



The Complete Calibration of the Color–Redshift Relation (C3R2) Survey: Survey Overview and Data Release 1

Daniel C. Masters¹, Daniel K. Stern², Judith G. Cohen³, Peter L. Capak⁴, Jason D. Rhodes^{2,5}, Francisco J. Castander⁶, and Stéphane Paltani⁷

¹Infrared Processing and Analysis Center, Pasadena, CA 91125, USA

²Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA

³California Institute of Technology, Pasadena, CA 91125, USA

⁴Spitzer Science Center, Pasadena, CA 91125, USA

⁵Kavli Institute for the Physics and Mathematics of the Universe, University of Tokyo, Chiba 277-8582, Japan

⁶Institut de Ciències de l'Espai (ICE, IEEC/CSIC), E-08193 Bellaterra (Barcelona), Spain

⁷Department of Astronomy, University of Geneva, Ch. d'Ecogia 16, 1290 Versoix, Switzerland

Received 2017 February 27; revised 2017 April 19; accepted 2017 April 19; published 2017 May 31

Abstract

A key goal of the Stage IV dark energy experiments *Euclid*, LSST, and *WFIRST* is to measure the growth of structure with cosmic time from weak lensing analysis over large regions of the sky. Weak lensing cosmology will be challenging: in addition to highly accurate galaxy shape measurements, statistically robust and accurate photometric redshift (photo- z) estimates for billions of faint galaxies will be needed in order to reconstruct the three-dimensional matter distribution. Here we present an overview of and initial results from the Complete Calibration of the Color–Redshift Relation (C3R2) survey, which is designed specifically to calibrate the empirical galaxy color–redshift relation to the *Euclid* depth. These redshifts will also be important for the calibrations of LSST and *WFIRST*. The C3R2 survey is obtaining multiplexed observations with Keck (DEIMOS, LRIS, and MOSFIRE), the Gran Telescopio Canarias (GTC; OSIRIS), and the Very Large Telescope (VLT; FORS2 and KMOS) of a targeted sample of galaxies that are most important for the redshift calibration. We focus spectroscopic efforts on undersampled regions of galaxy color space identified in previous work in order to minimize the number of spectroscopic redshifts needed to map the color–redshift relation to the required accuracy. We present the C3R2 survey strategy and initial results, including the 1283 high-confidence redshifts obtained in the 2016A semester and released as Data Release 1.

Key words: astronomical databases: catalogs – astronomical databases: surveys – cosmology: observations – galaxies: distances and redshifts

Supporting material: machine-readable table

1. Introduction

The upcoming large-scale cosmology experiments *Euclid* (Laureijs et al. 2011), LSST (Ivezic et al. 2008), and *WFIRST* (Spergel et al. 2015) will depend on robust photometric redshift (photo- z) estimates for billions of faint galaxies, in order to obtain a three-dimensional picture of the growth of cosmic structure. Small ($\gtrsim 0.2\%$) redshift biases can easily dominate the overall error budget in the cosmological parameters measured by these surveys (e.g., Huterer et al. 2006). Spectroscopic calibration efforts for these missions must therefore measure the color–redshift relation of galaxies with sufficient fidelity to reconstruct the redshift distributions of shear samples with negligible systematic bias. While photometric redshift estimation techniques have grown in sophistication and precision over the past few decades (e.g., Benítez 2000; Brammer et al. 2008; Ilbert et al. 2009; Carrasco Kind & Brunner 2013; Speagle et al. 2016), existing methods have not met the photo- z accuracy requirements set by weak lensing cosmology.

The relation of seven or eight galaxy broadband colors (referred to henceforth by the vector C) to redshift is ultimately an empirical question. In Masters et al. (2015; hereafter M15) we demonstrated a method, based on the *self-organizing map* (SOM; Kohonen 1990) algorithm, to constrain the empirical multidimensional color distribution of galaxies present in a survey. This technique allowed us to project the multicolor

distribution of galaxies in a topologically ordered way onto a two-dimensional grid. By applying this technique to a well-studied deep field with uniform *ugrizYJH* photometry, we were able to demonstrate that spectroscopic surveys to date do not sample the full color space of galaxies in a *Euclid*-like survey, and thus the color–redshift relation is not fully constrained with existing spectroscopy. This issue is of particular concern for machine-learning-based photo- z estimation, which requires color-complete training samples, but also affects the calibration of template-based techniques.

