

Fermilab


































































Deep Spectroscopy with DESI for Photometric Redshift
Training and Calibration

FERMILAB-PUB-26-0239-PPD

arXiv:2604.06143

This manuscript has been authored by Fermi Forward Discovery Group, LLC
under Contract No. 89243024CSC000002 with the U.S. Department of Energy,
Office of Science, Office of High Energy Physics.

Deep Spectroscopy with DESI for Photometric Redshift Training and Calibration

BIPRATEEP DEY ^{1,2,3,4} JEFFREY A. NEWMAN ⁴ TIANQING ZHANG ⁴ J. AGUILAR ⁵ S. AHLEN ⁶ A. ANAND ⁵
B. ANDREWS ⁷ S. BAILEY ⁵ D. BIANCHI ^{8,9} D. BROOKS ¹⁰ F. J. CASTANDER ^{11,12} T. CLAYBAUGH ⁵
A. CUCEU ⁵ K. S. DAWSON ¹³ A. DE LA MACORRA ¹⁴ J. DELLA COSTA ¹⁵ ARJUN DEY ¹⁵ P. DOEL ¹⁰
S. FERRARO ^{5,16} A. FONT-RIBERA ^{17,18} E. GAZTAÑAGA ^{11,19,12} SATYA GONTCHO A GONTCHO ²⁰
D. GRUEN ^{21,22} G. GUTIERREZ ²³ J. GUY ⁵ H. K. HERRERA-ALCANTAR ^{24,25} K. HONSCHEID ^{26,27}
M. ISHAK ²⁸ R. JOYCE ¹⁵ R. KEHOE ²⁹ D. KIRKBY ³⁰ T. KISNER ⁵ A. KREMIN ⁵ O. LAHAV ¹⁰
M. LANDRIAU ⁵ L. LE GUILLOU ³¹ A. LEAUTHAUD ^{32,33} M. E. LEVI ⁵ M. MANERA ^{34,18} P. MARTINI ^{26,35}
J. MCCULLOUGH ³⁶ A. MEISNER ¹⁵ R. MIQUEL ^{17,18} J. MOUSTAKAS ³⁷ A. D. MYERS ³⁸ J. MYLES ³⁶
S. NADATHUR ¹⁹ N. PALANQUE-DELABROUILLE ^{25,5} W. J. PERCIVAL ^{39,40,41} F. PRADA ⁴² I. PÉREZ-RÀFOLS ⁴³
G. ROSSI ⁴⁴ L. SAMUSHIA ^{45,46,47} E. SANCHEZ ⁴⁸ D. SCHLEGEL ⁵ M. SCHUBNEL ^{49,50} H. SEO ⁵¹
J. SILBER ⁵ D. SPRAYBERRY ¹⁵ G. TARLÉ ⁵⁰ B. A. WEAVER ¹⁵ N. WEAVERDYCK ⁵ R. H. WECHSLER ^{52,53,54}
R. ZHOU ⁵ AND H. ZOU ⁵⁵

ABSTRACT

Deep spectroscopic samples can be used to improve photometric redshift (photo- z) estimates and reduce uncertainties on redshift distributions. Such improvements can increase the cosmological constraining power of large imaging-based experiments such as the Vera C. Rubin Observatory’s Legacy Survey of Space and Time (LSST) and mitigate what may be a limiting systematic effect. We present results from the “DESI-Deep pilot” program, which was designed to assess the capability of the Dark Energy Spectroscopic Instrument (DESI) on the 4m Mayall telescope to measure redshifts of galaxies as faint as expected lensing samples for early LSST data ($m_i \leq 24.5$). We find that DESI is remarkably efficient at this task, with redshift success rates comparable to the results of observations from 10m-class telescopes with only $\sim 2\times$ longer integration time (rather than $\sim 8\times$ longer as would be expected from aperture-area scaling), while simultaneously achieving ~ 30 times larger multiplexing. We also find that the signal-to-noise ratio of the spectra scales as expected for background-limited observations even for the longest exposure times (~ 7 hours) and faintest targets in the program. These results demonstrate that DESI could provide the definitive redshift sample for the early years of LSST with a modest investment of observing time. Based upon the results of this program, we provide updated predictions for the time required to collect benchmark samples for photo- z training and calibration using a variety of spectroscopic facilities. Finally, we describe a potential “DESI-Deep” survey designed to train and calibrate photo- z ’s for imaging experiments, and provide forecasts of its impact on cosmological inference.

1. INTRODUCTION

Photometric redshifts (photo- z ’s) – i.e., estimates of galaxy redshifts obtained from photometric data alone – are essential to the success of ongoing and upcoming imaging-based astrophysical experiments. Despite the high precision of spectroscopic redshifts, acquiring spectra for billions of faint objects observed by modern imaging surveys is impractical. Therefore, projects like the Dark Energy Survey (DES; [The Dark Energy Survey Collaboration 2005](#)), the Hyper Suprime-Cam Subaru Strategic Program (HSC-SSP; [H. Aihara et al. 2018a](#)), the Euclid Mission ([R. Laureijs et al. 2011](#)), the Nancy

Grace Roman Space Telescope ([R. Akeson et al. 2019](#)), and the Vera C. Rubin Observatory’s Legacy Survey of Space and Time (LSST; [Ž. Ivezić et al. 2019](#)) all rely heavily on photometric redshifts to achieve their scientific objectives. Photo- z ’s enable such projects to construct three-dimensional maps of galaxy and matter distributions, study the formation and evolution of galaxy populations, and trace the Universe’s expansion and structure growth over time. Improved photo- z ’s with reduced scatter will ultimately lead to smaller random errors in the measurements of cosmological parameters from these experiments. However, if systematic errors

in the photo- z estimates are not controlled, they could overshadow all other sources of uncertainty in these experiments.

Beginning with [W. A. Baum \(1957\)](#), numerous techniques have been employed to estimate photometric redshifts. These include physics-based approaches that model galaxy spectral energy distributions (SEDs) against observed photometry, and empirical methods that establish a non-linear mathematical relationship between input photometry and redshift using a training set of galaxies with known z . For an overview of commonly used methods, the reader can refer to [M. Salvato et al. \(2019\)](#), [M. Brescia et al. \(2021\)](#) or [J. A. Newman & D. Gruen \(2022\)](#).

Regardless of the method used, large sets of spectroscopic redshift measurements that are representative of the target galaxy population are needed to improve algorithm performance and enhance the characterization of z distributions. For physics-based methods, these measurements are needed for refining galaxy SED templates and error models, devising priors, and correcting for telescope throughput and photometric zero-points, whereas empirical methods require a representative training set to work at all. Larger spectroscopic data sets can reduce random uncertainties in photometric redshift estimates for individual objects (i.e., improve *training*) and also enable more stringent estimates of the redshift distributions of galaxies in tomographic bins (i.e., improve *calibration*). The same data set can help improve on both fronts, provided that it is devoid of biases introduced by sample selection.

The need for such data sets was strongly emphasized in several submissions to recent community planning exercises. These included submissions to the Decadal Survey on Astronomy and Astrophysics 2020 (Astro2020) by [J. Newman et al. \(2019\)](#) and [R. Mandelbaum et al. \(2019\)](#) and to the US Community Study on the Future of Particle Physics (Snowmass 2021) by [J. A. Blazek et al. \(2022\)](#), [D. J. Schlegel et al. \(2022a\)](#) and [J. Annis et al. \(2022\)](#). This need was subsequently endorsed by the final report of the Particle Physics Project Prioritization Panel (P5; [S. Asai et al. 2024](#)). The importance of spectroscopic samples for redshift calibration was further demonstrated in the recent Kilo-Degree Survey (KiDS; [A. H. Wright et al. 2024](#)) Legacy analysis, which found that the incorporation of larger, higher-quality spectroscopic samples and improved methods of redshift calibration greatly reduced the level of tension with Planck cosmological results ([A. H. Wright et al. 2025a,b](#)).

Previous efforts, such as the Complete Calibration of the Color-Redshift Relation (C3R2; [D. C. Masters et al.](#)

[2019](#)) survey, have sought to optimize the targeting of spectroscopic samples by observing galaxies in specific regions of color space that are underrepresented or missing in existing data sets, thereby aiming to complete the map of the color-redshift relationship. Recent and proposed surveys such as the DESI Complete Calibration of the Color-Redshift Relation (DC3R2; [J. McCullough et al. 2024](#)) and 4MOST Complete Calibration of the Color-Redshift Relation (4C3R2; [D. Gruen et al. 2023](#)) also follow a similar approach. However, as the galaxy population evolves significantly over cosmic time, samples of brighter galaxies do not cover the full span of colors of fainter objects at the same z (as demonstrated by [M. R. Blanton \(2006\)](#)). Furthermore, redshift can vary with magnitude for fixed galaxy colors ([J. McCullough et al. 2024](#); [D. C. Masters et al. 2019](#)). Therefore, we cannot simply use spectroscopic data sets that are limited to brighter galaxies for training and calibration of redshifts of fainter galaxies. Moreover, heavily color- (or color-magnitude-) selected samples, which often involve nonlinear selection criteria, can introduce selection biases that affect our ultimate redshift estimates in complex ways ([D. Gruen & F. Brimiouille 2017](#); [W. G. Hartley et al. 2020](#); [J. Myles et al. 2021](#)). Therefore, magnitude-limited spectroscopic samples that match the depths of the photometric samples used in a study and include minimal selection cuts can offer a way forward to mitigate these biases.

[J. A. Newman et al. \(2015\)](#) laid out a proposal for such a sample arguing approximately 20,000 spectra, distributed over multiple wide sky fields and matching the depths of imaging surveys, should be sufficient to meet the photo- z training and calibration requirements for next-generation astrophysical experiments like LSST; both past theoretical forecasts and the scaling of errors with training set size for machine learning algorithms lead to numbers of this order. They also provided estimates of the time required to obtain such samples calculated by appropriately scaling the exposure times required by the DEIMOS spectrograph ([S. M. Faber et al. 2003](#)) on the Keck 10m telescope, while accounting for the multiplexing capabilities, fields of view, and collecting areas of various instruments. However, these scaling relations require testing now that a new generation of spectroscopic instruments, such as the Dark Energy Spectroscopic Instrument (DESI; [DESI Collaboration et al. 2022](#)) and the Subaru Prime Focus Spectrograph (PFS; [N. Tamura et al. 2016](#)), are becoming available. This work aims to provide such an update, by leveraging the capabilities of DESI.

The Dark Energy Spectroscopic Instrument (DESI; [DESI Collaboration et al. 2022](#)) is a robotic, fiber-fed,

highly multiplexed spectroscopic surveyor operating on the Mayall 4-meter telescope at Kitt Peak National Observatory. DESI can obtain spectra of nearly 5000 objects simultaneously over a $\sim 3^\circ$ diameter field (DESI Collaboration et al. 2016a,b; J. H. Silber et al. 2023; T. N. Miller et al. 2023; C. Poppett et al. 2024). DESI is currently conducting an eight-year survey of roughly 17,000 sq. degs. of the sky which will obtain spectra for over 60 million galaxies and quasars (compared to the initial forecasts of ~ 40 million). DESI is designed to enable a Stage IV dark energy experiment (by the definition of the Dark Energy Task Force; A. Albrecht et al. 2006) with its main objective of revealing the nature of dark energy by achieving the most accurate measurement of the Universe’s expansion history to date using baryonic acoustic oscillation measurements (M. Levi et al. 2013). The First Data Release (DR1, DESI Collaboration et al. 2025a) that includes spectra for more than 18 million unique targets is now public. Early results from DESI have already set significant constraints on models of dark energy (DESI Collaboration et al. 2025b,c).

Due to its high multiplexing capabilities, large field-of-view, and efficient design, DESI is well-positioned to deliver deep spectroscopic samples that could significantly enhance the scientific output from existing imaging-based dark energy experiments and early data from LSST. A secondary target program, the DESI Complete Calibration of the Color-Redshift Relation (DC3R2; J. McCullough et al. 2024), has successfully utilized data from the survey to calibrate photo- z distributions for color-selected and comparatively bright objects (with z -band magnitudes $\lesssim 22.1$).

However, DESI’s redshift measurement performance for fainter objects, particularly at LSST-like depths, is largely untested. This work addresses this need by studying DESI’s performance for galaxies at depths comparable to those anticipated for the LSST Year 1 weak lensing sample ($m_i \lesssim 24.05$). We specifically targeted objects fainter than past DESI targets while utilizing observing strategies that help us achieve exposure times longer than in the DESI main survey. Our goal is to assess the feasibility of using the current DESI hardware (or improved versions thereof) to obtain spectroscopic samples that can effectively train photometric redshift methods and calibrate redshift distributions for current and future imaging experiments.

This article is organized as follows. In Section 2, we describe our target selection, observations, and data processing. In Section 3, we detail our methods for estimating DESI’s redshift measurement efficiency, scaling of the signal-to-noise ratio (SNR) with exposure time and

source faintness, and creating a predictive model of the expected redshift success rate for a magnitude-limited sample. Section 4 then presents the corresponding results. Section 5 provides some intuition behind our results to aid in their interpretation and presents updated estimates of the time required to conduct a survey similar to that proposed by J. A. Newman et al. (2015) using various instruments. Section 5 also lays down an example plan for a future DESI survey to obtain such a sample and quantifies the benefits of such a data set in terms of cosmological constraining power. Finally, Section 6 summarizes the content of the article.

2. DATA

The sample presented in this article, hereafter referred to as the “DESI-Deep pilot sample” (or simply the “pilot sample”) was observed as part of a pilot program to characterize the potential to target new classes of objects with DESI after its current eight-year survey is complete. Two patches of the sky were chosen to carry out the observations, centered around the fields observed by the XMM-LSS (M. Pierre et al. 2004) and COSMOS (N. Scoville et al. 2007) surveys because of the availability of additional multi-wavelength observational data in these regions. The field of view of DESI (~ 8 sq. degs.) is significantly larger than the portions of these fields with the most extensive imaging. As a result, many of the DESI fibers could only be devoted to target classes that could be selected with more limited photometry. This has enabled the tests of DESI’s performance for broadly-selected samples of faint galaxies described in this paper.

For our purposes, we shall refer to the region between $33.5^\circ \leq \text{R.A.} \leq 37.5^\circ$ and $-7^\circ \leq \text{DEC.} \leq -3^\circ$ as the DESI-XMMLSS field and the region between $148^\circ \leq \text{R.A.} \leq 152^\circ$; $0^\circ \leq \text{DEC.} \leq 4^\circ$ as the DESI-COSMOS field. Some additional observations were carried out in a region between $244^\circ \leq \text{R.A.} \leq 248.5^\circ$ and $41.5^\circ \leq \text{DEC.} \leq 45^\circ$ which we label the DESI-Hercules field. Those observations were carried out as a filler program because the spectrographs were not properly cooled and, therefore, yielded poor data quality. Results from this field nevertheless do not change the overall conclusions of this work; for completeness, we present selected analyses of the DESI-Hercules data in Appendix A. All results in the main body of this paper are based only on data from the DESI-XMMLSS and DESI-COSMOS fields.

2.1. Target Selection

We have used photometry from the Hyper Suprime-Cam Subaru Strategic Program (HSC-SSP; H. Aihara et al. 2018b) public data release 3 (PDR3; H. Aihara

et al. 2022) wide catalog to select targets for the observations. We first queried for all unique sources (i.e., those with `isprimary=true`) from the PDR3 catalog located in our regions of interest in the sky and implemented several quality cuts. Specifically, we require that objects have a positive and finite `cmodel` flux measurement in the *i*-band and that the *i*-band photometry model fitting process did not fail (i.e., that `i_cmodel_flag=false` and `i_sdsscentroid_flag=false`). We then correct all apparent magnitudes in this work for Milky Way dust extinction based upon the D. J. Schlegel et al. (1998) reddening maps. As a proxy for DESI *i*-band fiber magnitudes ($m_{i\text{-fiber}}$), we use the quantity `i-convolvedflux-2-15-flux` from the HSC catalogs, which are the aperture magnitudes calculated from flux within a 1.5-arcsec (9.0-pixel) aperture diameter at a seeing of 1.09 arcsec (6.5 pixels FWHM). This definition differs slightly from the standard DESI convention for fiber magnitudes, which assumes the same aperture diameter but a fixed seeing of 1 arcsec. To assess the impact of this choice, we compared the *r*-band fiber magnitudes from HSC with those of the DESI Legacy Imaging Surveys DR9 (D. Schlegel et al. in prep.) for objects within the DESI-COSMOS and DESI-XMMLSS footprints. We found a median offset of approximately 0.2 mag, with the HSC-based fiber magnitudes appearing systematically fainter. Consequently, the results presented in this work using HSC-based magnitudes should be considered conservative.

We then selected galaxies that have extinction-corrected *i*-band magnitudes in the ranges of $22 \leq m_i \leq 24.5$ and $22 \leq m_{i\text{-fiber}} \leq 24.75$ in the DESI-XMMLSS field, or $22 \leq i \leq 24.5$ and $22 \leq m_{i\text{-fiber}} \leq 25$ for objects in the DESI-COSMOS field. In order to ensure that the sample would not be dominated by faint objects but rather more evenly sample the underlying population, we randomly subsampled the selected galaxies to have a uniform distribution in *i*-band magnitude. This resulted in a list of targets with a density of 6,250 deg⁻² selected from the parent catalog with a density of approximately 97,500 deg⁻². We did not apply any strategy to photometrically separate stars from galaxies for these observations in order to avoid missing any galaxy/AGN populations.

2.2. Observations and Data Processing

The observations were conducted using a series of individual exposures each lasting 1000 seconds of effective exposure time (E. F. Schlafly et al. 2023; D. Kirkby et al. in prep.). This requires adapting the instrument exposure time depending upon the observing conditions to achieve the equivalent signal-to-noise ra-

tio of what would have been obtained for observations with no moon, at the zenith, no Milky Way dust extinction, 1.1 arcsec seeing, and a sky background of 21.07 mag per square arcsec in the *r* band in 1000 seconds. After each exposure, the telescope was dithered by 1.5 arcmin and a new set of fibers were allocated to the same targets. This was done to reduce the impact of any systematic errors in the data collection and processing pipeline. The list of potential targets was assigned specific fibers using the DESI fiber assignment algorithm (A. Raichoor et al. et al. in prep.). To ensure that the fainter objects ($23 \leq m_i \leq 24.5$) received longer total exposure times, such objects were assigned a higher priority within the DESI fiber assignment algorithm, though still lower than targets from other observational programs conducted simultaneously. As a consequence, more bright objects were observed in total, but their exposure time was shorter than that of objects in the fainter magnitude range.

The telescope dithering followed a strategy similar to the DESI survey validation (SV; DESI Collaboration, et al. 2024) in which the center of the focal plane was moved along a circle of radius 0.12° around the center of the field. The raw data were processed using the DESI spectroscopic pipeline (J. Guy et al. 2023), and individual exposures for each object were co-added to obtain the final spectra. Subsequently, the redshifts were measured from each co-added spectrum using Redrock (S. Bailey et al. in prep.). We used the DESI Data Release-1 (DR1; DESI Collaboration et al. 2025a) software stack¹ for processing the DESI-Deep pilot data.

We observed a total of 4863 targets: 3095 in the DESI-XMMLSS field and 1768 in the DESI-COSMOS field. Observations for the DESI-XMMLSS field were conducted on October 29 and December 10, 14–16, 2022, while those for the DESI-COSMOS field occurred on March 16–17 and 24–26, 2023. Figure 1 shows the on-sky locations of the observed targets in the two fields along with their respective exposure times. Figure 2 shows the distribution of *i*-magnitude and *i*-fiber-magnitudes for the observed targets, and Figure 3 shows the distribution of exposure times for the objects. The median effective exposure time was approximately 94 min (17–308 min), with a maximum of 535 min. The numbers in parentheses denote the 5th and 95th percentile values. For the DESI-XMMLSS observations, the median seeing was 1.04 arcsec (0.75–1.24 arcsec), and for the DESI-COSMOS observations, it was 1.09 arcsec (0.98–1.70 arcsec). The median airmass was 1.30

¹ <https://data.desi.lbl.gov/doc/releases/dr1/software-version/>

(1.25-1.45) for the DESI-XMMLSS observations and 1.24 (1.15-1.48) for the DESI-COSMOS observations. Similarly, the median transparency was 1.09 (0.92–1.14) for DESI-XMMLSS observations and 0.97 (0.90–1.02) for DESI-COSMOS observations.

