *A 2.5 per cent measurement of the growth rate from small-scale redshift space clustering of SDSS-III CMASS galaxies*, *MNRAS* **444** (2014) 476 [[1404.3742](#)].
- [19] A. Burden, N. Padmanabhan, R.N. Cahn, M.J. White and L. Samushia, *Mitigating the impact of the DESI fiber assignment on galaxy clustering*, *J. Cosmology Astropart. Phys.* **2017** (2017) 001 [[1611.04635](#)].
- [20] R. Paviot, S. de la Torre, A. de Mattia, C. Zhao, J. Bautista, E. Burtin et al., *Angular systematics-free cosmological analysis of galaxy clustering in configuration space*, *MNRAS* **512** (2022) 1341 [[2110.10184](#)].
- [21] L. Anderson, E. Aubourg, S. Bailey, D. Bizyaev, M. Blanton, A.S. Bolton et al., *The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the Data Release 9 spectroscopic galaxy sample*, *MNRAS* **427** (2012) 3435 [[1203.6594](#)].
- [22] B. Reid et al., *SDSS-III Baryon Oscillation Spectroscopic Survey Data Release 12: galaxy target selection and large scale structure catalogues*, *Mon. Not. Roy. Astron. Soc.* **455** (2016) 1553 [[1509.06529](#)].
- [23] W.J. Percival and D. Bianchi, *Using angular pair upweighting to improve 3D clustering measurements*, *MNRAS* **472** (2017) L40 [[1703.02071](#)].
- [24] F.G. Mohammad, D. Bianchi, W.J. Percival, S. de la Torre, L. Guzzo, B.R. Granett et al., *The VIMOS Public Extragalactic Redshift Survey (VIPERS). Unbiased clustering estimate with VIPERS slit assignment*, *A&A* **619** (2018) A17 [[1807.05999](#)].
- [25] F.G. Mohammad, W.J. Percival, H.-J. Seo, M.J. Chapman, D. Bianchi, A.J. Ross et al., *The completed SDSS-IV extended baryon oscillation spectroscopic survey: pairwise-inverse probability and angular correction for fibre collisions in clustering measurements*, *MNRAS* **498** (2020) 128 [[2007.09005](#)].
- [26] D. Bianchi, A. Burden, W.J. Percival, D. Brooks, R.N. Cahn, J.E. Forero-Romero et al., *Unbiased clustering estimates with the DESI fibre assignment*, *MNRAS* **481** (2018) 2338 [[1805.00951](#)].
- [27] A. Smith, J.-h. He, S. Cole, L. Stothert, P. Norberg, C. Baugh et al., *Correcting for fibre assignment incompleteness in the DESI Bright Galaxy Survey*, *MNRAS* **484** (2019) 1285 [[1809.07355](#)].
- [28] F.G. Mohammad, S. de la Torre, D. Bianchi, L. Guzzo and J.A. Peacock, *Group-galaxy correlations in redshift space as a probe of the growth of structure*, *MNRAS* **458** (2016) 1948 [[1502.05045](#)].
- [29] P. Zarrouk, E. Burtin, H. Gil-Marín, A.J. Ross, R. Tojeiro, I. Pâris et al., *The clustering of the SDSS-IV extended Baryon Oscillation Spectroscopic Survey DR14 quasar sample: measurement of the growth rate of structure from the anisotropic correlation function between redshift 0.8 and 2.2*, *MNRAS* **477** (2018) 1639 [[1801.03062](#)].
- [30] M. Pinon, A. de Mattia, P. McDonald, E. Burtin, V. Ruhlmann-Kleider, M. White et al., *Mitigation of DESI fiber assignment incompleteness effect on two-point clustering with small angular scale truncated estimators*, *arXiv e-prints* (2024) [arXiv:2406.04804](#) [[2406.04804](#)].
- [31] DESI Collaboration, *DESI 2024 V: Analysis of the full shape of two-point clustering statistics from galaxies and quasars, in preparation* (2024) .