The analysis in M15 motivated a survey designed to systematically map the color–redshift relation over the currently undersampled regions of galaxy color space relevant to *Euclid*. M15 estimated that ~ 5000 new redshifts, carefully distributed in color space, would be sufficient to meet the stringent requirements for weak lensing cosmology. This “direct” approach to photo- z calibration is complementary to approaches based on spatial cross-correlation of photometric samples with spectroscopic samples (e.g., Newman 2008; Rahman et al. 2015). At least two independent methods to measure $N(z)$ for the tomographic shear samples will be required to ensure no systematic photo- z biases exist; these methods can therefore serve as useful checks on each other.

Here we describe the initial stage of what we are calling the Complete Calibration of the Color–Redshift Relation (C3R2) survey, designed to fill out the color space of galaxies with

secure redshifts to the *Euclid* weak lensing depth. By doing so, the empirical $P(z|C)$ relation obeyed by galaxies can be constrained with sufficient accuracy to meet the cosmological requirements of *Euclid*. The spectra will also be of significant value for the LSST and *WFIRST* calibrations, which will be more difficult than the calibration for *Euclid* due to the greater photometric depth of those surveys (Hemmati et al. 2017, in preparation). We estimate that ~ 40 Keck nights in total (or their equivalent) could achieve the fidelity required to meet the cosmological requirements for *Euclid*, when combined with extensive existing spectroscopy.

This paper gives an overview of the C3R2 survey and presents results from the 2016A semester, which constituted the first five nights of observing. All 2016A observations were done with Keck. The structure of this paper is as follows. In Section 2 we give an overview of the C3R2 survey strategy. In Section 3 we discuss the observations and data reduction for the first five nights of observing. In Section 4 we describe redshift determination and the identification of serendipitous sources. In Section 5 we present initial results from the survey. In Section 6 we conclude with a discussion. High-confidence redshifts from DR1 are provided in a machine readable table.

2. C3R2 Survey Overview

The Keck portion of the C3R2 survey is a joint effort between Caltech (PI: J. Cohen), NASA (PI: D. Stern), the University of California (PI: B. Mobasher), the University of Hawaii (PI: D. Sanders). European participation in C3R2 with the GTC (PI: C. Gutierrez) and VLT (PI: F. Castander), as well as Harvard participation with MMT (PI: D. Eisenstein), will commence in 2017. The first five nights of observing with Keck were allocated by Caltech in the 2016A semester. Here, we provide a brief overview of the C3R2 strategy for these observations.

2.1. The Self-organized Map of Galaxy Colors

In M15 we used COSMOS (Capak et al. 2007; Scoville et al. 2007; Laigle et al. 2016) *ugrizYJH* photometry of $\sim 130k$ galaxies, closely resembling what will be obtained by the *Euclid* survey, to map the color distribution of galaxies to the *Euclid* depth ($i \sim 24.5$ AB). We used the SOM algorithm (a manifold learning technique for nonlinear dimensionality reduction) to generate a topologically ordered 2D representation of the high-dimensional color distribution.⁸ Galaxies from COSMOS were then matched back to the self-organized map according to their best-matching color cell in the SOM. This sorting of galaxies enables a variety of analyses, including the density of galaxies in different parts of color space, the median 30-band photometric redshifts from COSMOS as a function of position in color space, and the distribution of spectroscopic redshifts on the map (Figure 1). Importantly, by placing all existing spectroscopy from the COSMOS field on the map, we reveal regions of color space for which no galaxies have existing high-confidence redshifts. Of greatest importance for the C3R2 survey are: (1) the current spectroscopic sampling across color space, and (2) the source density as a function of

⁸ The SOM algorithm was used mainly for its relative simplicity and visualization power; however, any technique that manages to quantify the density of galaxies in multicolor space would be equally appropriate. A number of other techniques for nonlinear dimensionality reduction (e.g., the generative topographic map, growing neural gas, and local linear embedding) may, in principle, offer some advantages over the SOM.

position in color space, as more common galaxies will contribute more to the cosmic shear signal.