Figure 4 shows four example spectra from the DESI-Deep pilot sample which yielded reliable redshift measurements. The top two panels display spectra of quiescent galaxies (characterized by absorption features), while the bottom two panels show star-forming galaxies with prominent emission lines. Within each class, the first object shown has a higher SNR than the second, illustrating performance across a range of conditions. All of these objects are fainter than typical targets in the main DESI survey, demonstrating DESI’s capability to secure reliable redshifts for galaxies with magnitudes comparable to lensing samples expected from early LSST data. While obtaining reliable redshifts is typically easier for galaxies with strong emission lines, these examples show that with sufficient exposure time, DESI can also reliably detect the absorption features of faint quiescent galaxies. This capability is crucial, as it enables secure redshift measurements for galaxy populations that are frequently challenging to measure.

3. METHODS

The primary objective of this work is to evaluate DESI’s effectiveness in measuring redshifts of objects substantially fainter than those typically observed during its main survey, reaching depths comparable to the faintest objects used for weak lensing analyses with LSST during its initial years of operation. We accomplish this by quantifying the redshift measurement efficiency of DESI as a function of both exposure time and target magnitude, while also assessing how the measured signal-to-noise ratios (SNRs) of the spectra scale with these parameters compared to the expectations for the background-limited regime. Additionally, we utilize the data to develop a model that establishes a relationship between exposure time, source brightness, and the probability of obtaining a successful redshift measurement, which could be used to plan future surveys with DESI or other similar instruments.

3.1. *Estimating Redshift Measurement Efficiency*

The DESI spectroscopic data reduction and redshift measurement pipelines, while extremely efficient, have only been optimized for the classes of objects observed during the main survey. Consequently, the standard automated metrics used to evaluate whether redshift measurements are secure were not well suited for the data obtained for this work. Instead, to determine whether

the redshift for a given object was measured securely, we carried out a visual inspection (VI) campaign similar to what was done for DESI survey validation (T.-W. Lan et al. 2023; D. M. Alexander et al. 2023).

Specifically, the processed spectra, along with an initial redshift measurement from Redrock, were used to create interactive web pages using the Prospect² spectrum visualization tool. Volunteers from the DESI collaboration then used the platform to visually inspect each spectrum with the Redrock fits and assigned a score of 0 to 4 to each spectra as a measure of the quality of the redshift measurement. A score of 4 indicated that two or more spectral features (e.g., emission or absorption lines or spectral breaks) are clearly and definitively visible, while 3 indicated the presence of at least one strong spectral feature as well as weaker spectral features that are highly likely to confirm the redshift. A score of 2 indicated that only one strong spectral feature was identifiable; 1 indicated that signal was present but the features could not be identified; and 0 indicated that no signal was present. If the automatic redshift measurement did not align with the observed spectrum, the inspectors manually determined the redshift by aligning a series of template spectra with the observed spectrum.

Each spectrum was inspected by two volunteers and their scores were averaged. If the average score was either 3 or higher, or 2 or lower, the score was retained. For average scores between 2 and 3, a third inspection was conducted by an individual considered to be expert in redshift assessment, and their score was considered final. Since securely measuring a redshift requires two or more spectral features be identified, only those objects whose redshifts were assigned a final score of 3 or higher were considered good enough to make a successful redshift measurement. This binary division (into “reliable”, i.e. quality ≥ 3 , versus “unreliable” redshifts) has been used to compute the instrument’s redshift measurement success rates for the DESI-Deep pilot sample (see Sect. 4.1). To isolate the impact of target selection from instrument performance, we entirely excluded objects that had reliable redshift estimates but were spectroscopically classified as stars when calculating redshift measurement success rates. Figure 5 illustrates the measured redshift distribution for objects deemed to have reliable redshift measurements after visual inspection. The redshift measurement success rate for some sample (e.g., objects within some bins in magnitude) is then simply the fraction of non-stellar objects in that sample that are assessed to have reliable redshift measurements.

² <https://github.com/desihub/prospect>

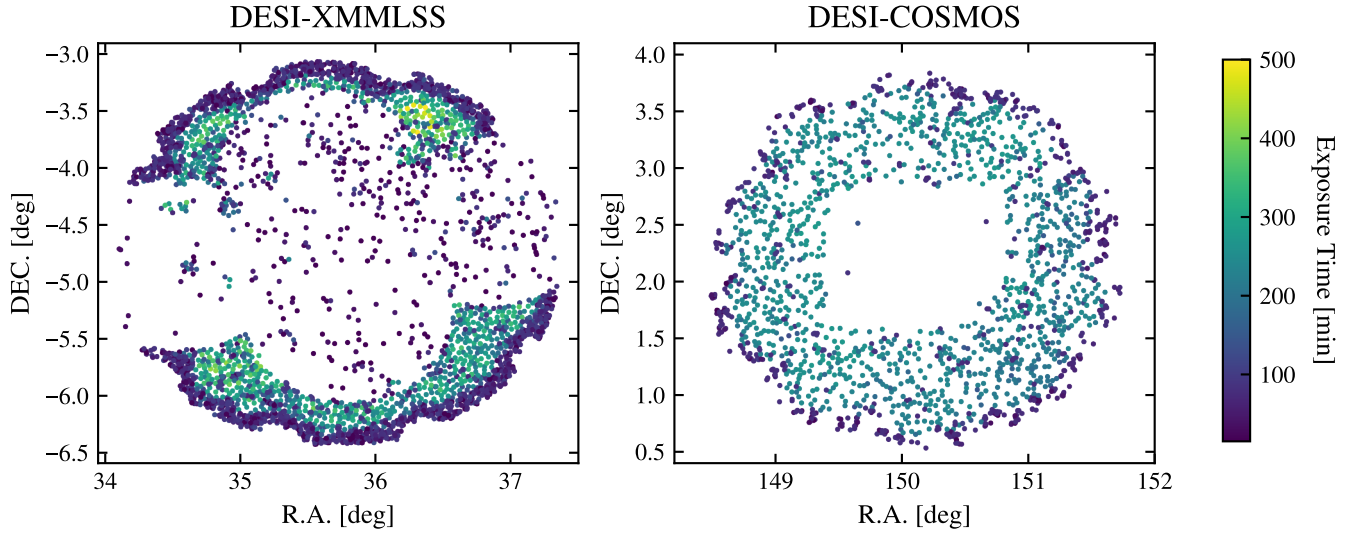


Figure 1. Locations of the 4863 objects observed in the DESI-XMMLSS ($33.5^\circ \leq \text{R.A.} \leq 37.5^\circ$; $-7^\circ \leq \text{DEC.} \leq -3^\circ$) and DESI-COSMOS ($148^\circ \leq \text{R.A.} \leq 152^\circ$; $0^\circ \leq \text{DEC.} \leq 4^\circ$) fields. The colors of the points indicate the total effective exposure time for each object. The empty regions at the center of each field were filled by targets for other pilot observations for potential future DESI programs.

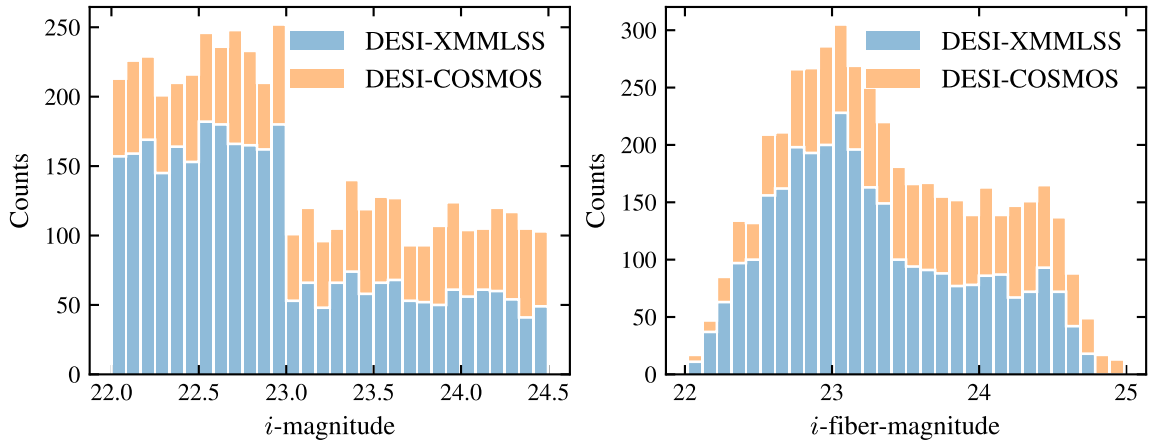


Figure 2. Distribution of i -magnitudes (left) and i -fiber-magnitudes (right) of the objects observed in the two fields, represented by stacked histograms. Objects with i -magnitude greater than 23 were allocated a higher priority in the DESI fiber assignment algorithm (A. Raichoor et al. et al. in prep.) resulting in them getting greater exposure times but for fewer objects.

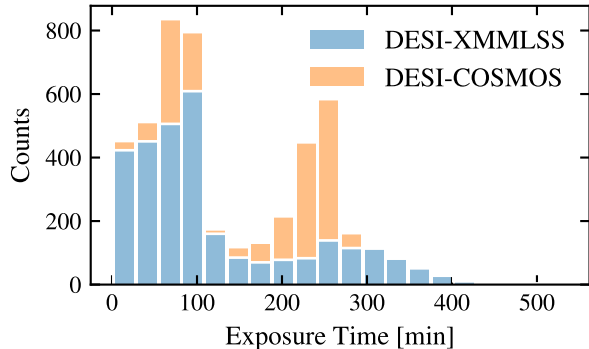


Figure 3. Distribution of effective exposure times represented by a stacked histogram. The median exposure time is about 94 min with the 5th and 95th percentiles of the distribution being 17 and 308 min respectively.

3.1.1. Adjusted Redshift Measurement Efficiency

Above $z \sim 1.6$, only spectral features that are weak in most galaxies will lie within the wavelength range covered by the DESI spectrographs. As a result, objects may lack reliable redshift measurements due to their spectrum having insufficient signal-to-noise ratio to identify spectral features that are typically strong, or alternatively because they are intrinsically at a redshift greater than 1.6 where prominent features are not covered by the spectrum. To decouple these two effects, we therefore calculate an adjusted redshift measurement success rate by estimating the number of objects that would be expected to be at redshifts greater than 1.6 in a sample (e.g., a magnitude bin) and subtracting that from the number of failures and total number of objects. The adjusted success rate can be used to separately assess DESI’s redshift measurement efficiency for a magnitude-limited sample with only objects at redshifts less than or equal to 1.6.

Since the redshift distribution of galaxies as a function of magnitude is not well measured at high redshifts, we have used two approaches to calculate the expected number of galaxies with redshift greater than 1.6 in a given bin, one analytic and one empirical. For an empirical adjustment, we used the COSMOS2020 data set (J. R. Weaver et al. 2022) with FARMER photometry (J. R. Weaver et al. 2023) and restricted the data to only the “clean” objects (FLAG_COMBINED=0). We used the redshift distribution of galaxies in this data set as a proxy for the true distribution. We used photometric redshift measurements obtained by the LePhare code (S. Arnouts et al. 2002; O. Ilbert et al. 2006) and selected objects that have good photometric measurements and are consistent with being galaxies or active galactic nuclei. The fraction of objects in a given magnitude bin

that are at $z > 1.6$ can then be calculated directly from these photo- z ’s.

For the analytic adjustment, we instead use the empirical relations for galaxy redshift distributions provided in the LSST Science Book (LSST Science Collaboration et al. 2009), which were derived using the methods presented in A. L. Coil et al. (2004) (but based on the final DEEP2 dataset). In this case, the redshift probability density function for a i -band magnitude limit $m_{i\text{-lim}}$, $p(z|m_{i\text{-lim}})$, is given by the formula:

$$p(z|m_{i\text{-lim}}) = \frac{1}{2z_0} \left(\frac{z}{z_0}\right)^2 \exp[-(z/z_0)^\alpha], \quad (1)$$

where $z_0 = 0.0417 \times m_{i\text{-lim}} - 0.744$ and $\alpha = 1$. The total number of galaxies per square degree present within a magnitude-limited sample (N_{gal}) is given by:

$$N_{\text{gal}}(m_{i\text{-lim}}) = 46 \times 10^{0.31(m_{i\text{-lim}} - 25)}. \quad (2)$$

By taking the integral of $p(z|m_{i\text{-lim}})$ within specific magnitude bins and scaling that with $N_{\text{gal}}(m_{i\text{-lim}})$, we can obtain the number counts of galaxies with redshifts greater than 1.6.

The analytical relations used are calibrated on optical spectroscopic data that become increasingly uncertain at $z > 1.5$. Similarly, photometric redshifts have not been fully validated for faint galaxy populations and higher redshift regimes. Therefore, these adjusted success rates should be considered to be estimates based on currently available data which may have inaccuracies when extrapolated.

3.2. Calculating Signal-to-Noise Ratios

In addition to assessing the redshift measurement efficiency, we also wish to investigate how the SNR of the observed spectra changes with exposure time and source brightness; this can then be compared to the expectations for observations in the background-limited regime (i.e., assuming that the noise in spectra is Poissonian and independent of source properties like flux, etc. but rather depends solely on the background count rate and exposure time). This analysis can provide insight into potential instrumental limitations, helping determine whether additional exposure time should yield meaningful improvements in SNR.

To ensure the robustness of our findings, we performed an empirical measurement of the SNR which is independent of the DESI spectroscopic pipeline. The spectroscopic pipeline makes assumptions about the intrinsic spectra of galaxies and is optimized for the main survey targets (J. Guy et al. 2023); we wish to avoid any impact from these assumptions.

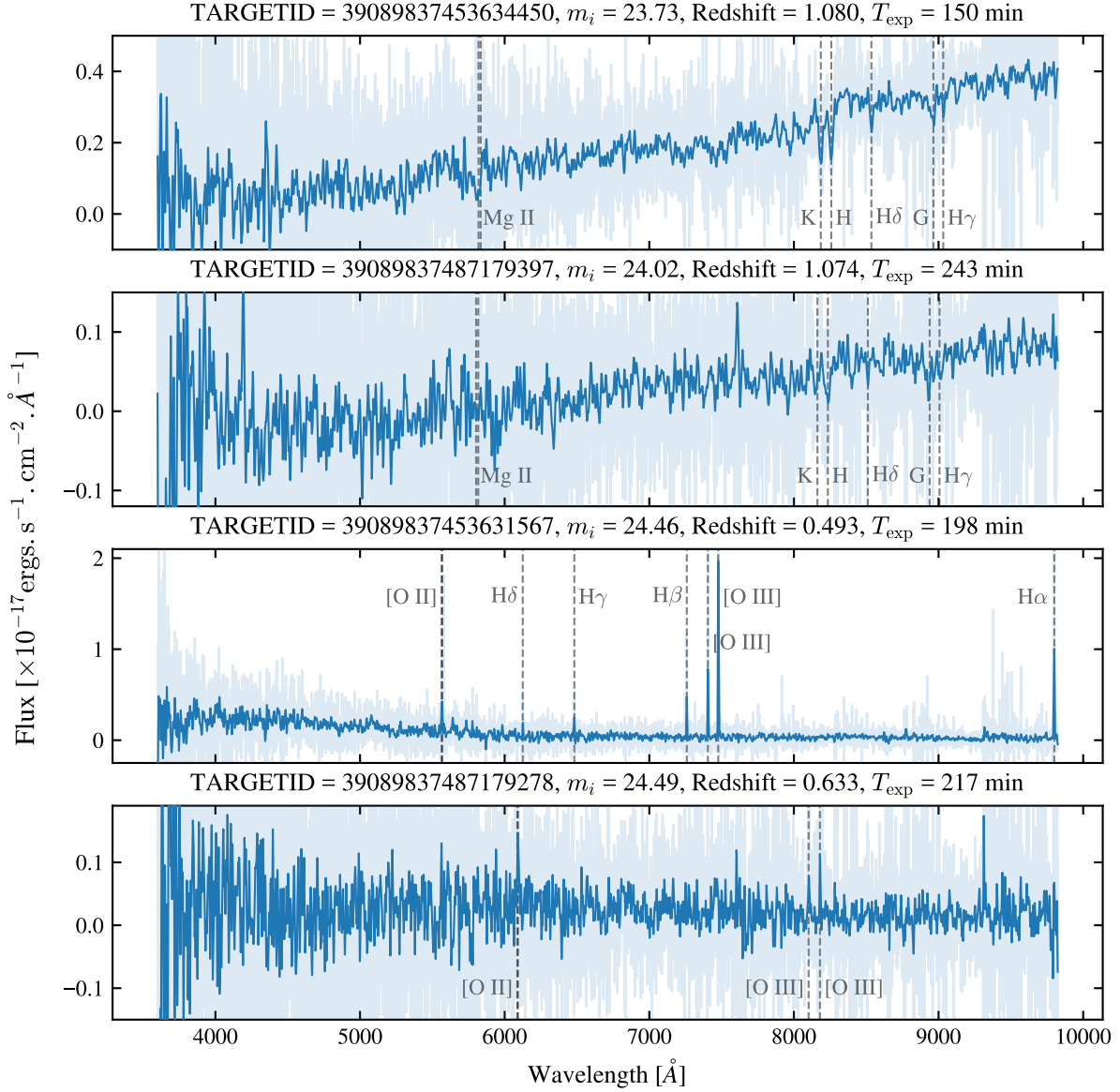


Figure 4. Spectra of four example galaxies from the DESI-Deep pilot sample, chosen to illustrate the diversity of galaxy types for which reliable redshifts were obtained. In each panel, the light blue shading represents the observed spectrum, while the solid dark blue line is the spectrum smoothed using an inverse-variance-weighted Gaussian kernel (with $\sigma = 4$ Å for the first two spectra and $\sigma = 2.4$ Å for the last two) to enhance spectral features for visualization. Data from the three DESI spectrograph arms have been coadded to form a single spectrum per object. Prominent absorption and emission lines are marked by labeled, gray dashed vertical lines. The TARGETID, i -band magnitude (m_i), measured redshift, and effective exposure time (T_{exp}) are listed above each spectrum. The top two panels show spectra of quiescent galaxies with absorption features dominating the redshift determination, while the bottom two show star-forming galaxies with prominent emission lines. The first object shown from each class has a higher signal-to-noise ratio than the second. All of these objects are fainter than typical targets in the main DESI survey, demonstrating the instrument’s ability to obtain reliable redshifts for galaxies as faint as those expected to be used for early lensing analyses with Rubin Observatory, given sufficient exposure time.

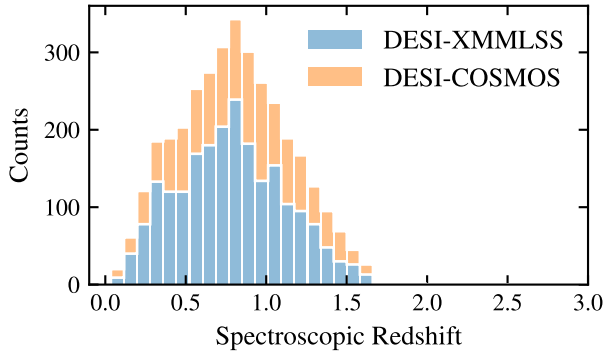


Figure 5. The redshift distribution of objects for which spectra were observed and deemed good enough for reliable redshift measurements. Beyond redshift 1.6, most prominent spectral features of a galaxy shift out of the rest-frame observing range of the DESI spectrographs, leading to a lack of redshift measurements beyond this z .

To calculate SNRs, we exclusively used data from the red arm of the DESI spectrographs in the wavelength range of 5900 Å to 7500 Å. We focus on this wavelength range as it has been found in previous work (J. Guy et al. 2023) to be the most reliable regime for characterizing DESI instrument performance across a range of scenarios. We first subtract the best-fit Redrock template spectrum (S. Bailey et al. in prep.) from the spectrum of a given object. We then divide the wavelength range into 64 bins of 25 Å width each and calculate the SNR for each bin as the ratio of the mean flux of the original spectrum within that bin divided by the standard deviation of the template-subtracted spectrum within it. The final empirical SNR is then calculated as the median of the binned SNRs. This yields a reliable quantification of the SNR of the spectral continuum, while being resilient to any anomalies, strong spectral features, etc. thanks to the robustness of the median statistic.