- [32] DESI Collaboration, *DESI 2024 VII: Cosmological constraints from full-shape analyses of the two-point clustering statistics measurements, in preparation* (2024) .
- [33] J.H. Silber, P. Fagrelus, K. Fanning, M. Schubnell, J.N. Aguilar, S. Ahlen et al., *The robotic*

- multiobject focal plane system of the dark energy spectroscopic instrument (DESI), *The Astronomical Journal* **165** (2022) 9.
- [34] Raichoor et al. in, *Dark Energy Spectroscopic Instrument Fiber Assignment, in preparation* (2024) .
- [35] A.D. Myers, J. Moustakas, S. Bailey, B.A. Weaver, A.P. Cooper, J.E. Forero-Romero et al., *The Target-selection Pipeline for the Dark Energy Spectroscopic Instrument*, *AJ* **165** (2023) 50 [2208.08518].
- [36] DESI Collaboration, *DESI 2024 I: Data Release 1 of the Dark Energy Spectroscopic Instrument, in preparation* (2025) .
- [37] Bailey et al., *in preparation* (2024) .
- [38] A. Raichoor, J. Moustakas, J.A. Newman, T. Karim, S. Ahlen, S. Alam et al., *Target Selection and Validation of DESI Emission Line Galaxies*, *AJ* **165** (2023) 126 [2208.08513].
- [39] R. Zhou, B. Dey, J.A. Newman, D.J. Eisenstein, K. Dawson, S. Bailey et al., *Target Selection and Validation of DESI Luminous Red Galaxies*, *AJ* **165** (2023) 58 [2208.08515].
- [40] E. Chaussidon, C. Yèche, N. Palanque-Delabrouille, D.M. Alexander, J. Yang, S. Ahlen et al., *Target Selection and Validation of DESI Quasars*, *ApJ* **944** (2023) 107 [2208.08511].
- [41] C. Hahn, M.J. Wilson, O. Ruiz-Macias, S. Cole, D.H. Weinberg, J. Moustakas et al., *The DESI Bright Galaxy Survey: Final Target Selection, Design, and Validation*, *AJ* **165** (2023) 253 [2208.08512].
- [42] A. Dey, D.J. Schlegel, D. Lang, R. Blum, K. Burleigh, X. Fan et al., *Overview of the DESI Legacy Imaging Surveys*, *AJ* **157** (2019) 168 [1804.08657].
- [43] Schlegel et al., *in preparation* (2024) .
- [44] DESI Collaboration, A.G. Adame, J. Aguilar, S. Ahlen, S. Alam, D.M. Alexander et al., *DESI 2024 IV: Baryon Acoustic Oscillations from the Lyman Alpha Forest*, *arXiv e-prints* (2024) arXiv:2404.03001 [2404.03001].
- [45] E. Chaussidon, *Constraining the local primordial non-gaussianity via the large scale-dependent bias with the DESI DR1 LRG and QSO, in preparation* (2024) .
- [46] DESI Collaboration, *DESI 2024 II: Sample definitions, characteristics and two-point clustering statistics, in preparation* (2024) .
- [47] J. Lasker, A.C. Rosell, A.D. Myers, A.J. Ross, D. Bianchi, M.M.S. Hanif et al., *Production of Alternate Realizations of DESI Fiber Assignment for Unbiased Clustering Measurement in Data and Simulations*, *arXiv e-prints* (2024) arXiv:2404.03006 [2404.03006].
- [48] C. Zhao et al., *Mock catalogues with survey realism for the DESI DR1, in preparation* (2024) .
- [49] N.A. Maksimova, L.H. Garrison, D.J. Eisenstein et al., *AbacusSummit: a massive set of high-accuracy, high-resolution N-body simulations*, *Monthly Notices of the Royal Astronomical Society* **508** (2021) 4017 [https://academic.oup.com/mnras/article-pdf/508/3/4017/40811763/stab2484.pdf].