2.2. Existing Spectroscopy across Galaxy Color Space

For C3R2 we need to identify the regions of galaxy color space for which spectroscopic redshifts already exist and where they are systematically missing. We collected existing spectroscopy in COSMOS to do this, as described in M15. These redshifts include (but are not limited to) those from VLT-VIMOS (Lilly et al. 2007; Le Fèvre et al. 2015), Keck-MOSFIRE (Kriek et al. 2015), Keck-DEIMOS (Kartaltepe et al. 2010), and *Magellan*-IMACS (Trump et al. 2007). For the 2016A run we used only the spectroscopy taken in the COSMOS survey to identify undersampled regions of color space. The reason we could not incorporate spectroscopy from other fields for these observations is that the photometry between fields has to be highly consistent in multiple bands to reliably place galaxies on the same color map; at the time this problem had not been solved. Significant subsequent work has been done to solve this problem for upcoming runs, to be described in a forthcoming paper. The fields that have subsequently been put on a highly consistent color frame in *ugrizYJH* to the *Euclid* depth are VVDS, SXDS, and EGS (in addition to COSMOS).

2.3. Target Prioritization

For the 2016A observations we used the SOM derived in M15 to prioritize regions of galaxy multicolor space that are currently undersampled by existing spectroscopic surveys. For observed fields in 2016A other than COSMOS (SXDS and EGS), we attempted to bring the photometry onto the COSMOS color system in order to select the targets in a consistent way. We used the CANDELS (Grogin et al. 2011) photometry in a subset of the COSMOS field, together with the CANDELS photometry in SXDS and EGS, to derive a rough color conversion between the fields.

Target prioritization for C3R2 is based on two main factors: (1) the usefulness of a galaxy for calibrating the $P(z|C)$ relation, and (2) the likelihood of obtaining a secure redshift given the instrument, exposure time, and expected galaxy properties. The usefulness of a particular galaxy to the redshift calibration effort depends both on how common its colors are in the data and whether high-confidence redshifts for galaxies with similar colors already exist.

Based on these considerations, we developed a prioritization scheme for galaxies that weights sources in unsampled cells of the SOM more heavily, and also gives preference to more common galaxy colors. The priorities for C3R2 are adaptive as new data is obtained and more of the color space is filled in. For the 2016A run our priority scheme was as follows:

1. We assign a initial priority value of 10 to objects occupying cells with no spectroscopic redshifts of even moderate quality (the gray regions of the SOM in the middle panel of Figure 1), a starting priority of 3 to objects in cells with a spectroscopic redshift(s) of only moderate confidence, and a starting priority of 1 to galaxies in color cells that already have one or more high-confidence spectroscopic redshifts. Galaxies with existing redshifts of at least moderate confidence were not targeted.