3.3. Modeling the Redshift Success Rate

Using the DESI-Deep pilot survey results, we have developed a simple model for the probability of obtaining a successful redshift measurement given the total DESI effective exposure time (T_{exp}) and the brightness of a source. This model will be particularly valuable for planning future surveys with DESI or other similar instruments (cf. Sections 5.2 & 5.3). To relate the exposure time and brightness to the success probability, we begin by defining a new parameter which we call the adjusted magnitude:

$$\widetilde{m}_i = m_i - 1.25 \log \left(\frac{T_{\text{exp}} [\text{sec}]}{6000 [\text{sec}]} \right), \quad (3)$$

where m_i is the extinction corrected i -band magnitude. This quantity is related to the negative logarithm of the SNR (Eq. 6) up to an additive constant.

The probability of success calculated from the outcome of a series of success–failure experiments (Bernoulli trials) follows a Binomial distribution. We therefore use a generalized linear model with a Binomial likelihood (also known as logistic regression; see P. McCullagh & J. A. Nelder 1989 for a discussion on the topic) to model the observed success rate as a function of adjusted magnitude. For the systematic component of the model, we chose a linear transformation of \widetilde{m}_i , as it is the simplest model that well represents the data. The probability of obtaining a successful redshift measurement is then given by:

$$p_{\text{success}} = \sigma(A \cdot \widetilde{m}_i + B), \quad (4)$$

where $\sigma(x)$ is the inverse of the logit function (also called the sigmoid function) and is defined as:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (5)$$

and A and B are the parameters of the systematic component of the generalized linear model, which are determined using maximum likelihood estimation. We employ Bootstrap Resampling (B. Efron 1979) with 1000 resamples, for the construction of 95% confidence and prediction intervals.

4. RESULTS

4.1. Redshift Success Rates

The DESI pipelines are optimized for the main survey targets, which are typically color-selected and brighter than the objects in the DESI-Deep pilot sample, potentially making the standard metrics used to assess the quality of a redshift measurement unreliable for our purposes. Although the main survey primarily employs $\Delta\chi^2$ thresholds (DESI Collaboration, et al. 2024) to assess redshift reliability (where $\Delta\chi^2$ represents the difference between the χ^2 values for the best and second-best redshift fits), we observe from Figure 6 that there is significant scatter in the relation between visual inspection quality flags and $\Delta\chi^2$ values; e.g., even some objects with $\Delta\chi^2 > 100$ (much larger than the threshold used for DESI samples) were found to have insecure redshift measurements, while many other objects with $\Delta\chi^2 < 10$ (below the typical DESI threshold) had reliable redshifts.

Therefore, there is no clear threshold in $\Delta\chi^2$ that separates reliable redshifts from unreliable ones with high purity and completeness. Consequently, we rely on VI labels for quality assessment (as described in Section 3.1) for this work. We are also working to de-

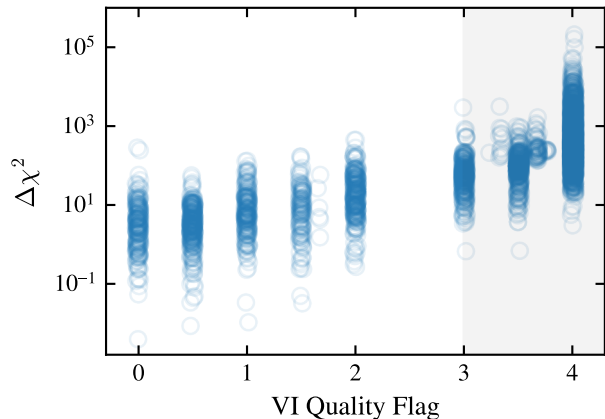


Figure 6. The $\Delta\chi^2$ between the best and second-best redshift fit plotted against visual inspection (VI) quality flags for our data set. The shaded gray region shows the flags corresponding to objects with reliable redshift measurements. While the DESI main survey uses $\Delta\chi^2$ for determining redshift measurement reliability, we observe a significant scatter between VI quality classifications and $\Delta\chi^2$ values, without any clear threshold that separates reliable redshift measurements (i.e., VI Quality Flag ≥ 3) from unreliable ones for the DESI-Deep pilot data set. Therefore, we exclusively use VI labels for quality assessment in this work.

velop automated methods for identifying reliable redshifts based on alternative metrics, which will be discussed in a future work.

We assess the instrument’s redshift measurement performance using the redshift measurement success rate, defined as the fraction of objects that have reliable redshift estimates (i.e., VI quality flag ≥ 3) and are spectroscopically classified as galaxies relative to the full number of non-stellar objects (i.e., excluding those with spectra reliably classified as stars from the denominator). This provides a conservative estimate of instrumental capability while isolating out any target selection effects.

We have evaluated the redshift measurement success rate in bins of i -band magnitude, i -band fiber magnitude, or effective exposure time. Since these success rate estimates are derived from finite samples, we calculate the 95% confidence limits using the Clopper-Pearson intervals (C. J. Clopper & E. S. Pearson 1934), which are appropriate for the binomial distribution that governs the probability of a particular spectrum yielding a reliable redshift. Figure 7 shows DESI’s redshift measurement success rate as a function of i -band magnitude and in bins of effective exposure time. Figure 8 shows the same but as a function of the i -band fiber magnitude. Only observations with exposure times longer than one hour are included in each plot; the widths of the time

bins were chosen such that the varying exposure times within a bin do not change the SNR by more than a factor of 1.2.

We compare the DESI redshift measurement success rates with those of the DEEP2 and DEEP3 galaxy redshift surveys (J. A. Newman et al. 2013; R. Zhou et al. 2019). Data for these surveys were collected using the DEIMOS spectrograph (S. M. Faber et al. 2003) on the Keck-II 10m telescope at W. M. Keck Observatory and represents the largest available (as of now) spectroscopic data set of galaxies at magnitudes similar to those of the DESI-Deep pilot sample.

We performed a spatial cross-match of the DEEP2/3 data with overlapping HSC-SSP PDR3 wide survey photometry with a tolerance of 1 arcsecond to perform the comparison. To create the parent catalog, we select objects with a photometric cross-match, spectroscopic quality flag (ZQUALITY) greater than 0, and measured redshifts greater than 0.001. We consider a DEEP2/3 redshift measurement to be successful if ZQUALITY is 3 or higher. The redshift measurement success rate is then defined as the fraction of objects with a successful redshift measurement in a given magnitude bin. All DEEP2/3 observations were performed with an on-sky exposure time of approximately 1 hour; therefore we do not bin this sample in time.

We also compared the success rates with those from the zCOSMOS galaxy redshift survey (S. J. Lilly et al. 2007); however, this test is limited by their sample that lacks a color selection (zCOSMOS-bright; S. J. Lilly et al. 2009) having a magnitude limit ($m_i = 22.5$) only a little fainter than the bright end of the DESI-Deep pilot sample. However, where there is an overlap, the zCOSMOS-bright sample exhibits significantly lower redshift measurement success rates (roughly 50% for objects with $22 < m_i \leq 22.5$, when determined using similar criteria for a reliable measurement as this work) compared to either DEEP2/3 or this work at the same magnitudes.

Figures 7 and 8 demonstrate that DESI consistently achieves a redshift measurement success rate greater than 75% for brighter objects in our sample with relatively short exposure times of about 1.5 hours. Similar success rates are achieved for the fainter objects with high enough exposure times. Remarkably, we also see that with about 2 hours of exposure time, DESI can match the redshift measurement success rate of the DEEP2 and DEEP3 surveys. The observed trends are the same when the success rates are checked as a function of i -band magnitude or i -band fiber magnitudes, suggesting that the success rates are not significantly af-

ected by the reduced amount of light entering the DESI fiber-fed spectrographs.

Figure 9 shows the redshift measurement success rate as a function of exposure time in bins of i -band magnitude. As expected, we observe that the success rate increases with increasing exposure times at a given magnitude. We see that for the brighter objects in our sample over 75% success rate can be achieved with about 2 hours of exposure time, whereas similar success rates require over 7 hours of exposure time for the faint objects in our sample. We have used our data to create a quantitative model of the success rate as discussed in Sections 3.3 and 4.3.

To estimate the redshift measurement success rate when the failures are only due to objects with redshift less than 1.6 – i.e., objects with potential prominent spectral features within the rest-frame wavelength range accessible to DESI – we can adjust the measured success rates following the procedure described in Section 3.1.1. Figure 10 shows the measured and adjusted success rates as a function of i -band magnitude and in bins of exposure time with the adjustment done based upon data with deep multi-band photometry (COSMOS2020) or instead when based upon empirical analytic relations. The effect of the $z > 1.6$ failures becomes prominent for the fainter objects in our sample, as the fraction of high-redshift objects increases in this regime. For the fainter objects in our data set, we find that roughly 5 to 10% of the redshift measurement failures are due to $z > 1.6$ objects, so the redshift failures are consistent with being dominated by objects with $z \leq 1.6$.

4.2. Scaling of Signal-to-Noise Ratios

We next test how the scaling of noise in DESI spectra with exposure time compares to what would be expected if the data are background-limited. To do this we can compare the observed scaling of the continuum SNR of the spectra (measured as described in Section 3.2) with exposure time and observed flux to what would be expected in the background-limited regime. For that case, we expect:

$$\text{SNR} \propto \text{Flux} \times (\text{Exposure Time})^{1/2},$$

as the total number of signal counts will be proportional to flux times exposure time, whereas the noise is proportional to the square root of the number of background counts, which will be given by the background flux times the exposure time. Hence, for an object with i -band fiber magnitude $m_{i\text{-fiber}}$ that is observed for an effective exposure time of T_{exp} , the expected SNR can be written as:

$$\text{Expected SNR} = \kappa \times 10^{-(m_{i\text{-fiber}}/2.5)} \times T_{\text{exp}}^{1/2}, \quad (6)$$

where κ is a proportionality constant that we can determine from the data. In practice, we calculate its value separately for each object with an empirical SNR measurement, and then set κ to be the median of these results. We use the fiber magnitude rather than overall magnitude of an object within Eq. 6 as the former is more closely connected to the amount of light that enters the DESI spectrographs. We use Milky Way extinction-corrected fiber magnitudes for this calculation as the DESI dynamic exposure time calculator accounts for the effect of Galactic dust when scaling exposure times.

Figures 11 and 12 show that the ratio of empirical to expected SNR is close to one for the entire range of galaxy brightnesses and exposure times in our data set. DESI follows the expected SNR scaling of $T_{\text{exp}}^{0.5}$, matching the scaling demonstrated using other similar fiber-fed spectroscopic instruments (e.g., C. Lidman et al. 2020). Achieving high redshift measurement success rates for target samples similar to lensing samples in LSST data will require even longer exposure times than those we have tested so far (cf. Section 5.2). However, the current results are very promising and indicate that it is likely that still deeper observations may be performed without immediately hitting a ceiling set by instrument systematics such as imperfect photon counting in the spectrograph CCDs, sky subtraction, or losses incurred due to the aperture of the fibers.

4.3. Model for the Redshift Success Rate

Following the methodology described in Section 3.3, we obtain the best-fit values and 95% confidence intervals for the parameters in Equation 4 to be $A = -1.20_{-0.11}^{+0.10}$ and $B = 28.74_{-2.28}^{+2.63}$. Therefore, our predictive model for quantifying the expected redshift success rate as a function of source magnitude and exposure time is given by:

$$p_{\text{success}} = \sigma(-1.20 \widetilde{m}_i + 28.74) \quad (7)$$

where p_{success} is the probability a particular object's redshift will be securely measured, $\widetilde{m}_i = m_i - 1.25 \log \left(\frac{T_{\text{exp}}[\text{sec}]}{6000[\text{sec}]} \right)$ and σ is the sigmoid function defined as $\sigma(x) = \frac{1}{1+e^{-x}}$.

For planning future surveys, it can be useful to invert this equation and reframe the same model to predict the required exposure time ($\widetilde{T}_{\text{exp}}$) to achieve a success rate of p_{success} for an i -band magnitude limit of m_i . This is

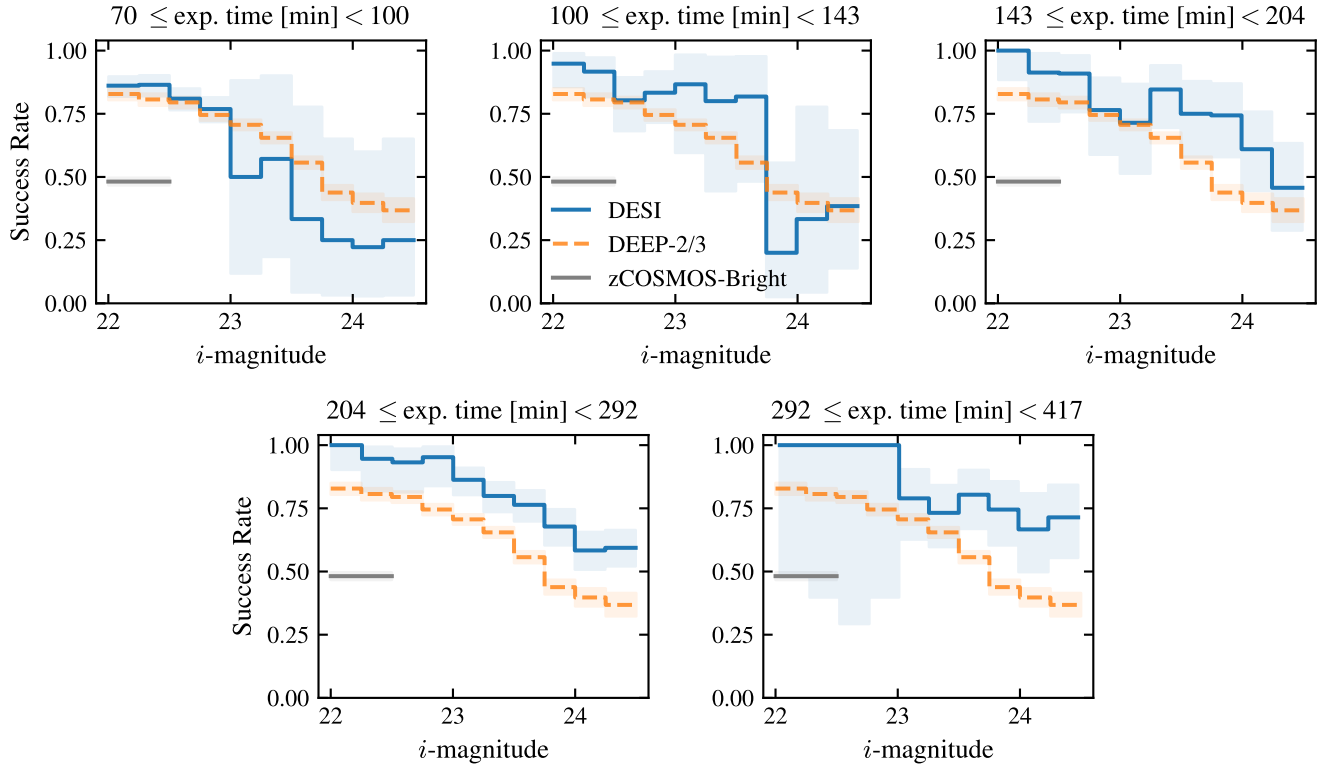


Figure 7. Redshift measurement success rate for the DESI-Deep pilot sample as a function of i -band magnitude, with separate panels corresponding to different ranges of effective exposure time. The blue curve shows the success rate obtained from our observations, with the 95% confidence interval shown using the shaded blue region. The success rates as a function of i -magnitude for the combined DEEP2 and DEEP3 survey datasets are plotted in orange for comparison, with the shaded orange area showing the corresponding 95% confidence interval. The gray line denotes the average success rate for the zCOSMOS-Bright sample for objects with i -band magnitudes between 22 and 22.5. DEEP2, DEEP3 and zCOSMOS-Bright had typical exposure times of roughly one hour, and hence the same curve is plotted for all bins of exposure time. We observe that at exposure times of about 2 hours DESI’s success rate matches or exceeds that of the DEEP2 and DEEP3 galaxy redshift surveys.

given by:

$$\tilde{T}_{\text{exp}}(m_i, p_{\text{success}}) [\text{sec}] = 10^{-15.38+0.8 m_i+0.67 \times \text{logit}(p_{\text{success}})} \quad (8)$$

where the logit function is the inverse of the sigmoid function, defined as $\text{logit}(x) = \ln\left(\frac{x}{1-x}\right)$.

Figure 13 shows our best-fit predictive model along with the measured success rates in bins of effective magnitude. We observe that this simple model is able to represent the observed success rates very well and will allow us to make estimates of exposure times required for any future sky survey with a DESI-like instrument.

5. DISCUSSION

Our results have shown that DESI has achieved a high redshift measurement efficiency and excellent signal-to-noise scaling compared to prior instruments. In this section, we provide insight into these results by discussing key differences in instrumentation and data processing pipelines that may explain these differences. We then use our measurements to update the survey time esti-

mates for obtaining the redshift training and calibration samples proposed by J. A. Newman et al. (2015) and J. A. Blazek et al. (2022). Finally, building on these findings, we lay out a plan for a new magnitude-limited survey, which we call the “DESI-Deep” program, to leverage the instrument’s powerful capabilities for this purpose.

5.1. Explaining Differences in Redshift Measurement Efficiency

The results from the DESI-Deep pilot sample indicate that DESI must be significantly more efficient than the Keck/DEIMOS-based DEEP2 and DEEP3 surveys at measuring redshifts for faint galaxies, when accounting for differences in telescope aperture. The same holds for the VLT/VIMOS-based zCOSMOS survey, but in this section we will focus on DEEP2/DEEP3 where there is greater overlap in magnitude range and more detailed comparisons are feasible. This significant gap warrants investigation into its underlying causes. While a de-

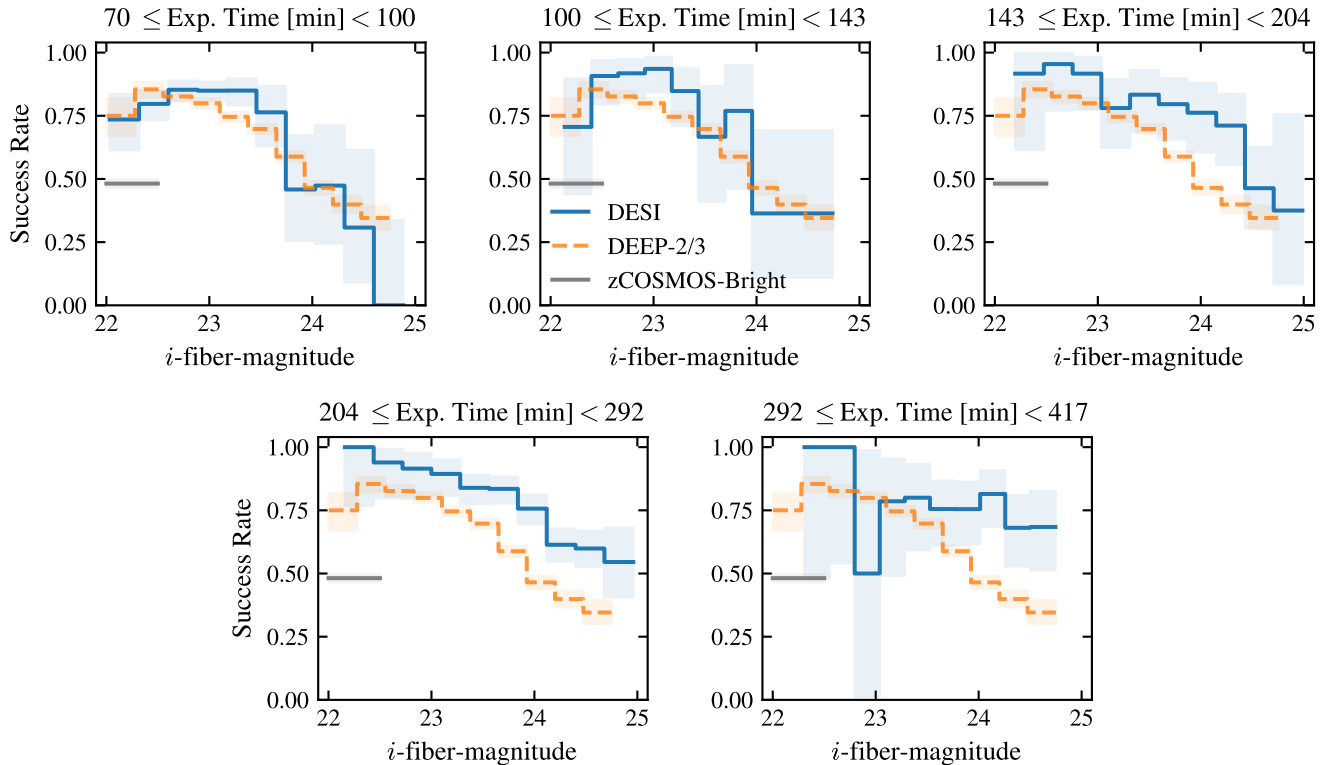


Figure 8. Redshift measurement success rate for the DESI-Deep pilot sample as a function of i -band fiber magnitude, with separate panels corresponding to different ranges of effective exposure time. The blue curve shows the success rate obtained for our observations, with the corresponding 95% confidence interval shown as the shaded blue region. The success rates for the combined DEEP2 and DEEP3 survey dataset is plotted in orange for comparison, with the shaded orange area showing the corresponding 95% confidence interval. The line denotes the average success rate for the zCOSMOS-Bright sample for objects with i -band magnitudes between 22 and 22.5. DEEP2, DEEP3 and zCOSMOS-Bright had typical exposure times of roughly one hour, and hence the same curve is plotted for all bins of exposure time. We observe that at exposure times of roughly 2 hours DESI’s redshift measurement success rate matches that of DEEP2/DEEP3.

tailed, component-by-component comparison of the two facilities is complex and beyond the scope of this work, we discuss a few key factors in instrumentation, data processing and observing strategy that likely contribute to these differences, based on our experience with both DESI and DEIMOS.