- [50] L.H. Garrison, D.J. Eisenstein, D. Ferrer et al., *The abacus cosmological N-body code*, *Monthly Notices of the Royal Astronomical Society* **508** (2021) 575 [https://academic.oup.com/mnras/article-pdf/508/1/575/40458823/stab2482.pdf].
- [51] C.-H. Chuang, F.-S. Kitaura, F. Prada, C. Zhao and G. Yepes, *EZmocks: extending the Zel'dovich approximation to generate mock galaxy catalogues with accurate clustering statistics*, *MNRAS* **446** (2015) 2621 [1409.1124].
- [52] S. Yuan, H. Zhang, A.J. Ross et al., *The DESI One-Percent Survey: Exploring the Halo Occupation Distribution of Luminous Red Galaxies and Quasi-Stellar Objects with AbacusSummit*, *arXiv e-prints* (2023) arXiv:2306.06314 [2306.06314].

- [53] A. Rocher, V. Ruhlmann-Kleider, E. Burtin et al., *The desi one-percent survey: exploring the halo occupation distribution of emission line galaxies with abacussummit simulations*, *Journal of Cosmology and Astroparticle Physics* **2023** (2023) 016.
- [54] S.D. Landy and A.S. Szalay, *Bias and Variance of Angular Correlation Functions*, *ApJ* **412** (1993) 64.
- [55] D. Bianchi, H. Gil-Marín, R. Ruggeri and W.J. Percival, *Measuring line-of-sight-dependent Fourier-space clustering using FFTs*, *MNRAS* **453** (2015) L11 [1505.05341].
- [56] R. Scoccimarro, *Fast estimators for redshift-space clustering*, *Phys. Rev. D* **92** (2015) 083532 [1506.02729].
- [57] N. Hand, Y. Feng, F. Beutler, Y. Li, C. Modi, U. Seljak et al., *nbodykit: An Open-source, Massively Parallel Toolkit for Large-scale Structure*, *AJ* **156** (2018) 160 [1712.05834].
- [58] K. Yamamoto, M. Nakamichi, A. Kamino, B.A. Bassett and H. Nishioka, *A Measurement of the Quadrupole Power Spectrum in the Clustering of the 2dF QSO Survey*, *Publications of the Astronomical Society of Japan* **58** (2006) 93.
- [59] A.J. Ross, J. Aguilar, S. Ahlen, S. Alam, A. Anand, S. Bailey et al., *The Construction of Large-scale Structure Catalogs for the Dark Energy Spectroscopic Instrument*, *arXiv e-prints* (2024) arXiv:2405.16593 [2405.16593].
- [60] M.J. Wilson, J.A. Peacock, A.N. Taylor and S. de la Torre, *Rapid modelling of the redshift-space power spectrum multipoles for a masked density field*, *MNRAS* **464** (2017) 3121 [1511.07799].
- [61] F. Beutler and P. McDonald, *Unified galaxy power spectrum measurements from 6dFGS, BOSS, and eBOSS*, *J. Cosmology Astropart. Phys.* **2021** (2021) 031 [2106.06324].
- [62] E. Castorina and M. White, *Beyond the plane-parallel approximation for redshift surveys*, *MNRAS* **476** (2018) 4403 [1709.09730].
- [63] P. Mc Donald, , *in preparation* (2024) .
- [64] E. Keihänen, H. Kurki-Suonio, V. Lindholm, A. Viitanen, A.S. Suur-Uski, V. Allevato et al., *Estimating the galaxy two-point correlation function using a split random catalog*, *A&A* **631** (2019) A73 [1905.01133].
- [65] Y.P. Jing, *Correcting for the Alias Effect When Measuring the Power Spectrum Using a Fast Fourier Transform*, *ApJ* **620** (2005) 559 [astro-ph/0409240].