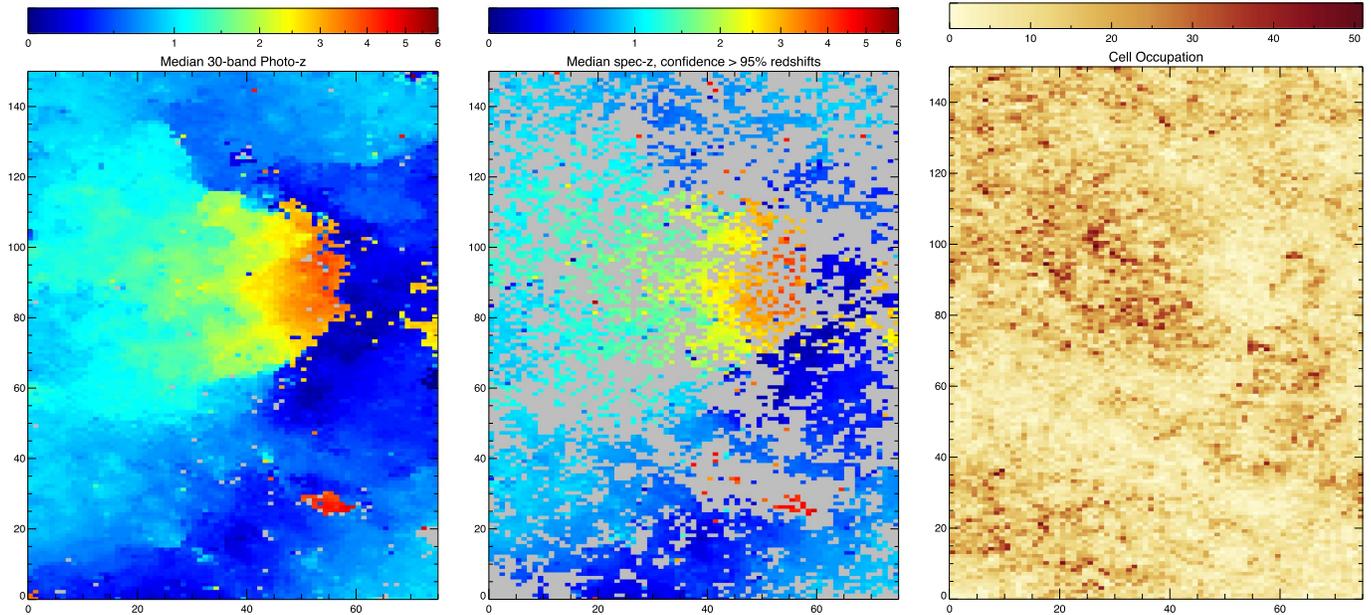


Figure 1. Self-organizing map (SOM) forming the basis for the C3R2 targeting strategy in the 2016A semester. Each one of the 11,250 cells represents a particular galaxy SED that shows up in the data with some degree of regularity; see Masters et al. (2015) for details. Note that the axes are not physical, but are merely indices into the map. Left: the map colored by the median 30-band photometric redshifts of galaxies associating with each cell, using the full input sample of sources. Center: our imperfect knowledge of the *empirical* $P(z|C)$ mapping from existing spectroscopy in the field. The cells are colored according to the median redshift of spec- z objects in the color cell with high-confidence redshifts. Gray cells have no galaxies with existing high-confidence redshifts. Right: the density of galaxies across the map. These are the key ingredients of the C3R2 survey strategy.

2. We multiply each galaxy’s priority by the number of objects in its color cell, effectively upweighting sources with common SEDs.⁹
3. We penalize objects that are color outliers within their color cell in order to avoid using them for calibration. A small fraction of objects in the sample are not represented well in the SOM, either because they have abnormal colors from photometric errors or the superposition of two or more sources, or are truly rare objects (e.g., X-ray sources). We want to avoid calibrating with these.

As will be described in future data releases, this prioritization scheme has been refined for the 2016B and later observations to more efficiently map the color–redshift relation. In M15 we pointed out that spectroscopic effort could also be intentionally directed at regions of color space with intrinsically higher redshift uncertainty (e.g., with double-peaked redshift PDFs). For now we have not prioritized based on redshift uncertainty; however, as the survey progresses and the color map is filled in we may incorporate this quantity.

2.4. Estimating Required Instruments and Exposure Times

A crucial element of the C3R2 survey is the use of best-fit spectral templates to the galaxies to predict the exposure times with different instruments needed to obtain a secure redshift. If we then fail to obtain a redshift under nominal observing conditions we can prioritize the target further for follow-up. This potential re-targeting is important to avoid systematic biases in the redshifts obtained in different parts of color space.