One of the primary factors determining an instrument’s efficiency for measuring redshifts is its total throughput, defined as the ratio of the number of photons captured by the spectrograph’s detectors to those collected by the telescope’s primary mirror. The Keck/DEIMOS system used in the DEEP2/3 surveys had a throughput of 15–25%³. In contrast, the DESI instrument achieves a total throughput of 30–40% (DESI Collaboration et al. 2022), a notable figure given that fiber-fed spectrographs typically have had lower throughput compared to slit-based spectrographs. This

enhanced throughput results from improvements across nearly all aspects of instrumentation, including advancements in corrector lens materials and designs, improvements in spectrograph design and stability, and the use of charge-coupled device (CCD) detectors with higher quantum efficiency compared to previous generations.

The effective throughput of spectroscopy will also be affected by any offsets between an object’s true position on the sky and its targeted location, which would cause a significant fraction of light to fail to be captured by a fiber or slitlet. The imaging catalogs used for targeting by the DEEP2/3 surveys employed astrometry tied to the USNO-A2.0 catalog (A. Monet et al. 1998); due to the sparseness of USNO-A2.0 stars, systematic errors in that catalog, and imperfections in the algorithms used to map this astrometry onto CFHT-12K imaging, those catalogs exhibited systematic astrometric errors of up to 1 arcsec, modulated over scales of 5–10 arcmin, when compared to SDSS catalogs (A. L. Coil et al. 2004; D. J. Matthews et al. 2013). Since the

³ <https://www2.keck.hawaii.edu/inst/deimos/ripisc.html>

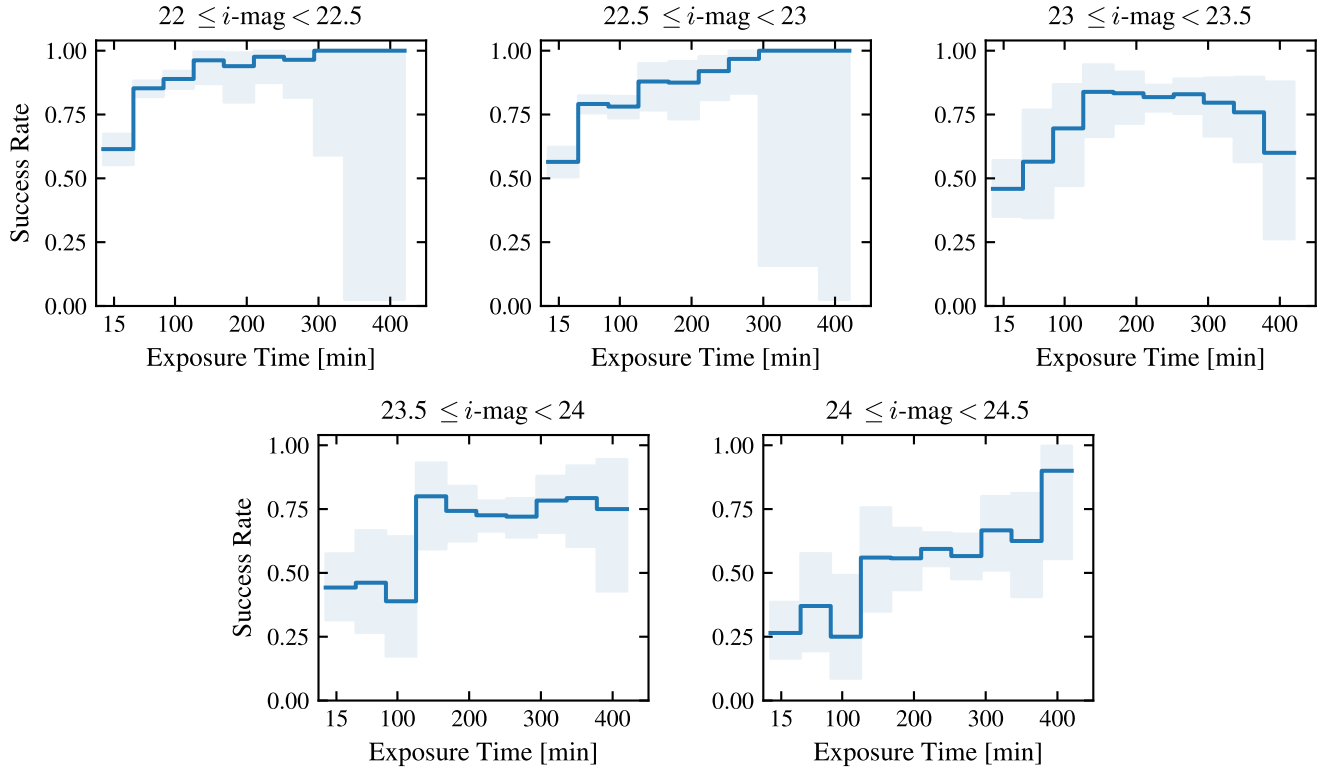


Figure 9. Redshift measurement success rate for the DESI-Deep pilot sample as a function of exposure time, with separate panels corresponding to different bins of i -band magnitude. The blue curve shows the success rate obtained for our observations, with 95% confidence intervals shown using the shaded blue region. For every i magnitude bin, we observe that the success rate increases with increase in exposure time as expected for a background-dominated measurement. The figure also provides us with a rough estimate of the success rate to be expected for a given i -magnitude limit and exposure time.

slitmasks used by Keck/DEIMOS are 16 arcmin long and astrometric reference stars were near the edges of the masks, this could cause the masks to be misaligned compared to some objects. Stars were typically centered within their alignment boxes to ~ 0.1 arcsec, but doing so could cause galaxies to be off center if they were in a different area of the imaging with relative astrometric offsets.

In contrast the HSC-SSP PDR3 catalog we use for targeting has been able to base its global astrometry on the Gaia DR1 catalog (Gaia Collaboration et al. 2016), which contributes negligibly (~ 1 milliarcsec) to the total error budget; the RMS scatter of bright star positions in HSC-SSP PDR3 relative to Gaia is 0.013 arcsec (H. Aihara et al. 2022). The RMS error in positioning DESI fibers on the sky is 0.19 arcsec RMS (J. H. Silber et al. 2023) and contributes the most to the astrometric error budget. This large difference in astrometric accuracy could lead to substantial light losses with the narrow slits used for DEEP2/DEEP3 observations compared to DESI.

DESI adjusts the actual wall-clock exposure time for each observation dynamically, based on real-time ob-

serving conditions and Milky Way dust extinction, in order to achieve a relatively uniform SNR across exposures (E. F. Schlafly et al. 2023). As a consequence, the actual exposure times are generally somewhat longer than the effective exposure time values. In contrast, the DEEP2/3 surveys did not employ a dynamic exposure time calculator, although exposure times were occasionally extended on an ad hoc basis to compensate for degraded observing conditions. A typical DEEP2/3 exposure time was approximately 1 hour. For the DESI-Deep pilot sample, the median wall-clock time for observations was approximately 1.5 times larger than the corresponding effective exposure time. These differing strategies for determining exposure times may partly explain the observed differences in redshift measurement efficiency. However, even under the most favorable assumptions, a factor of 1.5 in effective exposure time is insufficient to fully explain the total difference in redshift measurement efficiency between the surveys.

Higher redshift measurement success rates can also be facilitated by better data processing, particularly the subtraction of night sky lines. DESI has significant instrumental advantages for this over DEIMOS, since the

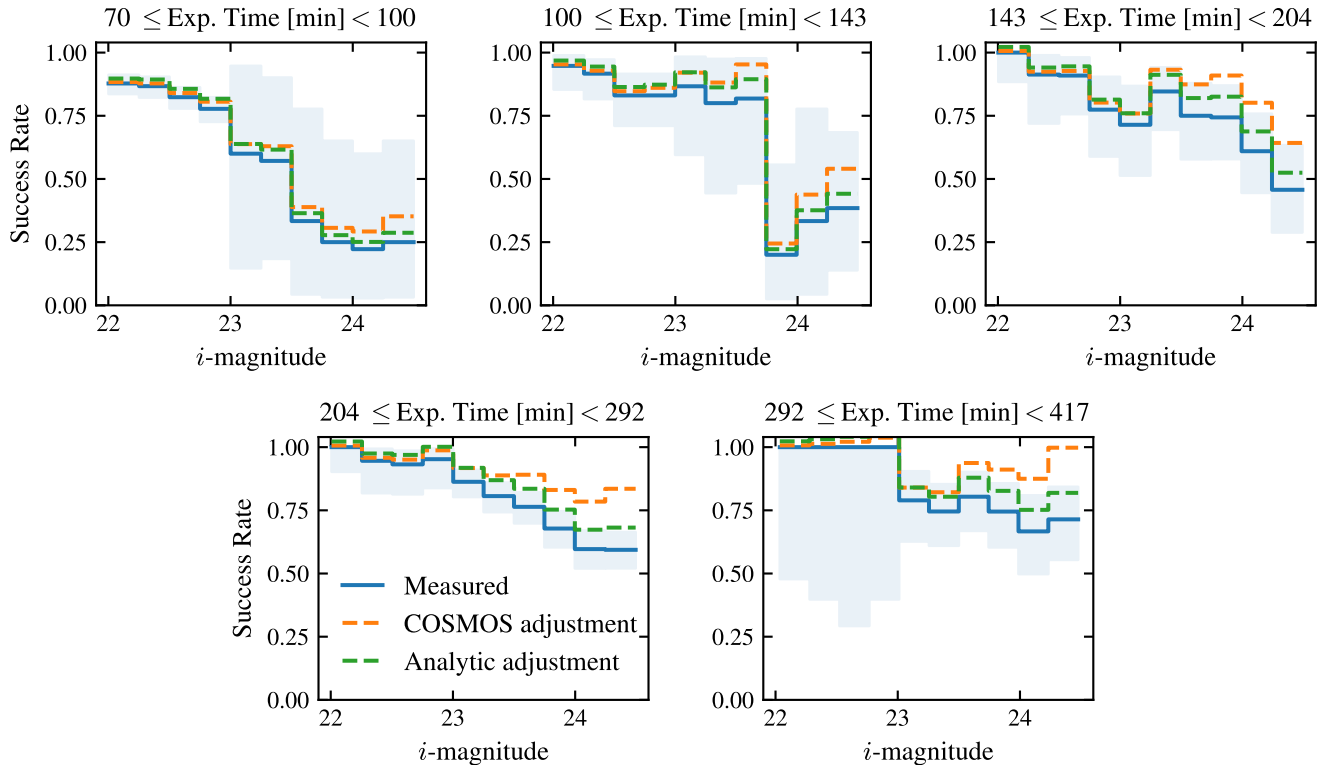


Figure 10. Test of the impact of adjusting for the abundance of $z > 1.6$ objects when calculating redshift success rates. The measured and adjusted redshift success rate are plotted as a function of i -band magnitude, with separate panels corresponding to different ranges of exposure time. As before, the solid blue curve shows the measured success rate as measured with 95% confidence intervals shown using the shaded blue region. The dashed curves adjust this success rate to instead estimate the success rate for a magnitude-limited sample that only includes objects with redshifts less than or equal to 1.6. The orange dashed curve represents the success rate when the adjustment is calculated using the COSMOS2020 data set, while the dashed green line instead shows the success rate when the adjustment is done based upon analytic fits (see Section 3.1.1 for details). The adjusted curves should be indicative of DESI’s redshift success rate when all the observed objects have prominent spectral features within the wavelength coverage of DESI. The effect of the $z > 1.6$ failures becomes more prominent for the fainter objects in our sample, but even in that regime most redshift failures appear to be due to objects at redshifts below 1.6.

fibers are always the same (as opposed to the custom slitmasks used for each DEIMOS observation) and the spectrographs are kept stable (rather than having to rotate as the telescope moves in the case of DEIMOS). The DESI spectroscopic data reduction pipelines also incorporate significant technical advances (J. Guy et al. 2023) which should contribute to its high redshift measurement efficiency. Two key improvements are the methods used for spectral extraction and sky subtraction.

The DEEP2/3 pipelines used the “optimal extraction” method from K. Horne (1986), which is based upon the assumption that the 2D spectrograph point-spread function (PSF) is separable, a condition often not met in practice. In contrast, DESI employs the “spectro-perfectionism” algorithm of A. S. Bolton & D. J. Schlegel (2010). This technique performs a “perfect” extraction of one-dimensional spectra by creating a 2D forward model of the raw pixel data, using a precise

model of the 2D PSF instead of a projected 1D profile. This approach yields several advantages, including minimizing variance, incorporating a common wavelength grid for all fibers, and having a well-defined resolution matrix. These improvements are crucial for subsequent processing steps, leading to better sky subtraction, flux and wavelength calibration, and ultimately, redshift estimation.

These advances in spectral extraction directly enable a more sophisticated approach to sky subtraction. The DEEP2/3 pipelines primarily utilized a “local” sky subtraction, where an interpolated signal from sky regions adjacent to the science target on a given slitlet was subtracted. By leveraging the common wavelength grid and resolution matrix from spectral extractions, the DESI pipeline utilizes “non-local” sky subtraction methods instead. It constructs a coherent model of the sky spectrum for an entire spectrograph image using data

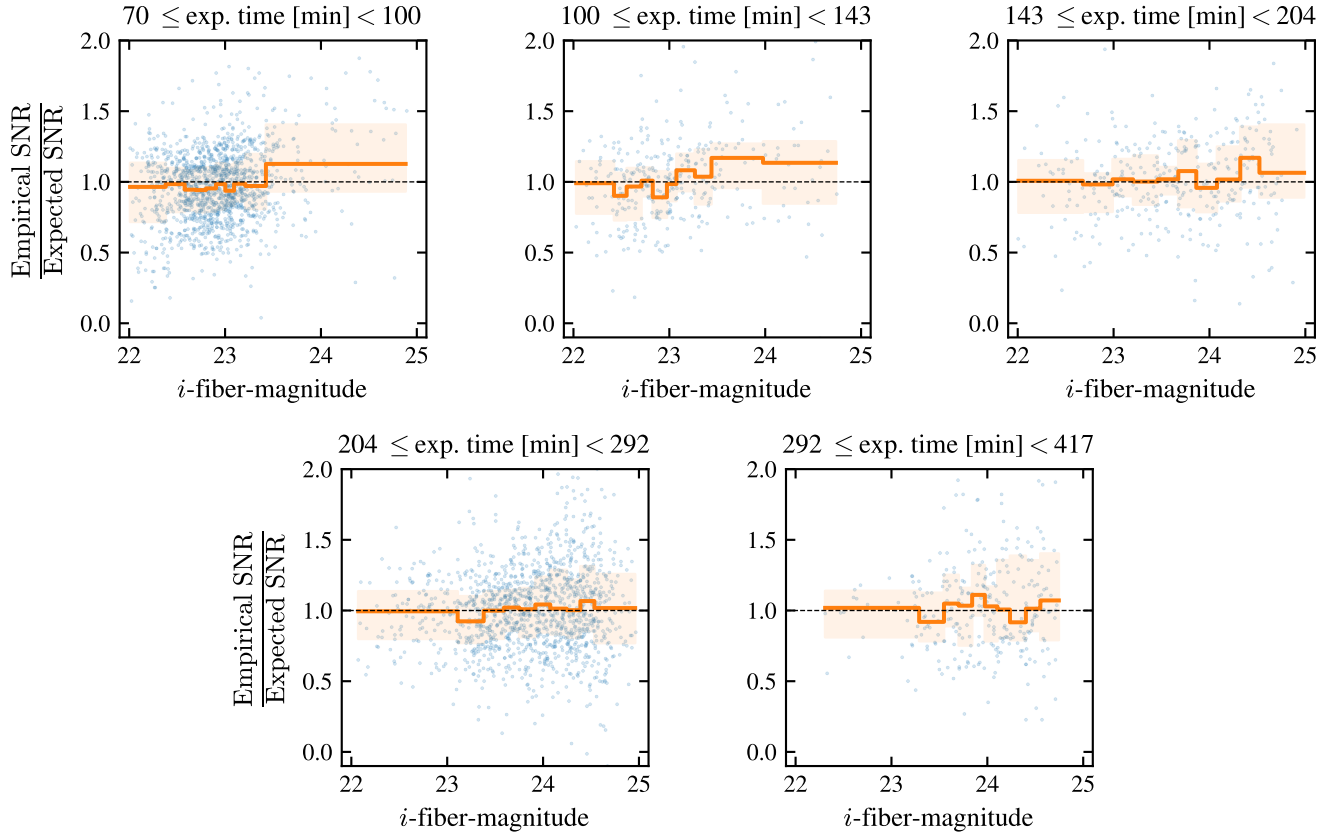


Figure 11. The ratio of empirically measured SNR and what we would expect from a background limited regime plotted as a function of i -fiber-magnitude in bins of exposure time. Each blue dot represents a single object and the orange curve denotes the average in ten equal population bins. The shaded orange region shows the 95% confidence interval on the mean. We observe that the instrument performance follows the background limited expectations for the entire range of magnitude and exposure times tested and even the faintest galaxies and for the longest exposure times.

from numerous sky fibers distributed across the focal plane. The created model is further adjusted using a principal component analysis (PCA) derived model of strong sky lines (c.f. section 4.7 of [J. Guy et al. 2023](#) for details). This comprehensive model is then subtracted from each science spectrum. The combination of “spectro-perfectionism” and non-local sky modeling allows the DESI pipeline to perform significantly better sky subtraction. Similar algorithms that utilize non-local sky subtraction with PCA-based methods have been shown to be effective in other fiber-fed spectroscopic instruments, enabling long exposure times without a significant reduction in expected SNR ([R. Sharp & H. Parkinson 2010](#); [M. Hart 2019](#)). Residuals from imperfect sky subtraction are the dominant source of systematic errors for ground-based, fiber-fed spectrographs that prevent efficient observations of faint sources ([K. Bundy et al. 2022](#)). Reducing this primary source of errors helps to enable DESI to carry out deeper observations while still tracking the expected SNR scaling relations for background-limited spectroscopy.

In addition to improvements in throughput and data processing, DESI’s broad wavelength coverage significantly enhances its redshift measurement efficiency compared to previous surveys. The DESI spectrographs cover a range of nearly 6200 Å (from 3600 Å to 9800 Å); in contrast, the DEEP2 survey with DEIMOS was limited to a span of about 2600 Å. While the subsequent DEEP3 survey utilized a lower-resolution grating yielding a broader wavelength range, its effective coverage of approximately 4000 – 4500 Å remains considerably smaller than that of DESI. This limited wavelength range restricts the spectral features observable at a given redshift, thereby lowering the overall redshift success rate. For instance, even if a galaxy has strong spectral features, those features may be redshifted out of the instrument’s observable window. For the DEIMOS setups used in DEEP2/3, there was a practical maximum redshift limit of $z \approx 1.45$ compared to $z \approx 1.6$ for DESI, contributing to DESI’s higher redshift measurement efficiency compared to the DEEP2/3 surveys.