- [66] E. Sefusatti, M. Crocce, R. Scoccimarro and H.M.P. Couchman, *Accurate estimators of correlation functions in Fourier space*, *MNRAS* **460** (2016) 3624 [1512.07295].
- [67] H.A. Feldman, N. Kaiser and J.A. Peacock, *Power spectrum analysis of three-dimensional redshift surveys*, *Astrophys. J.* **426** (1994) 23 [astro-ph/9304022].
- [68] M. Rashkovetskyi, D. Forero-Sánchez, A. de Mattia, D.J. Eisenstein, N. Padmanabhan, H. Seo et al., *Semi-analytical covariance matrices for two-point correlation function for DESI 2024 data*, *arXiv e-prints* (2024) arXiv:2404.03007 [2404.03007].
- [69] D. Forero-Sanchez et al., *Analytical and EZmock covariance validation for the DESI 2024 results*, *in preparation* (2024) .
- [70] M. M. S Hanif et al., *Fast Fiber Assign: Emulating fiber assignment effects for realistic DESI catalogs*, *in preparation* (2024) .
- [71] DESI Collaboration, A.G. Adame, J. Aguilar, S. Ahlen, S. Alam, G. Aldering et al., *Validation of the Scientific Program for the Dark Energy Spectroscopic Instrument*, *AJ* **167** (2024) 62 [2306.06307].
- [72] DESI Collaboration, A.G. Adame, J. Aguilar, S. Ahlen, S. Alam, G. Aldering et al., *The Early Data Release of the Dark Energy Spectroscopic Instrument*, *AJ* **168** (2024) 58 [2306.06308].

- [73] A.N. Pearl, A.R. Zentner, J.A. Newman, R. Bezanson, K. Wang, J. Moustakas et al., *The DESI One-percent Survey: Evidence for Assembly Bias from Low-redshift Counts-in-cylinders Measurements*, *ApJ* **963** (2024) 116 [2309.08675].
- [74] A. Rocher, V. Ruhlmann-Kleider, E. Burtin, S. Yuan, A. de Mattia, A.J. Ross et al., *The DESI One-Percent survey: exploring the Halo Occupation Distribution of Emission Line Galaxies with ABACUSSUMMIT simulations*, *J. Cosmology Astropart. Phys.* **2023** (2023) 016 [2306.06319].
- [75] S. Yuan, H. Zhang, A.J. Ross, J. Donald-McCann, B. Hadzhiyska, R.H. Wechsler et al., *The DESI one-per cent survey: exploring the halo occupation distribution of luminous red galaxies and quasi-stellar objects with ABACUSSUMMIT*, *MNRAS* **530** (2024) 947 [2306.06314].
- [76] J. Yu, C. Zhao, V. Gonzalez-Perez, C.-H. Chuang, A. Brodzeller, A. de Mattia et al., *The DESI One-Percent Survey: exploring a generalized SHAM for multiple tracers with the UNIT simulation*, *MNRAS* **527** (2024) 6950 [2306.06313].
- [77] H. Gao, Y.P. Jing, K. Xu, D. Zhao, S. Gui, Y. Zheng et al., *The DESI One-Percent Survey: A Concise Model for the Galactic Conformity of Emission-line Galaxies*, *ApJ* **961** (2024) 74 [2309.03802].
- [78] DESI Collaboration, A.G. Adame, J. Aguilar, S. Ahlen, S. Alam, D.M. Alexander et al., *DESI 2024 III: Baryon Acoustic Oscillations from Galaxies and Quasars*, *arXiv e-prints* (2024) arXiv:2404.03000 [2404.03000].
- [79] DESI Collaboration, A.G. Adame, J. Aguilar, S. Ahlen, S. Alam, D.M. Alexander et al., *DESI 2024 VI: Cosmological Constraints from the Measurements of Baryon Acoustic Oscillations*, *arXiv e-prints* (2024) arXiv:2404.03002 [2404.03002].