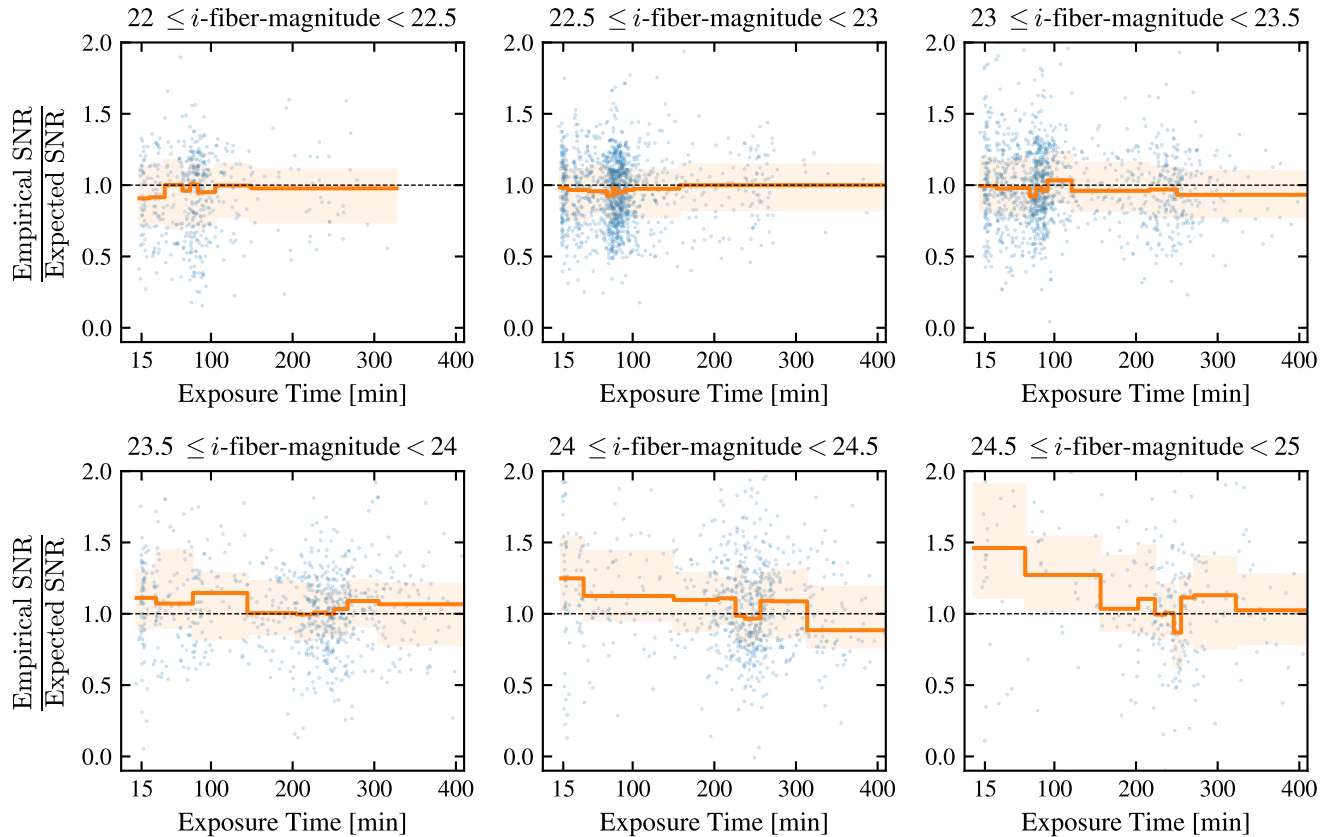


Figure 12. The ratio of empirically measured SNR and what we would expect from a background limited regime plotted as a function of exposure time in bins of i -fiber-magnitude. Each blue dot represents a single object and the orange curve denotes the average in ten equal population bins. The shaded orange region shows the 95% confidence interval on the mean. We observe that the instrument performance follows the background limited expectations for the entire range of magnitude and exposure times tested and even the faintest galaxies and for the longest exposure times.

Furthermore, DESI is able to provide this broader wavelength coverage while still maintaining a relatively high spectral resolution, particularly for the infrared arm of the spectrograph (covering $6560 - 9800 \text{ \AA}$ with $R \sim 4000$). DESI can therefore resolve narrow features such as the [O II] $\lambda\lambda 3726, 3729$ doublet, which is a key feature for redshift determination for galaxies with emission lines, at $z > 0.76$. In contrast, for the lower-resolution grating used in DEEP3, this feature is often unresolved. The contrast is even greater with lower-resolution instruments like the VIMOS spectrograph (O. Le Fèvre et al. 2003); e.g., zCOSMOS-bright observations had resolution $R \sim 600$, much too low to split the two components of the [O II] doublet (J. Comparat et al. 2013). Since any pair of spectral features (including both components of a single doublet) is sufficient to securely determine a redshift, this greatly benefits DESI’s efficiency at delivering robust z measurements. The relative impact of resolution versus wavelength range is not entirely clear, however; both DEEP2 and DEEP3 delivered similar overall redshift measurement success rates

to each other, despite differing by roughly a factor of two in spectral resolution and wavelength range.

A final source of differences between the DESI and DEEP2/3 observations lies in the observing strategies used. For DEEP2/3 the same slitmask was typically used for all observations of a given object. Furthermore, the instrument’s flexure compensation system is designed to ensure that observations of an object are repeatably placed at the same position on the detector. This can cause persistent patterns in sky subtraction residuals across observations (see Fig. 26 of J. A. Newman et al. 2013 for examples).

In contrast, our pilot survey implemented an observing strategy designed to mitigate systematics in deep integrations. Total integrations, reaching up to a maximum of 7 hours, were broken into units of 1000-sec effective exposure time (the DESI exposure time calculator [ETC; E. F. Schlafly et al. 2023; D. Kirkby et al. in prep.] is used to actively adjust exposure times to account for real-time observing conditions, including seeing, sky transparency, airmass, and Galactic reddening,

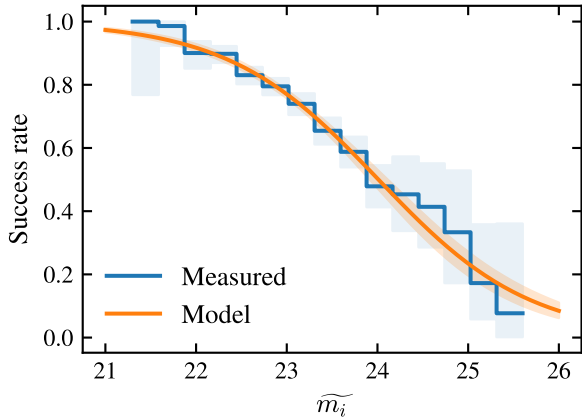


Figure 13. Redshift success rate versus the adjusted magnitude ($\tilde{m}_i = m_i - 1.25 \log\left(\frac{T_{\text{exp}} [\text{sec}]}{6000 [\text{sec}]}\right)$). The blue curve shows the success rate in 15 equal width bins, with the blue shaded region representing the 95% confidence interval on the counts. The orange curve shows the generalized linear model fit to the data with the shaded orange region denoting the 95% prediction interval obtained from bootstrap resampling. We have chosen the simplest model that well represents the observations.

keeping the depth equivalent between observations). Between each exposure, the telescope was slightly dithered, which places the same target onto a different fiber for subsequent observations. Thanks to DESI’s fast readout and fiber positioning speeds, this strategy adds negligible overhead to the total observing time. The crucial benefit of this dithering approach is that it averages over individual fiber- and position-dependent instrumental systematics, including any residuals from sky subtraction, suppressing their impact on the final coadded spectrum by a factor of $(\text{Number of exposures})^{-\frac{1}{2}}$. We believe that this multi-exposure dithering strategy is a primary reason why the signal-to-noise ratio in our data scales as expected for purely background-limited observations, even for the longest total exposure times on the faintest objects.

It is uncertain which if any of these factors is the dominant explanation for why the redshift efficiency of DESI is so high compared to what would be expected from scaling prior Keck and VLT surveys by relative aperture areas. However, all of them work in the same direction. It is therefore not surprising that DESI should be more efficient than prior surveys, but it was not expected that the difference in redshift measurement efficiency would be so high.

5.2. Implications for Spectroscopy in Support of Photometric Redshifts

Previous work by J. A. Newman et al. (2015) and J. A. Blazek et al. (2022) estimated the total telescope time that would be required to obtain a fiducial magnitude-limited spectroscopic sample designed to enable accurate photo- z training and calibration using a variety of existing and proposed facilities. However, those estimates were scaled from the performance of the Keck/DEIMOS spectrograph as used in the DEEP2 survey. However, we have found that the redshift measurement efficiency of DESI is significantly higher than what would be expected from this scaling; accordingly, estimates of the required time should be reassessed for any instrument whose performance would be similar to DESI’s.

In this section we therefore provide time estimates for the fiducial survey from J. A. Blazek et al. (2022) which are scaled from the measured performance of DESI rather than Keck/DEIMOS, which should be more appropriate for any instrument and observing strategy which more closely matches the DESI-Deep pilot program. We closely adhere to all other assumptions made in that work. We also add survey time estimates for several new facilities and have updated instrument specifications to reflect their current designs.

Following J. A. Blazek et al. (2022), we calculate the amount of time required for a benchmark survey designed to obtain spectroscopic samples for photo- z training and calibration for the 10-year LSST sample. The survey goals are to acquire spectra for at least 20,000 objects down to a magnitude of $m_i = 25.3$ with a $\sim 90\%$ redshift measurement success rate. To mitigate the effects of cosmic variance, these targets must be distributed across either 15 widely separated fields for instruments with a field-of-view (FoV) $\leq 1 \text{ deg}^2$, 6 fields for $1 \text{ deg}^2 < \text{FoV} < 3 \text{ deg}^2$, or 4 fields for instruments with $\text{FoV} \geq 3 \text{ deg}^2$. At a minimum, each field is required to be at least 0.1 deg^2 in area, which translates to a minimum total sky area of 1.5 deg^2 for such a survey.

To calculate the total survey duration for different instruments, we first determine the required per-object exposure time with DESI (\tilde{T}_{exp}) using Equation 8. We scale the time required for DESI by the ratio of the effective light collecting areas of the Mayall telescope and the telescope hosting a given instrument. For any multiplexed observation, we assume that the exposure time required is dictated by the faintest objects in the fiducial survey (i.e., that faint objects are common enough that they will be included within any pointing/instrument configuration; since galaxy number counts rise geometrically with magnitude, this should be the case). The total survey duration is then calculated by consider-

ing the time required to meet a given requirement of the fiducial survey – the number of fields required (the “Field-limited” case), the total survey area (the “FoV-limited” scenario), or the total number of objects with spectra (the “multiplex-limited” case). The survey time required to meet all requirements will then be the maximum of the time required for each of these scenarios.

Field-Limited Case: For an instrument with a large FoV and high multiplexing, the total observing time needed, T_{obs} , will be determined by the number of fields, N_{fields} , required to mitigate sample/cosmic variance:

$$T_{\text{obs}} = N_{\text{fields}} \times \tilde{T}_{\text{exp}}(m_i, p_{\text{success}}) \quad (9)$$

$$\times \frac{\text{Collecting Area}_{\text{DESI}}}{\text{Collecting Area}_{\text{Instrument}}}$$

where $\text{Collecting Area}_{\text{DESI}}$ is set to 8.66 m^2 . We take N_{fields} to be 15, 6, or 4, depending on the instrument’s FoV, as described above.

FoV-Limited Case. For an instrument with high multiplexing but a small FoV, the total observing time may be determined by the number of pointings needed to cover the required sky area:

$$T_{\text{obs}} = \frac{\text{Total Survey Area}}{\text{FoV Area}} \times \tilde{T}_{\text{exp}}(m_i, p_{\text{success}}) \quad (10)$$

$$\times \frac{\text{Collecting Area}_{\text{DESI}}}{\text{Collecting Area}_{\text{Instrument}}}$$

where the Total Survey Area is set to 1.5 deg^2 .

Multiplex-Limited Case. For an instrument with a large FoV but low multiplexing, the total observing time will be determined by the number of pointings needed to observe the total number of targets in the fiducial sample:

$$T_{\text{obs}} = \frac{\text{Number of Objects}}{\text{Multiplexing}} \times \tilde{T}_{\text{exp}}(m_i, p_{\text{success}}) \quad (11)$$

$$\times \frac{\text{Collecting Area}_{\text{DESI}}}{\text{Collecting Area}_{\text{Instrument}}}$$

where the total Number of Objects is set to 20,000.

In Table 1, we provide the final time estimates for a list of telescope and spectroscopic instrument pairs. This represents an update of Table 2 of J. A. Blazek et al. (2022). All estimates are given in units of “dark years”, where one dark year is defined as 365 5.33-hour dark-time nights (assuming an average of eight hours per night with one-third of the time lost due to weather and instrumental overheads). Dark-time is generally defined as the time after sunset when the sun is significantly below ($\gtrsim 15 \text{ deg}$) the horizon, the moon phase is not near full, and the moon is significantly far away ($\gtrsim 50 \text{ deg}$) from the field being observed (cf. E. F. Schlafly et al. 2023 for DESI’s definition of dark-time).

From Table 1, we find that the required number of dark years is significantly ($\sim 3\times$) reduced for each facility compared to the estimates in J. A. Blazek et al. (2022) when assuming DESI-like efficiency. Given the ongoing progress in technology, it could be achieved with sufficient instrumental stability and data processing quality equivalent to what has been implemented with DESI. It is therefore reasonable to assume that future spectroscopic facilities will be at least as efficient, if not more. This provides reassurance that spectroscopic samples for photo- z training and calibration of faint galaxy populations can be obtained in a feasible amount of time. We also note that DESI itself will be one of the most efficient instruments for obtaining such a sample.

5.3. Design of a Potential photo- z Survey with DESI (DESI-Deep)

The efficiency of DESI for measuring redshifts, as demonstrated in this work, can enable a dedicated survey of faint galaxies that would provide a transformative data set to the community with a modest investment of DESI observing time. The resulting data set would enhance the cosmological constraints of major imaging surveys and benefit nearly all areas of extragalactic science by characterizing the spectral energy distributions of a large, magnitude-limited sample of galaxies up to high redshifts. This data set could be used not only to train and calibrate photo- z ’s but also to develop similar methods for determining galaxies’ physical properties from photometric information (e.g., training algorithms for predicting stellar masses such as that presented in R. Zhou et al. 2023).

In this subsection, we describe an example design for a dedicated survey program with DESI, which we call the “DESI-Deep Survey” hereafter, which provides an example of what sort of training sets the instrument would provide. This design builds on the pilot program used in this work and addresses the critical lack of representative spectroscopic samples reaching depths comparable to lensing samples from early LSST data (let alone the full 10-year survey depth).

An ideal DESI-Deep survey would target a magnitude-limited sample of galaxies with $18 \leq m_i \leq 24.5$ and employ minimal selection cuts to ensure a representative sample. Targets can be selected from LSST photometry when available or from existing wide-area surveys like the HSC-SSP to enable immediate implementation; analysis of the sample would require LSST-based color information (e.g., for creating tomographic redshift bins or forward modeling any incompleteness) but targeting would not. The magnitude-limited sample would be

Table 1. Total effective exposure time required to obtain spectroscopic data sets for photo- z training and calibration using various spectroscopic facilities, assuming a DESI-like redshift measurement efficiency.

Telescope/Instrument	Collecting Area [m ²]	FoV Area [deg ²]	Multiplex	Total time (Dark Years)
Mayall / DESI	8.66	8	5,000	0.55
VISTA / 4MOST	11.99	4.2	1,624	1.21
WHT / WEAVE	12.7	3.1	960	1.94
6m-class / Spec-S5*	19	3.7	12850	0.12 ^a
6.5m-class / MegaMapper*	28.0	7.1	26,100	0.17
6.5m-class / MUST*	30	6.16	21170	0.16
VLT / MOONS*	52	0.55	1000	0.45
Subaru / PFS	52.8	1.25	2,400	0.19
Keck / DEIMOS	72.0	0.018	150	6.8 ^b
Keck / FOBOS*	72.0	0.1	1,800	0.25
12m-class / WST*	113	3.1	20000	0.04
GMT / MANIFEST + GMACS*	368	0.087	380	0.17
TMT / WFOS*	633.45	0.007	100	0.4
ELT / MOSAIC Optical*	978.0	0.01	200	0.18 ^c
ELT / MOSAIC NIR*	978.0	0.01	100	0.24

^aSpec-S5 is proposed to be two identical telescopes and spectroscopic instruments. The quoted time assumes that both facilities are observing in parallel. The telescope parameters quoted are for each individual observatory.

^bThe estimate for time required with Keck/DEIMOS has not been updated to reflect DESI-like efficiency as the original numbers from J. A. Blazek et al. (2022) were calculated based on Keck/DEIMOS data.

^cFor ELT, observations are required with both the optical and near-IR settings to achieve the required wavelength coverage, increasing total time required.

*The telescope/instrument has been proposed but is not operational as of writing this document (i.e., early 2026).

NOTE—The acronyms used in the table are defined as follows, DESI: Dark Energy Spectroscopic Instrument, VISTA: Visible and Infrared Survey Telescope for Astronomy, 4MOST: 4-metre Multi-Object Spectroscopic Telescope, WHT: William Herschel Telescope, WEAVE: WHT Enhanced Area Velocity Explorer, Spec-S5: The Stage-5 Spectroscopic Experiment, MUST: Multiplexed Survey Telescope, VLT: Very Large Telescope, MOONS: Multi-Object Optical and Near-IR Spectrograph, PFS: Prime Focus Spectrograph, DEIMOS: DEep Imaging Multi-Object Spectrograph, FOBOS: Fiber-Optic Broadband Optical Spectrograph, WST: Widefield Spectroscopic Telescope, GMT: Giant Magellan Telescope, MANIFEST: Many Instrument Fiber System, GMACS: GMT Multi-object Astronomical and Cosmological Spectrograph, TMT: Thirty Meter Telescope, WFOS: Wide-Field Optical Spectrometer, ELT: Extremely Large Telescope, MOSAIC: Multi-Object Spectrograph for Astrophysics, Intergalactic medium studies and Cosmology, NIR: Near Infrared.

References—DESI: DESI Collaboration et al. (2022), 4MOST: R. S. de Jong et al. (2019), WEAVE: S. Jin et al. (2024), Spec-S5: R. Besuner et al. (2025), MegaMapper: D. J. Schlegel et al. (2022b), MUST: C. Zhao et al. (2024), MOONS: M. Cirasuolo et al. (2020), PFS: N. Tamura et al. (2016), DEIMOS: S. M. Faber et al. (2003), FOBOS: K. Bundy et al. (2019), WST: R. Bacon et al. (2024), MANIFEST: T. Zafar et al. (2022), GMACS: D. L. DePoy et al. (2018), WFOS: C. Steidel et al. (2022), MOSAIC: R. Pelló et al. (2024).

downsampled to have a more uniform distribution in i -band magnitude to ensure that targets are not dominated by the faintest objects. In order to mitigate the effects of sample/cosmic variance, the proposed survey would perform observations on two ~ 10 square degree fields with a total exposure time of 25 hours per field, observing at least 1000 objects per unit magnitude and a total of 10,000 objects per field. The first phase of the survey is proposed to be conducted in two fields for easy logistics and running it in parallel to the main DESI survey but could be expanded to four or more fields in a future phase.

The target of observing 20,000 objects in total (10,000 per field) is motivated by the requirements for calibrating the redshift distributions for tomographic redshift bins used in cosmological analyses. Assuming a typical choice of five tomographic bins, this total sample size provides approximately 4,000 objects per bin. The primary goal is to precisely characterize the mean redshift of each tomographic bin, $\langle z \rangle$, as well as the spread in redshifts, $((z - \langle z \rangle)^2)^{1/2}$. Under the assumption that the observed sample is a random subsample of the underlying true distribution, the standard error on the population mean scales as $\sigma_{n(z)}/\sqrt{N}$, where $\sigma_{n(z)}$ is the standard deviation of the redshift distribution within the bin (where typically $\sigma_{n(z)} \approx 0.1$ for a redshift bin) and N is the number of objects. Similarly, the standard error on the standard deviation scales as $\sigma_{n(z)}/\sqrt{2N}$, so it is always smaller (but has similar calibration requirements to the mean, as found in [LSST Dark Energy Science Collaboration et al. 2018](#)). A sample of 4,000 objects per bin with a success rate of 75% would thus reduce the standard error of $\langle z \rangle$ to $\sim 2 \times 10^{-3}$ and the standard error of $((z - \langle z \rangle)^2)^{1/2}$ to $\sim 1.4 \times 10^{-3}$, meeting the requirements for stage IV dark energy experiments ([LSST Dark Energy Science Collaboration et al. 2018](#)).

Photometric redshift training and calibration are most effectively done in fields with deep photometry and rich multi-wavelength imaging, which enables the breaking of redshift degeneracies and the characterization of the impact of photometric errors ([A. H. Wright et al. 2020](#); [J. Myles et al. 2021](#)). The ideal fields for a DESI-Deep program are therefore the two equatorial Rubin Observatory LSST Deep Drilling Fields (DDFs) — COSMOS and XMM-LSS, as they will have exceptionally deep photometry from both Rubin and the Roman Space Telescope and extensive ancillary data. Of the LSST DDFs, the only other field accessible at all to DESI is ECDF-S, but it could only be observed at high air masses, greatly reducing observational efficiency.

A DESI-Deep survey can employ an observing strategy that is very similar to that used for the pilot sam-

ple. Effective exposure times of 1000 seconds per observation, as are used for the main DESI survey, should still be a good choice for such a program. The full field area would be tiled with dithered pointings in a rosette-like pattern in order to ensure that each target is observed with different fibers in different exposures. This serves to suppress any instrumental systematics that vary from fiber to fiber while maximizing both the number of unique targets observed and the cumulative exposure time for high-priority targets. The observations could be spread over multiple months or years and potentially be paired with other programs — such as instrument calibration or transient host redshift measurement — that require multiple visits to the same patch of sky over a long period of time.

The faint magnitude limit for the proposed DESI-Deep survey was chosen to balance two goals: creating a dataset that is immediately useful for near-term science while ensuring that the exposure time per object needed to obtain a high redshift success rate is not too long. Based on our predictive model (Equation 7), we expect a DESI-Deep survey with 25-hour maximum exposure times per target to achieve an overall success rate of at least 75% for objects with $m_i = 24.5$ (which is the same as the lensing sample depth for the HSC Year 3 data set). We anticipate this will exceed 80% at the anticipated LSST Year 1 lensing sample depth ($m_i = 24.05$) and be 90% or higher for targets with i -band magnitudes below 23. In [Figure 14](#) we show the expected success rates as a function of exposure time required for a variety of different magnitude limits. The magnitude limits are chosen to roughly represent expected source samples from various frontier weak lensing experiments.

Any spectroscopic survey will have redshift measurement failures, even with very long exposure times. Moreover, these failures do not occur randomly but tend to affect galaxies with specific colors or at particular redshifts, which can introduce biases in the inferred redshift distributions if not accounted for. We note that galaxies for which a reliable redshift measurement could not be obtained with DESI could be followed up in the future with larger telescopes with infrared spectrographs to determine the nature of redshift failures and further increase the value of a DESI-Deep data set.

5.4. *Impact of Incompleteness of Spectroscopic Samples on Cosmology*

While a complete quantification of the complex and non-linear effects of redshift failures is beyond the scope of this work, it is valuable to assess their potential impact on cosmological parameter constraints. To this end, we here use a cosmological forecasting framework to as-

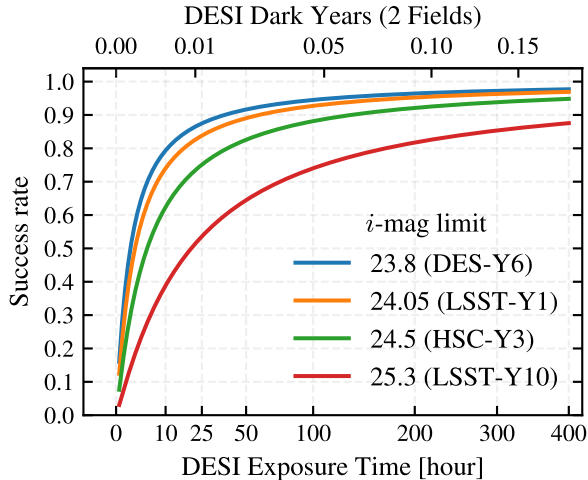


Figure 14. Predicted redshift measurement success rates plotted as a function of the total amount of dark time required for various magnitude limits. The magnitude limits are chosen to roughly represent the magnitude limits of the samples of various frontier weak lensing experiments. The lower horizontal axis displays the exposure times needed per object, while the upper horizontal axis shows the total time required to collect a magnitude-limited sample in the field-limited regime for 2 fields, measured in units of dark-years (8 hours of observations per night for 365 days of which 1/3rd is lost due to weather and instrument overheads). The extrapolation of the success rate beyond ~ 7 hours assumes the SNR continues to follow expectations for background-dominated observations and is not affected by instrumental systematics.

sess the impact of such failures for DESI-Deep-like surveys under two extreme scenarios that represent best- and worst-cases for occurrences in redshift failures.

Specifically, we use the Fisher forecasting framework developed in T. Zhang et al. (2025) to predict the Dark Energy Task Force Figure of Merit (FoM; A. Albrecht et al. 2006) for LSST Year-1 and Year 10 3×2 pt analyses (i.e., the combination of weak lensing, galaxy-galaxy lensing, and galaxy clustering measurements) in the w_0-w_a CDM model. We calculate the FoM as the inverse of the square root of the determinant of the covariance matrix for the w_0 and w_a parameters to be inferred in a cosmological analysis. It is therefore proportional to the Dark Energy Task Force Figure of Merit defined as the reciprocal of the area of the error ellipse enclosing the 95% confidence limit in the w_0-w_a plane. A higher value of FoM corresponds to a lower area for the error ellipse and corresponds to a cosmological experiment with greater constraining power. Our forecasts use a fiducial cosmology based on Planck 2018 Λ CDM parameters (Planck Collaboration et al. 2020), and the baseline source redshift distribution, $n(z)$, excludes catastrophic

outliers, following T. Zhang et al. (2025) and LSST Dark Energy Science Collaboration et al. (2018).

The forecasts are computed for various redshift failure scenarios by modifying the effective number density, n_{eff} , of source galaxies in the five tomographic bins assumed to be used for the 3×2 pt analysis. We assume that the reduction of n_{eff} is entirely due to the impact of redshift measurement failures, so that the value of n_{eff} relative to the ideal case corresponds to the success rate of the redshift measurements. This corresponds to the assumption that we can (e.g., based on colors) separate out the population of galaxies which has complete redshift characterization from spectroscopy, and that lensing and clustering analyses would only use those objects; thus, a survey with higher redshift measurement success rates would enable lensing analyses that utilize more objects. Although this constitutes a strong assumption, the availability of high SNR photometry across approximately six bands, as will be provided by LSST, should render such characterization feasible.

Specifically, we scale the covariance matrix for LSST 3×2 pt analyses assuming that shape noise dominates the lensing contribution; in that case, the noise per shear component scales as $1/\sqrt{n_{\text{eff}}}$. We assume five tomographic source bins divided such that each contains the same number of (pre-failure) galaxies, and redshift failures in one bin are treated as independent of the others. The priors on redshift bias and scatter are fixed across all scenarios, derived from $N_{\text{samples}} = 5000$ reference redshift samples (see Section 3.5 of T. Zhang et al. 2025 for details). The forecasts include a redshift-dependent intrinsic alignment model and marginalize over intrinsic alignment, photometric redshift, and galaxy bias parameters when computing the FoM on cosmological parameters.

We forecast the relative FoM under two extreme scenarios; realistic observations should yield results somewhere between these possibilities. In the first, redshift failures occur evenly across all tomographic bins, resulting in a uniform decrease in $n_{i,\text{eff}}$ for all bins. This represents a best-case scenario in which redshift failures occur randomly and equally at all redshifts. In the second case, all failures are confined to the highest redshift bins; e.g., we assume that all redshift failures happen either in the last redshift bin; or, if the failure rate exceeds the number of objects in that bin, that they occur in the last two bins, or the last three bins if the number is even greater; etc. For failure rates that are not a multiple of 20%, we interpolate between the FoM values for 0%, 20%, 40%, etc. failure rates (i.e., the cases where a bin is entirely lost); this approximates the results from dropping the highest-redshift objects from analysis en-

tirely, rather than still using a subset of objects at very high z . This scenario represents the other extreme where redshift failures are all concentrated at the highest redshifts, potentially causing entire redshift ranges to be impossible to utilize in analyses. In all cases, we define the relative FoM as the ratio between the FoM when $n_{i,\text{eff}}$ is reduced to the FoM when there has been no reduction.

The results of this forecasting exercise are presented in Figure 15. This figure shows the relative FoM as a function of the overall spectroscopic success rate. We show the FoM degradation for the best-case and worst-case failure scenarios, evaluated for both the LSST Year-1 and Year-10 survey depths. Using the predictive model developed in this work, we can directly link the success rate to a required survey exposure time and thereby connect the observing time invested in a photo- z survey directly to its impact on cosmological constraining power. While this plot informs the design of our proposed DESI-Deep survey, it also serves as a general tool for forecasting the cosmological impact of redshift failures for other future spectroscopic photo- z training surveys.

We note that for an LSST Year-1 like analysis, achieving an 80% spectroscopic success rate as is the goal for the DESI-Deep survey (Section 5.3) results in a relative FoM value between 0.96 (for uniform failures) and 0.64 (for worst-case failures). Assuming a roughly circular error ellipse in the w_0 - w_a plane, this FoM degradation corresponds to an increase in the size of cosmological parameter error bars of 2% to 25% compared to an ideal survey with no redshift failures. This level of statistical uncertainty is well within the error budget for current and near-term weak lensing experiments, demonstrating that a spectroscopic data set from the proposed DESI-Deep survey provides a powerful and feasible path toward robust cosmological constraints. Future surveys with longer exposure times and correspondingly higher completeness rates can further reduce the impact of failures.

6. SUMMARY

In this work, we have presented the results from the DESI-Deep pilot program, which was designed to obtain deep spectroscopy for a faint, magnitude-limited sample of galaxies ($22 \leq m_i \leq 24.5$). Our goal was to assess DESI’s redshift measurement performance for such samples and its potential to provide the crucial, representative spectroscopic samples needed for photometric redshift training and calibration for next-generation imaging surveys such as LSST. We find that DESI is remarkably efficient at this task; with just two hours of exposure time, it can match or exceed the redshift success

rates of the DEEP2 and DEEP3 surveys conducted with the DEIMOS spectrograph on the Keck 10m telescope. Furthermore, we demonstrated that the signal-to-noise ratio of the spectra scales as expected for background-limited observations even for our longest exposures (~ 7 hours) and faintest targets. This indicates that DESI has not yet hit a systematic floor and that even deeper observations are feasible. We attribute this high performance to a combination of DESI’s excellent throughput, superior astrometry, advanced data processing algorithms, and a dithering strategy that minimizes instrumental systematics.

Building on these results, we have developed a quantitative model that predicts the redshift success rate as a function of galaxy magnitude and exposure time. Using this model, we have updated estimates of the time required to obtain a benchmark spectroscopic sample for the full 10-year LSST dataset, finding that a Stage V spectroscopic facility with DESI-like performance but a larger mirror (e.g., Spec-S5) could achieve this in as few as 50 dark nights. We propose a dedicated “DESI-Deep Survey” to obtain a more limited version of such a sample, targeting $\sim 20,000$ galaxies down to $m_i = 24.5$ in two LSST Deep Drilling Fields. A cosmological forecasting analysis shows that the resulting dataset, even with realistic redshift failure rates, would be transformative for cosmology, dramatically improving the precision of cosmological constraints from LSST and providing a lasting legacy resource for all of extragalactic astronomy.

DATA AVAILABILITY

Code and data used in this work are available at: <https://biprateep.github.io/desi-deep-pilot/>. At present, only the targeting, redshift, and visual-inspection catalogs, along with spectral-derived data products, are being made available. The raw spectra will be distributed as part of the DESI Data Release 2. The redshift catalog for this data set is also included in the compilation of spectroscopic redshifts in the DESI-COSMOS and DESI-XMMLSS fields presented by J. Ratajczak et al. (2026).

ACKNOWLEDGMENTS

B. Dey is a postdoctoral fellow at the University of Toronto in the Eric and Wendy Schmidt AI in Science Postdoctoral Fellowship Program, a program of Schmidt Sciences. B. Dey and B. H. Andrews acknowledge the support of the National Science Foundation under Grant No. AST-2009251. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the

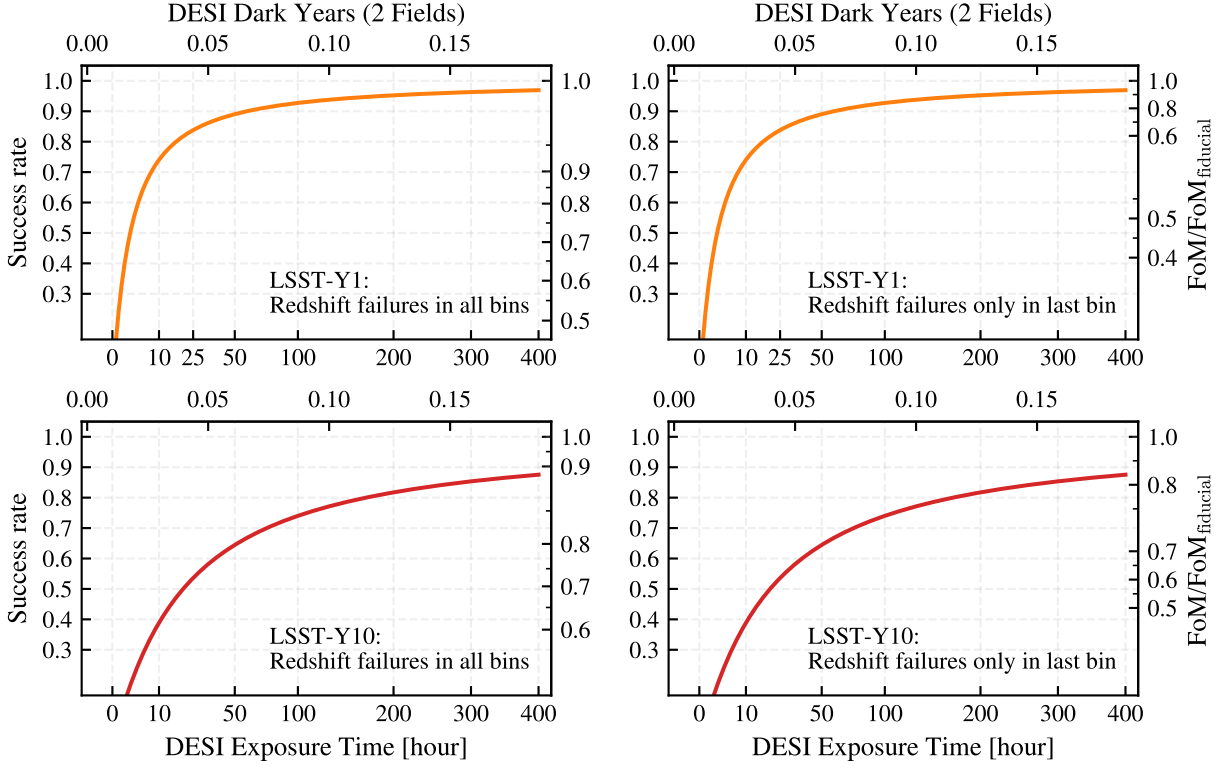


Figure 15. The relationship between exposure time with the DESI instrument, redshift measurement success rate, and the LSST survey cosmological constraining power as quantified by the relative DETF figure of merit ($\text{FoM}/\text{FoM}_{\text{fiducial}}$), degraded due to parts of parameter space lacking calibration redshifts due to spectroscopic measurement failures. The top panels correspond to an LSST Year-1-like 3×2 pt (weak lensing plus galaxy clustering) analysis, while the bottom panels correspond to an LSST Year-10-like analysis. The left panels show the predicted redshift success rate (left-vertical axis) or relative FoM (right-vertical axis) as a function of DESI exposure time (lower-horizontal axis) for the case where redshift failures occur randomly at all redshifts, while the right panels display the same but for the case where the redshift failures only occur at the very highest redshifts. The left panels correspond to the best case scenario for a spectroscopic survey while the right panels correspond to a worst case scenario; real surveys will fall somewhere in between these limits. We have added the total time required for a (field-limited) spectroscopic survey on the upper-horizontal axis of each of these plots for easy comparison to the discussions in Section 5.2. This figure directly connects the investment in spectroscopic observing time to its impact on cosmology, providing a tool to assess the trade-offs in designing a spectroscopic survey to obtain data for photo- z training and calibration. For an LSST Year-1 like analysis, achieving an 80% spectroscopic success rate as is the goal for the DESI-Deep survey (Section 5.3), results in a relative FoM value between 0.96 (for uniform failures) and 0.64 (for worst-case failures). Assuming a roughly circular error ellipse in the w_0 - w_a plane, this FoM degradation corresponds to an increase in the size of cosmological parameter error bars of 2% to 25% compared to an ideal survey with no redshift failures. This level of statistical uncertainty is well within the error budget for current and near-term weak lensing experiments.

views of the National Science Foundation. The efforts of J. A. Newman were supported by grant DE-SC0007914 from the U.S. Department of Energy Office of Science, Office of High Energy Physics.

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 (CONACYT); 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 I’oligam Du’ag (Kitt Peak), a mountain with particular significance to the Tohono O’odham Nation.

The Hyper Suprime-Cam (HSC) collaboration includes the astronomical communities of Japan and Taiwan, and Princeton University. The HSC instrumentation and software were developed by the National Astronomical Observatory of Japan (NAOJ), the Kavli Institute for the Physics and Mathematics of the Universe (Kavli IPMU), the University of Tokyo, the High Energy Accelerator Research Organization (KEK), the Academia Sinica Institute for Astronomy and Astrophysics in Taiwan (ASIAA), and Princeton University. Funding was contributed by the FIRST program from the Japanese Cabinet Office, the Ministry of Education, Culture, Sports, Science and Technology (MEXT), the Japan Society for the Promotion of Science (JSPS), Japan Science and Technology Agency (JST), the Toray Science Foundation, NAOJ, Kavli IPMU, KEK, ASIAA, and Princeton University.

This paper makes use of software developed for Vera C. Rubin Observatory. We thank the Rubin Observa-

tory for making their code available as free software at <http://pipelines.lsst.io/>.

This paper is based on data collected at the Subaru Telescope and retrieved from the HSC data archive system, which is operated by the Subaru Telescope and Astronomy Data Center (ADC) at NAOJ. Data analysis was in part carried out with the cooperation of Center for Computational Astrophysics (CfCA), NAOJ. We are honored and grateful for the opportunity of observing the Universe from Maunakea, which has the cultural, historical and natural significance in Hawaii.

This research was supported by the Munich Institute for Astro-, Particle and BioPhysics (MIAPbP) which is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy – EXC-2094 – 390783311.

The authors advocate for the judicious and ethical use of artificial intelligence (AI) based tools in scientific discourse. In accordance with the [editorial guidelines](#) of the American Astronomical Society on the appropriate use of AI-based writing tools, the authors confirm that AI assistance (with Google Gemini; [Gemini Team et al. 2023](#)) was utilized in the preparation of this manuscript solely to polish, condense, and edit our original research. The authors retain the full responsibility for the accuracy, integrity, and content of this work.

Facilities: Dark Energy Spectroscopic Instrument (DESI; [DESI Collaboration et al. 2022](#)), Hyper Suprime-Cam (HSC; [S. Miyazaki et al. 2018](#)), DEIMOS ([S. M. Faber et al. 2003](#)).

Software: Astropy ([Astropy Collaboration et al. 2013, 2018, 2022](#)), Google Gemini ([Gemini Team et al. 2023](#)), Matplotlib ([J. D. Hunter 2007](#)), Numpy ([C. R. Harris et al. 2020](#)), Pandas ([Wes McKinney 2010](#); [The pandas development team 2020](#)), Scikit-Learn ([F. Pedregosa et al. 2011](#)), Scipy ([P. Virtanen et al. 2020](#)), Statsmodels ([S. Seabold & J. Perktold 2010](#)).

APPENDIX

A. ADDITIONAL OBSERVATIONS IN THE DESI-HERCULES FIELD

As part of the DESI-Deep pilot program, we conducted additional observations in the DESI-Hercules field (defined in section 2). Due to the low quality of the data obtained in this field, it was not used for the results in the main body of the article. In this appendix, we document this sample and present selected results which include this dataset along with the other two fields.

The targets in the DESI-Hercules field were selected based upon HSC photometry using the same criteria as applied in the other regions, and all data processing and analysis procedures were identical to those used for our other results. Figure 16 shows the on-sky locations of the observed targets in all three fields, along with the effective exposure times per object. The empty regions in the focal plane correspond to spectrographs that were unavailable during these observations. The data collected came from spectrographs that had not been cooled to their optimal operating temperatures.

Figures 17 and 18 show DESI’s redshift measurement success rate and the ratio of empirical to expected SNR, respectively, as a function of magnitude and exposure time, combining data from all three fields. Including these additional observations increases the total number of targets to approximately 7,370, which reduces the statistical errors on our binned measurements. However, the lower quality of this dataset also reduces the overall redshift measurement efficiency and weakens the SNR scaling due to the known issues with the Hercules data. Despite this, all main conclusions drawn in this work remain unchanged even after including this additional data.

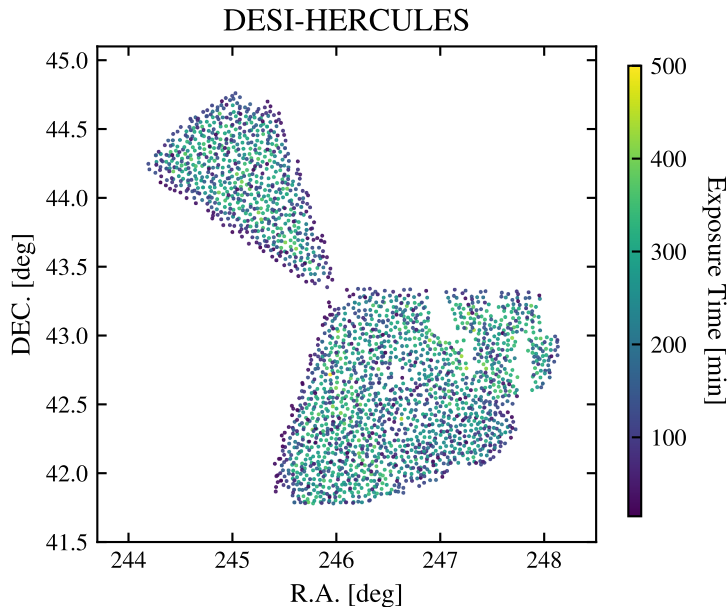


Figure 16. Locations of the 2511 objects observed in the DESI-Hercules ($244^{\circ} \leq \text{R.A.} \leq 248.5^{\circ}$; $41.5^{\circ} \leq \text{DEC.} \leq 45^{\circ}$) field. The colors indicate the total effective exposure time for the object. The empty region in the focal plane is due to certain spectrographs not being operational.

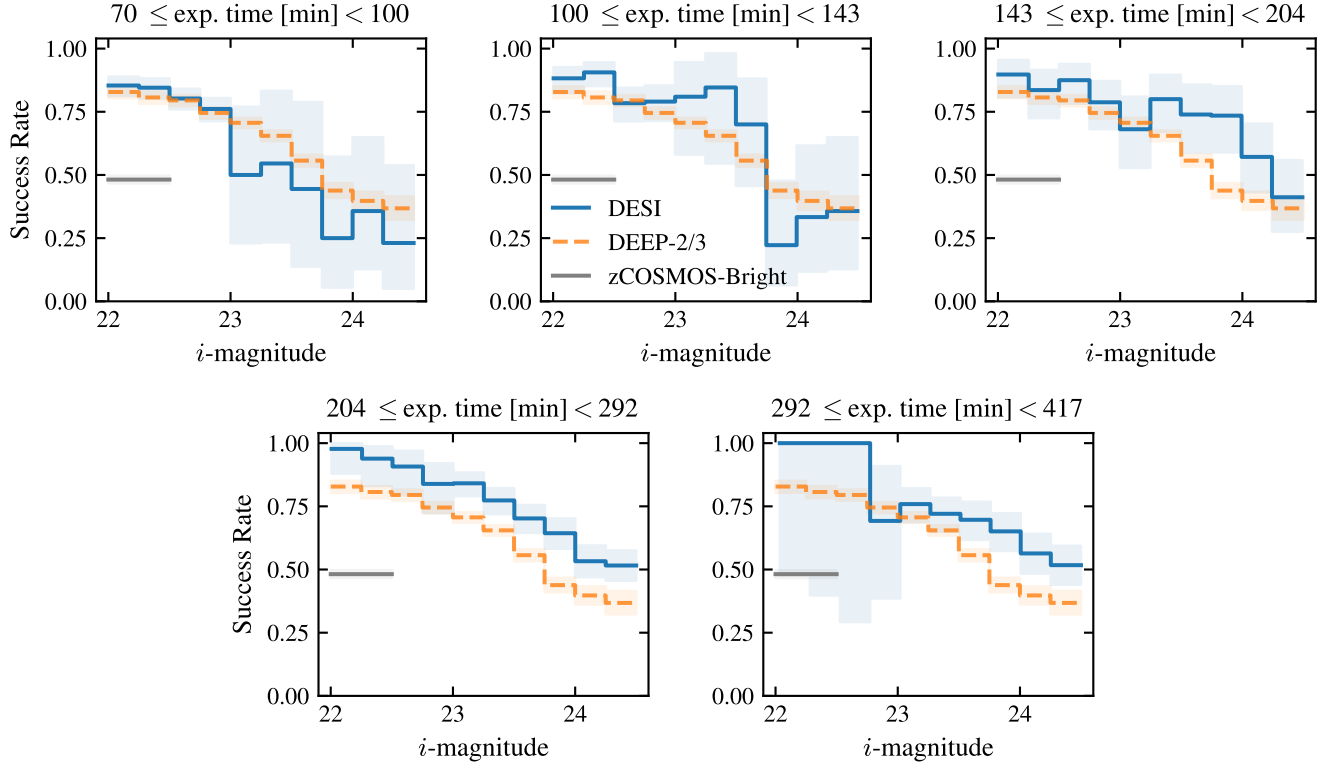


Figure 17. As Figure 7, but with data from the DESI-Hercules field included. The plot shows measurement success rate for the full DESI-Deep pilot sample as a function of i -band magnitude, with separate panels corresponding to different ranges of effective exposure time. The blue curve shows the success rate obtained from our observations, with the 95% confidence interval shown using the shaded blue region. The success rates as a function of i -magnitude for the combined DEEP2 and DEEP3 survey datasets are plotted in orange for comparison, with the shaded orange area showing the corresponding 95% confidence interval. The gray line denotes the average success rate for the zCOSMOS-Bright sample for objects with i -band magnitudes between 22 and 22.5. Compared to Figure 7, the inclusion of DESI-HERCULES data lowers the standard errors on the success rate measurements, but results in slightly lower success rates. However, at exposure times of about 2 hours, DESI’s success rate still matches that of the DEEP2 and DEEP3 galaxy redshift surveys.

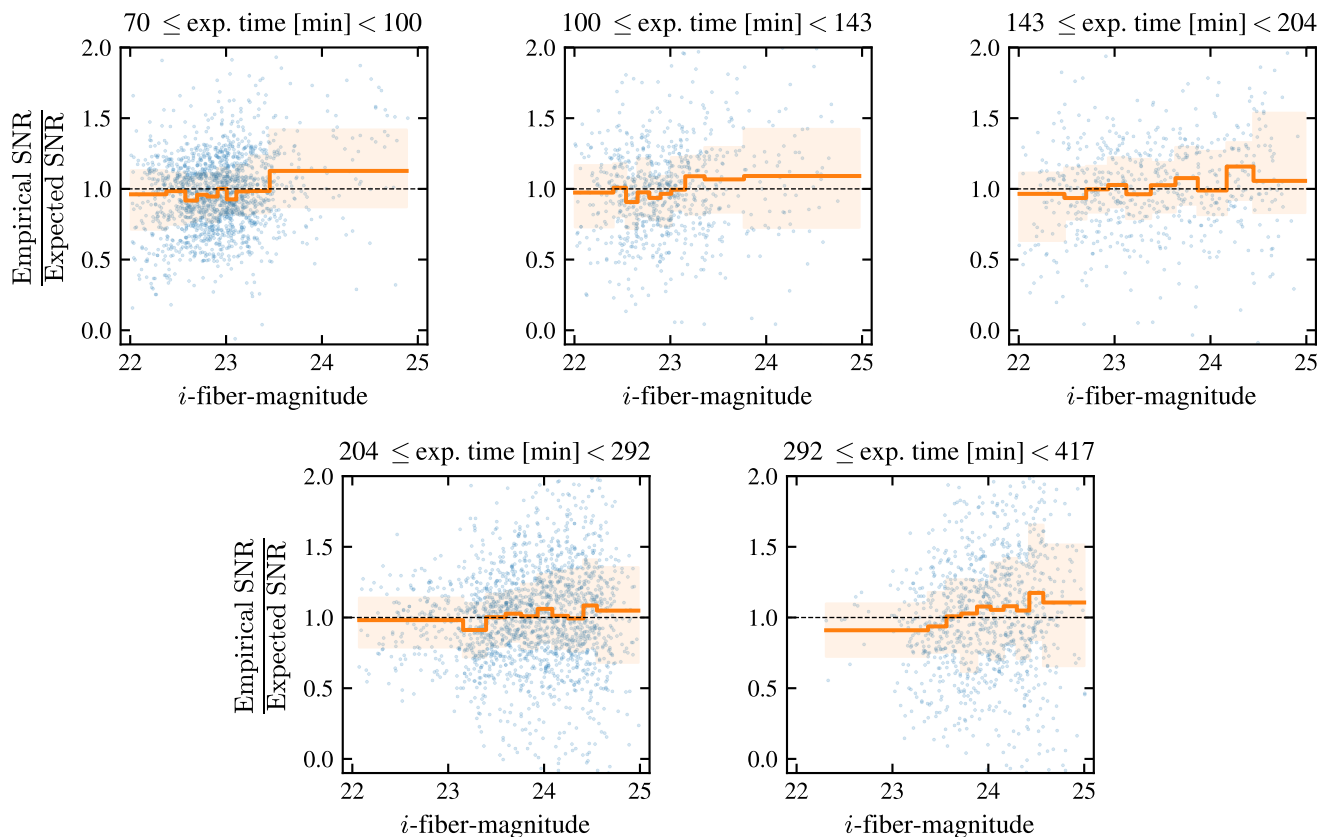


Figure 18. As Figure 11, but with data from the DESI-Hercules field included. The ratio of empirically measured SNR and what we would expect from a background limited regime plotted as a function of i -fiber-magnitude in bins of exposure time for the full DESI-Deep pilot sample. Each blue dot represents a single object and the orange curve denotes the average in ten equal population bins. The shaded orange region shows the 95% confidence interval on the mean. Compared to Figure 11 we observe a slightly worse scaling of the SNR especially at the fainter magnitudes. However, this is expected, given the non-optimal status of the instrument during the DESI-HERCULES observations.


































































REFERENCES

- Aihara, H., Arimoto, N., Armstrong, R., et al. 2018a, PASJ, 70, S4, doi: [10.1093/pasj/psx066](https://doi.org/10.1093/pasj/psx066)
- Aihara, H., Arimoto, N., Armstrong, R., et al. 2018b, PASJ, 70, S4, doi: [10.1093/pasj/psx066](https://doi.org/10.1093/pasj/psx066)
- Aihara, H., AlSayyad, Y., Ando, M., et al. 2022, PASJ, 74, 247, doi: [10.1093/pasj/psab122](https://doi.org/10.1093/pasj/psab122)
- Akeson, R., Armus, L., Bachelet, E., et al. 2019, arXiv e-prints, arXiv:1902.05569, doi: [10.48550/arXiv.1902.05569](https://doi.org/10.48550/arXiv.1902.05569)
- Albrecht, A., Bernstein, G., Cahn, R., et al. 2006, arXiv e-prints, astro, doi: [10.48550/arXiv.astro-ph/0609591](https://doi.org/10.48550/arXiv.astro-ph/0609591)
- Alexander, D. M., Davis, T. M., Chaussidon, E., et al. 2023, AJ, 165, 124, doi: [10.3847/1538-3881/acacfc](https://doi.org/10.3847/1538-3881/acacfc)
- Annis, J., Newman, J. A., & Slosar, A. 2022, arXiv e-prints, arXiv:2209.08049. <https://arxiv.org/abs/2209.08049>
- Arnouts, S., Moscardini, L., Vanzella, E., et al. 2002, MNRAS, 329, 355, doi: [10.1046/j.1365-8711.2002.04988.x](https://doi.org/10.1046/j.1365-8711.2002.04988.x)
- Asai, S., Ballarino, A., Bose, T., et al. 2024, arXiv e-prints, arXiv:2407.19176, doi: [10.48550/arXiv.2407.19176](https://doi.org/10.48550/arXiv.2407.19176)
- Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33, doi: [10.1051/0004-6361/201322068](https://doi.org/10.1051/0004-6361/201322068)
- Astropy Collaboration, Price-Whelan, A. M., SipHocz, B. M., et al. 2018, aj, 156, 123, doi: [10.3847/1538-3881/aabc4f](https://doi.org/10.3847/1538-3881/aabc4f)
- Astropy Collaboration, Price-Whelan, A. M., Lim, P. L., et al. 2022, ApJ, 935, 167, doi: [10.3847/1538-4357/ac7c74](https://doi.org/10.3847/1538-4357/ac7c74)
- Bacon, R., Mauneiri, V., Randich, S., et al. 2024, arXiv e-prints, arXiv:2405.12518, doi: [10.48550/arXiv.2405.12518](https://doi.org/10.48550/arXiv.2405.12518)
- Bailey, S., DESI Collaboration, & et al. in prep.
- Baum, W. A. 1957, AJ, 62, 6, doi: [10.1086/107433](https://doi.org/10.1086/107433)
- Besuner, R., Dey, A., Drlica-Wagner, A., et al. 2025, arXiv e-prints, arXiv:2503.07923, doi: [10.48550/arXiv.2503.07923](https://doi.org/10.48550/arXiv.2503.07923)
- Blanton, M. R. 2006, ApJ, 648, 268, doi: [10.1086/505628](https://doi.org/10.1086/505628)
- Blazek, J. A., Clowe, D., Collett, T. E., et al. 2022, arXiv e-prints, arXiv:2204.01992. <https://arxiv.org/abs/2204.01992>
- Bolton, A. S., & Schlegel, D. J. 2010, PASP, 122, 248, doi: [10.1086/651008](https://doi.org/10.1086/651008)
- Brescia, M., Cavuoti, S., Razim, O., et al. 2021, Frontiers in Astronomy and Space Sciences, 8, 70, doi: [10.3389/fspas.2021.658229](https://doi.org/10.3389/fspas.2021.658229)
- Bundy, K., Westfall, K., MacDonald, N., et al. 2019, in Bulletin of the American Astronomical Society, Vol. 51, 198, doi: [10.48550/arXiv.1907.07195](https://doi.org/10.48550/arXiv.1907.07195)
- Bundy, K., Law, D., MacDonald, N., et al. 2022, AJ, 164, 94, doi: [10.3847/1538-3881/ac76cc](https://doi.org/10.3847/1538-3881/ac76cc)
- Cirasuolo, M., Fairley, A., Rees, P., et al. 2020, The Messenger, 180, 10, doi: [10.18727/0722-6691/5195](https://doi.org/10.18727/0722-6691/5195)
- Clopper, C. J., & Pearson, E. S. 1934, Biometrika, 26, 404
- Coil, A. L., Newman, J. A., Kaiser, N., et al. 2004, ApJ, 617, 765, doi: [10.1086/425676](https://doi.org/10.1086/425676)
- Comparat, J., Kneib, J.-P., Bacon, R., et al. 2013, A&A, 559, A18, doi: [10.1051/0004-6361/201322452](https://doi.org/10.1051/0004-6361/201322452)
- de Jong, R. S., Agertz, O., Berbel, A. A., et al. 2019, The Messenger, 175, 3, doi: [10.18727/0722-6691/5117](https://doi.org/10.18727/0722-6691/5117)
- DePoy, D. L., Schmidt, L. M., Ribeiro, R., et al. 2018, in Ground-based and Airborne Instrumentation for Astronomy VII, ed. C. J. Evans, L. Simard, & H. Takami, Vol. 10702, International Society for Optics and Photonics (SPIE), 107021X, doi: [10.1117/12.2313940](https://doi.org/10.1117/12.2313940)
- DESI Collaboration, Aghamousa, A., Aguilar, J., et al. 2016a, arXiv e-prints, arXiv:1611.00036, doi: [10.48550/arXiv.1611.00036](https://doi.org/10.48550/arXiv.1611.00036)
- DESI Collaboration, Aghamousa, A., Aguilar, J., et al. 2016b, arXiv e-prints, arXiv:1611.00037, doi: [10.48550/arXiv.1611.00037](https://doi.org/10.48550/arXiv.1611.00037)
- DESI Collaboration, Abareshi, B., Aguilar, J., et al. 2022, AJ, 164, 207, doi: [10.3847/1538-3881/ac882b](https://doi.org/10.3847/1538-3881/ac882b)
- DESI Collaboration, Adame, A. G., Aguilar, J., et al. 2024, AJ, 167, 62, doi: [10.3847/1538-3881/ad0b08](https://doi.org/10.3847/1538-3881/ad0b08)
- DESI Collaboration, Abdul-Karim, M., Adame, A. G., et al. 2025a, arXiv e-prints, arXiv:2503.14745, doi: [10.48550/arXiv.2503.14745](https://doi.org/10.48550/arXiv.2503.14745)
- DESI Collaboration, Adame, A. G., Aguilar, J., et al. 2025b, JCAP, 2025, 028, doi: [10.1088/1475-7516/2025/07/028](https://doi.org/10.1088/1475-7516/2025/07/028)
- DESI Collaboration, Abdul Karim, M., Aguilar, J., et al. 2025c, PhRvD, 112, 083515, doi: [10.1103/tr6y-kpc6](https://doi.org/10.1103/tr6y-kpc6)
- Efron, B. 1979, The Annals of Statistics, 7, 1. <http://www.jstor.org/stable/2958830>
- Faber, S. M., Phillips, A. C., Kibrick, R. I., et al. 2003, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 4841, Instrument Design and Performance for Optical/Infrared Ground-based Telescopes, ed. M. Iye & A. F. M. Moorwood, 1657–1669, doi: [10.1117/12.460346](https://doi.org/10.1117/12.460346)
- Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2016, A&A, 595, A2, doi: [10.1051/0004-6361/201629512](https://doi.org/10.1051/0004-6361/201629512)
- Gemini Team, Anil, R., Borgeaud, S., et al. 2023, arXiv e-prints, arXiv:2312.11805, doi: [10.48550/arXiv.2312.11805](https://doi.org/10.48550/arXiv.2312.11805)
- Gruen, D., & Brimiouille, F. 2017, MNRAS, 468, 769, doi: [10.1093/mnras/stx471](https://doi.org/10.1093/mnras/stx471)
- Gruen, D., McCullough, J., Amon, A., et al. 2023, The Messenger, 190, 28, doi: [10.18727/0722-6691/5307](https://doi.org/10.18727/0722-6691/5307)

- Guy, J., Bailey, S., Kremin, A., et al. 2023, *AJ*, 165, 144, doi: [10.3847/1538-3881/acb212](https://doi.org/10.3847/1538-3881/acb212)
- Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, *Nature*, 585, 357–362, doi: [10.1038/s41586-020-2649-2](https://doi.org/10.1038/s41586-020-2649-2)
- Hart, M. 2019, *AJ*, 157, 221, doi: [10.3847/1538-3881/ab1a35](https://doi.org/10.3847/1538-3881/ab1a35)
- Hartley, W. G., Chang, C., Samani, S., et al. 2020, *MNRAS*, 496, 4769, doi: [10.1093/mnras/staa1812](https://doi.org/10.1093/mnras/staa1812)
- Horne, K. 1986, *Publications of the Astronomical Society of the Pacific*, 98, 609, doi: [10.1086/131801](https://doi.org/10.1086/131801)
- Hunter, J. D. 2007, *Computing in Science Engineering*, 9, 90
- Ilbert, O., Arnouts, S., McCracken, H. J., et al. 2006, *A&A*, 457, 841, doi: [10.1051/0004-6361:20065138](https://doi.org/10.1051/0004-6361:20065138)
- Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, *ApJ*, 873, 111, doi: [10.3847/1538-4357/ab042c](https://doi.org/10.3847/1538-4357/ab042c)
- Jin, S., Trager, S. C., Dalton, G. B., et al. 2024, *MNRAS*, 530, 2688, doi: [10.1093/mnras/stad557](https://doi.org/10.1093/mnras/stad557)
- Kirkby, D., DESI Collaboration, & et al. in prep.
- Lan, T.-W., Tojeiro, R., Armengaud, E., et al. 2023, *ApJ*, 943, 68, doi: [10.3847/1538-4357/aca5fa](https://doi.org/10.3847/1538-4357/aca5fa)
- Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, arXiv e-prints, arXiv:1110.3193, doi: [10.48550/arXiv.1110.3193](https://doi.org/10.48550/arXiv.1110.3193)
- Le Fèvre, O., Saisse, M., Mancini, D., et al. 2003, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 4841, Instrument Design and Performance for Optical/Infrared Ground-based Telescopes, ed. M. Iye & A. F. M. Moorwood, 1670–1681, doi: [10.1117/12.460959](https://doi.org/10.1117/12.460959)
- Levi, M., Bebek, C., Beers, T., et al. 2013, arXiv e-prints, arXiv:1308.0847, doi: [10.48550/arXiv.1308.0847](https://doi.org/10.48550/arXiv.1308.0847)
- Lidman, C., Tucker, B. E., Davis, T. M., et al. 2020, *MNRAS*, 496, 19, doi: [10.1093/mnras/staa1341](https://doi.org/10.1093/mnras/staa1341)
- Lilly, S. J., Le Fèvre, O., Renzini, A., et al. 2007, *ApJS*, 172, 70, doi: [10.1086/516589](https://doi.org/10.1086/516589)
- Lilly, S. J., Le Brun, V., Maier, C., et al. 2009, *ApJS*, 184, 218, doi: [10.1088/0067-0049/184/2/218](https://doi.org/10.1088/0067-0049/184/2/218)
- LSST Dark Energy Science Collaboration, Mandelbaum, R., Eifler, T., et al. 2018, arXiv e-prints, arXiv:1809.01669. <https://arxiv.org/abs/1809.01669>
- LSST Science Collaboration, Abell, P. A., Allison, J., et al. 2009, arXiv e-prints, arXiv:0912.0201, doi: [10.48550/arXiv.0912.0201](https://doi.org/10.48550/arXiv.0912.0201)
- Mandelbaum, R., Blazek, J., Chisari, N. E., et al. 2019, *BAAS*, 51, 363. <https://arxiv.org/abs/1903.09323>
- Masters, D. C., Stern, D. K., Cohen, J. G., et al. 2019, *ApJ*, 877, 81, doi: [10.3847/1538-4357/ab184d](https://doi.org/10.3847/1538-4357/ab184d)
- Matthews, D. J., Newman, J. A., Coil, A. L., Cooper, M. C., & Gwyn, S. D. J. 2013, *ApJS*, 204, 21, doi: [10.1088/0067-0049/204/2/21](https://doi.org/10.1088/0067-0049/204/2/21)
- McCullagh, P., & Nelder, J. A. 1989, *Generalized Linear Models* (London: Chapman & Hall / CRC)
- McCullough, J., Gruen, D., Amon, A., et al. 2024, *MNRAS*, 531, 2582, doi: [10.1093/mnras/stae1316](https://doi.org/10.1093/mnras/stae1316)
- Miller, T. N., Doel, P., Gutierrez, G., et al. 2023, arXiv e-prints, arXiv:2306.06310, doi: [10.48550/arXiv.2306.06310](https://doi.org/10.48550/arXiv.2306.06310)
- Miyazaki, S., Komiyama, Y., Kawanomoto, S., et al. 2018, *PASJ*, 70, S1, doi: [10.1093/pasj/psx063](https://doi.org/10.1093/pasj/psx063)
- Monet, A., Bird, J., Canzian, B., et al. 1998, *VizieR On-line Data Catalog: I/252*. Originally published in: U.S. Naval Observatory Flagstaff Station (USNOFS) and, Universities Space Research Association (USRA) stationed at USNOFS. , (1998)
- Myles, J., Alarcon, A., Amon, A., et al. 2021, *MNRAS*, 505, 4249, doi: [10.1093/mnras/stab1515](https://doi.org/10.1093/mnras/stab1515)
- Newman, J., Blazek, J., Chisari, N. E., et al. 2019, *BAAS*, 51, 358. <https://arxiv.org/abs/1903.09325>
- Newman, J. A., & Gruen, D. 2022, *ARA&A*, 60, 363, doi: [10.1146/annurev-astro-032122-014611](https://doi.org/10.1146/annurev-astro-032122-014611)
- Newman, J. A., Cooper, M. C., Davis, M., et al. 2013, *ApJS*, 208, 5, doi: [10.1088/0067-0049/208/1/5](https://doi.org/10.1088/0067-0049/208/1/5)
- Newman, J. A., Abate, A., Abdalla, F. B., et al. 2015, *Astroparticle Physics*, 63, 81, doi: [10.1016/j.astropartphys.2014.06.007](https://doi.org/10.1016/j.astropartphys.2014.06.007)
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *Journal of Machine Learning Research*, 12, 2825
- Pelló, R., Puech, M., Prieto, É., et al. 2024, in *Ground-based and Airborne Instrumentation for Astronomy X*, ed. J. J. Bryant, K. Motohara, & J. R. D. Vernet, Vol. 13096, International Society for Optics and Photonics (SPIE), 1309615, doi: [10.1117/12.3019047](https://doi.org/10.1117/12.3019047)
- Pierre, M., Valtchanov, I., Altieri, B., et al. 2004, *JCAP*, 2004, 011, doi: [10.1088/1475-7516/2004/09/011](https://doi.org/10.1088/1475-7516/2004/09/011)
- Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, *A&A*, 641, A6, doi: [10.1051/0004-6361/201833910](https://doi.org/10.1051/0004-6361/201833910)
- Poppett, C., Tyas, L., Aguilar, J., et al. 2024, *AJ*, 168, 245, doi: [10.3847/1538-3881/ad76a4](https://doi.org/10.3847/1538-3881/ad76a4)
- Raichoor et al., A., DESI Collaboration, & et al. in prep.
- Ratajczak, J., Dawson, K. S., Weaverdyck, N., et al. 2026, *AJ*, 171, 71, doi: [10.3847/1538-3881/ae1fde](https://doi.org/10.3847/1538-3881/ae1fde)
- Salvato, M., Ilbert, O., & Hoyle, B. 2019, *Nature Astronomy*, 3, 212, doi: [10.1038/s41550-018-0478-0](https://doi.org/10.1038/s41550-018-0478-0)
- Schlafly, E. F., Kirkby, D., Schlegel, D. J., et al. 2023, *AJ*, 166, 259, doi: [10.3847/1538-3881/ad0832](https://doi.org/10.3847/1538-3881/ad0832)
- Schlegel, D., DESI Collaboration, & et al. in prep.

- Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, *ApJ*, 500, 525, doi: [10.1086/305772](https://doi.org/10.1086/305772)
- Schlegel, D. J., Ferraro, S., Aldering, G., et al. 2022a, arXiv e-prints, arXiv:2209.03585.
<https://arxiv.org/abs/2209.03585>
- Schlegel, D. J., Kollmeier, J. A., Aldering, G., et al. 2022b, arXiv e-prints, arXiv:2209.04322, doi: [10.48550/arXiv.2209.04322](https://doi.org/10.48550/arXiv.2209.04322)
- Scoville, N., Aussel, H., Brusa, M., et al. 2007, *ApJS*, 172, 1, doi: [10.1086/516585](https://doi.org/10.1086/516585)
- Seabold, S., & Perktold, J. 2010, in 9th Python in Science Conference
- Sharp, R., & Parkinson, H. 2010, *MNRAS*, 408, 2495, doi: [10.1111/j.1365-2966.2010.17298.x](https://doi.org/10.1111/j.1365-2966.2010.17298.x)
- Silber, J. H., Fagrelus, P., Fanning, K., et al. 2023, *AJ*, 165, 9, doi: [10.3847/1538-3881/ac9ab1](https://doi.org/10.3847/1538-3881/ac9ab1)
- Steidel, C., Peng, E., Fucik, J., et al. 2022, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 12184, Ground-based and Airborne Instrumentation for Astronomy IX, ed. C. J. Evans, J. J. Bryant, & K. Motohara, 1218423, doi: [10.1117/12.2629464](https://doi.org/10.1117/12.2629464)
- Tamura, N., Takato, N., Shimono, A., et al. 2016, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9908, Ground-based and Airborne Instrumentation for Astronomy VI, ed. C. J. Evans, L. Simard, & H. Takami, 99081M, doi: [10.1117/12.2232103](https://doi.org/10.1117/12.2232103)
- The Dark Energy Survey Collaboration. 2005, arXiv e-prints, astro, doi: [10.48550/arXiv.astro-ph/0510346](https://doi.org/10.48550/arXiv.astro-ph/0510346)
- The pandas development team. 2020,, latest Zenodo, doi: [10.5281/zenodo.3509134](https://doi.org/10.5281/zenodo.3509134)
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *Nature Methods*, 17, 261, doi: [10.1038/s41592-019-0686-2](https://doi.org/10.1038/s41592-019-0686-2)
- Weaver, J. R., Kauffmann, O. B., Ilbert, O., et al. 2022, *ApJS*, 258, 11, doi: [10.3847/1538-4365/ac3078](https://doi.org/10.3847/1538-4365/ac3078)
- Weaver, J. R., Zalesky, L., Kokorev, V., et al. 2023, *ApJS*, 269, 20, doi: [10.3847/1538-4365/acf850](https://doi.org/10.3847/1538-4365/acf850)
- Wes McKinney. 2010, in Proceedings of the 9th Python in Science Conference, ed. Stéfan van der Walt & Jarrod Millman, 56 – 61, doi: [10.25080/Majora-92bf1922-00a](https://doi.org/10.25080/Majora-92bf1922-00a)
- Wright, A. H., Hildebrandt, H., van den Busch, J. L., & Heymans, C. 2020, *A&A*, 637, A100, doi: [10.1051/0004-6361/201936782](https://doi.org/10.1051/0004-6361/201936782)
- Wright, A. H., Kuijken, K., Hildebrandt, H., et al. 2024, *A&A*, 686, A170, doi: [10.1051/0004-6361/202346730](https://doi.org/10.1051/0004-6361/202346730)
- Wright, A. H., Stözlner, B., Asgari, M., et al. 2025a, arXiv e-prints, arXiv:2503.19441, doi: [10.48550/arXiv.2503.19441](https://doi.org/10.48550/arXiv.2503.19441)
- Wright, A. H., Hildebrandt, H., van den Busch, J. L., et al. 2025b, arXiv e-prints, arXiv:2503.19440, doi: [10.48550/arXiv.2503.19440](https://doi.org/10.48550/arXiv.2503.19440)
- Zafar, T., Lawrence, J., Zheng, J., et al. 2022, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 12184, Ground-based and Airborne Instrumentation for Astronomy IX, ed. C. J. Evans, J. J. Bryant, & K. Motohara, 1218417, doi: [10.1117/12.2628142](https://doi.org/10.1117/12.2628142)
- Zhang, T., Almoubayyed, H., Mandelbaum, R., et al. 2025, arXiv e-prints, arXiv:2507.01374, doi: [10.48550/arXiv.2507.01374](https://doi.org/10.48550/arXiv.2507.01374)
- Zhao, C., Huang, S., He, M., et al. 2024, arXiv e-prints, arXiv:2411.07970, doi: [10.48550/arXiv.2411.07970](https://doi.org/10.48550/arXiv.2411.07970)
- Zhou, R., Cooper, M. C., Newman, J. A., et al. 2019, *MNRAS*, 488, 4565, doi: [10.1093/mnras/stz1866](https://doi.org/10.1093/mnras/stz1866)
- Zhou, R., Dey, B., Newman, J. A., et al. 2023, *AJ*, 165, 58, doi: [10.3847/1538-3881/aca5fb](https://doi.org/10.3847/1538-3881/aca5fb)

All Authors and Affiliations

BIPRATEEP DEY ^{1,2,3,4} JEFFREY A. NEWMAN ⁴ TIANQING ZHANG ⁴ J. AGUILAR ⁵ S. AHLÉN ⁶ A. ANAND ⁵
B. ANDREWS ⁷ S. BAILEY ⁵ D. BIANCHI ^{8,9} D. BROOKS ¹⁰ F. J. CASTANDER ^{11,12} T. CLAYBAUGH ⁵
A. CUCEU ⁵ K. S. DAWSON ¹³ A. DE LA MACORRA ¹⁴ J. DELLA COSTA ¹⁵ ARJUN DEY ¹⁵ P. DOEL ¹⁰
S. FERRARO ^{5,16} A. FONT-RIBERA ^{17,18} E. GAZTAÑAGA ^{11,19,12} SATYA GONTCHO A GONTCHO ²⁰
D. GRUEN ^{21,22} G. GUTIERREZ ²³ J. GUY ⁵ H. K. HERRERA-ALCANTAR ^{24,25} K. HONSCHEID ^{26,27}
M. ISHAK ²⁸ R. JOYCE ¹⁵ R. KEHOE ²⁹ D. KIRKBY ³⁰ T. KISNER ⁵ A. KREMIN ⁵ O. LAHAV ¹⁰
M. LANDRIAU ⁵ L. LE GUILLOU ³¹ A. LEAUTHAUD ^{32,33} M. E. LEVI ⁵ M. MANERA ^{34,18} P. MARTINI ^{26,35}
J. MCCULLOUGH ³⁶ A. MEISNER ¹⁵ R. MIQUEL ^{17,18} J. MOUSTAKAS ³⁷ A. D. MYERS ³⁸ J. MYLES ³⁶
S. NADATHUR ¹⁹ N. PALANQUE-DELABROUILLE ^{25,5} W. J. PERCIVAL ^{39,40,41} F. PRADA ⁴² I. PÉREZ-RÀFOLS ⁴³
G. ROSSI ⁴⁴ L. SAMUSHIA ^{45,46,47} E. SANCHEZ ⁴⁸ D. SCHLEGEL ⁵ M. SCHUBNEL ^{49,50} H. SEO ⁵¹
J. SILBER ⁵ D. SPRAYBERRY ¹⁵ G. TARLÉ ⁵⁰ B. A. WEAVER ¹⁵ N. WEAVERDYCK ⁵ R. H. WECHSLER ^{52,53,54}
R. ZHOU ⁵ AND H. ZOU ⁵⁵

¹Department of Statistical Sciences, University of Toronto, Toronto, ON M5G 1Z5, Canada

²Canadian Institute for Theoretical Astrophysics (CITA), University of Toronto, Toronto, ON M5S 3H8

³Dunlap Institute for Astronomy & Astrophysics, University of Toronto, Toronto, ON M5S 3H4, Canada

⁴Department of Physics and Astronomy and PITT PACC, University of Pittsburgh, Pittsburgh, PA 15260, USA

⁵Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA

⁶Department of Physics, Boston University, 590 Commonwealth Avenue, Boston, MA 02215 USA

⁷Department of Physics & Astronomy and Pittsburgh Particle Physics, Astrophysics, and Cosmology Center (PITT PACC), University of Pittsburgh, 3941 O'Hara Street, Pittsburgh, PA 15260, USA

⁸Dipartimento di Fisica "Aldo Pontremoli", Università degli Studi di Milano, Via Celoria 16, I-20133 Milano, Italy

⁹INAF-Osservatorio Astronomico di Brera, Via Brera 28, 20122 Milano, Italy

¹⁰Department of Physics & Astronomy, University College London, Gower Street, London, WC1E 6BT, UK

¹¹Institut d'Estudis Espacials de Catalunya (IEEC), c/ Esteve Terradas 1, Edifici RDIT, Campus PMT-UPC, 08860 Castelldefels, Spain

¹²Institute of Space Sciences, ICE-CSIC, Campus UAB, Carrer de Can Magrans s/n, 08913 Bellaterra, Barcelona, Spain

¹³Department of Physics and Astronomy, The University of Utah, 115 South 1400 East, Salt Lake City, UT 84112, USA

¹⁴Instituto de Física, Universidad Nacional Autónoma de México, Circuito de la Investigación Científica, Ciudad Universitaria, Cd. de México C. P. 04510, México

¹⁵NSF NOIRLab, 950 N. Cherry Ave., Tucson, AZ 85719, USA

¹⁶University of California, Berkeley, 110 Sproul Hall #5800 Berkeley, CA 94720, USA

¹⁷Institució Catalana de Recerca i Estudis Avançats, Passeig de Lluís Companys, 23, 08010 Barcelona, Spain

¹⁸Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Edifici Cn, Campus UAB, 08193, Bellaterra (Barcelona), Spain

¹⁹Institute of Cosmology and Gravitation, University of Portsmouth, Dennis Sciama Building, Portsmouth, PO1 3FX, UK

²⁰University of Virginia, Department of Astronomy, Charlottesville, VA 22904, USA

²¹Excellence Cluster ORIGINS, Boltzmannstrasse 2, D-85748 Garching, Germany

²²University Observatory, Faculty of Physics, Ludwig-Maximilians-Universität, Scheinerstr. 1, 81677 München, Germany

²³Fermi National Accelerator Laboratory, PO Box 500, Batavia, IL 60510, USA

²⁴Institut d'Astrophysique de Paris. 98 bis boulevard Arago. 75014 Paris, France

²⁵IRFU, CEA, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France

²⁶Center for Cosmology and AstroParticle Physics, The Ohio State University, 191 West Woodruff Avenue, Columbus, OH 43210, USA

²⁷Department of Physics, The Ohio State University, 191 West Woodruff Avenue, Columbus, OH 43210, USA

²⁸Department of Physics, The University of Texas at Dallas, 800 W. Campbell Rd., Richardson, TX 75080, USA

²⁹Department of Physics, Southern Methodist University, 3215 Daniel Avenue, Dallas, TX 75275, USA

³⁰Department of Physics and Astronomy, University of California, Irvine, 92697, USA

³¹Sorbonne Université, CNRS/IN2P3, Laboratoire de Physique Nucléaire et de Hautes Energies (LPNHE), FR-75005 Paris, France

³²Department of Astronomy and Astrophysics, UCO/Lick Observatory, University of California, 1156 High Street, Santa Cruz, CA 95064, USA

³³Department of Astronomy and Astrophysics, University of California, Santa Cruz, 1156 High Street, Santa Cruz, CA 95065, USA

³⁴Departament de Física, Serra Hünter, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain

³⁵Department of Astronomy, The Ohio State University, 4055 McPherson Laboratory, 140 W 18th Avenue, Columbus, OH 43210, USA

³⁶Department of Astrophysical Sciences, Princeton University, Princeton NJ 08544, USA

³⁷Department of Physics and Astronomy, Siena University, 515 Loudon Road, Loudonville, NY 12211, USA

³⁸Department of Physics & Astronomy, University of Wyoming, 1000 E. University, Dept. 3905, Laramie, WY 82071, USA

³⁹*Department of Physics and Astronomy, University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1, Canada*

⁴⁰*Perimeter Institute for Theoretical Physics, 31 Caroline St. North, Waterloo, ON N2L 2Y5, Canada*

⁴¹*Waterloo Centre for Astrophysics, University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1, Canada*

⁴²*Instituto de Astrofísica de Andalucía (CSIC), Glorieta de la Astronomía, s/n, E-18008 Granada, Spain*

⁴³*Departament de Física, EEBE, Universitat Politècnica de Catalunya, c/Eduard Maristany 10, 08930 Barcelona, Spain*

⁴⁴*Department of Physics and Astronomy, Sejong University, 209 Neungdong-ro, Gwangjin-gu, Seoul 05006, Republic of Korea*

⁴⁵*Abastumani Astrophysical Observatory, Tbilisi, GE-0179, Georgia*

⁴⁶*Department of Physics, Kansas State University, 116 Cardwell Hall, Manhattan, KS 66506, USA*

⁴⁷*Faculty of Natural Sciences and Medicine, Ilia State University, 0194 Tbilisi, Georgia*

⁴⁸*CIEMAT, Avenida Complutense 40, E-28040 Madrid, Spain*

⁴⁹*Department of Physics, University of Michigan, 450 Church Street, Ann Arbor, MI 48109, USA*

⁵⁰*University of Michigan, 500 S. State Street, Ann Arbor, MI 48109, USA*

⁵¹*Department of Physics & Astronomy, Ohio University, 139 University Terrace, Athens, OH 45701, USA*

⁵²*Kavli Institute for Particle Astrophysics and Cosmology, Stanford University, Menlo Park, CA 94305, USA*

⁵³*Physics Department, Stanford University, Stanford, CA 93405, USA*

⁵⁴*SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, USA*

⁵⁵*National Astronomical Observatories, Chinese Academy of Sciences, A20 Datun Road, Chaoyang District, Beijing, 100101, P. R. China*