⁹ We have since substantially lessened the extent to which we weight by cell occupation, because it is effectively accomplished already by the source density on the sky—i.e., more common sources will find their way onto masks more frequently.

We use a technique developed for the proposed SPHEREx mission (Doré et al. 2014; Stickley et al. 2016) to predict the spectrum of galaxies based on their broadband photometry. In brief, this method fits a set of templates based on the libraries of Brown et al. (2014; for galaxies) and Salvato et al. (2009; for AGNs) to deep multiband photometry. Based on the analysis in Stickley et al. (2016), we can estimate the continuum to within 20% and the emission line strengths to within a factor of two. We then use the instrumental response curves for each telescope and instrument to estimate the required integration time to obtain a redshift to that galaxy, given its estimated photometric redshift. Primary objects for a mask are those expected to yield a redshift within a factor of two of the intended mask integration time. The time estimates were compared with previous observations to verify their accuracy. As described in Section 4.2, we use a flagging scheme to keep track of objects for which a redshift was expected but not obtained. These sources can then be prioritized for additional observations.

3. Observations and Data Reductions

Five nights were allocated by Caltech in the 2016A semester: three nights on DEIMOS (Faber et al. 2003), and one night each with LRIS (Oke et al. 1995) and MOSFIRE (McLean et al. 2012). Tables 1 and 2 summarize the nights and observed slitmasks. All five nights had excellent observing conditions. Here we describe the observations and data reduction.

3.1. DEIMOS

DEIMOS observations were conducted using the 600 groove mm^{-1} grating blazed at 7200 Å and the GG400 blocking filter, with dithering performed to improve sky subtraction. In the initial observing run, we experimented with minimum slit

Table 1
List of Observing Nights

UT Date	Code	Instrument	# Masks	Observing Conditions
2015 Dec 15	N01-D	DEIMOS	4	clear, 0 ^h 65 seeing
2016 Feb 28	N02-M	MOSFIRE	6	clear, 0 ^h 5–0 ^h 65 seeing
2016 Feb 29	N03-D	DEIMOS	4	clear, 0 ^h 65 seeing; moon
2016 Mar 01	N04-D	DEIMOS	7	clear, 1 ^h 0 seeing; moon
2016 Apr 09	N05-L	LRIS	4	thin cirrus, 0 ^h 97 seeing

Table 2
List of Observed Slitmasks

Mask ID/Name	Night	R.A. (J2000)	Decl. (J2000)	PA (°)	Exposure (s)	# Targets (total/ $Q = 4$ /ser)
16A-D01/UDS-m1n1	N01-D	2:17:27.0	−5:15:07	90.0	2 × 1800	86/51/5
16A-D02/UDS-m3n1	N01-D	2:17:27.0	−5:14:07	90.0	2 × 1800	100/66/5
16A-D03/COSMOS-m3n1	N01-D	10:00:22.0	+2:20:00	90.0	4 × 1800	104/72/7
16A-D04/COSMOS-m4n1	N01-D	10:00:22.0	+2:35:00	90.0	4 × 1800	70/60/19
16A-M05/COSMOS-m1-Y	N02-M	10:00:57.2	+1:48:40	85.0	20 × 180	24/8/1
16A-M06/COSMOS-m2-Y	N02-M	10:00:54.4	+2:01:47	55.0	20 × 180	29/3/0
16A-M07/COSMOS-m3-Y	N02-M	10:00:57.7	+2:14:38	40.0	20 × 180	24/5/1
16A-M08/COSMOS-m1-K	N02-M	10:00:10.5	+2:14:20	30.0	20 × 180	12/5/0
16A-M09/COSMOS-m4-Y	N02-M	10:00:14.2	+2:03:34	30.0	16 × 180	25/6/1
16A-M10/EGS-m1-K	N02-M	14:17:57.4	+52:35:51	25.0	22 × 180	23/3/0
16A-D11/COSMOS-m1n2	N03-D	9:58:43.2	+1:42:00	90.0	3 × 1200	93/62/2
16A-D12/COSMOS-m8n2	N03-D	9:58:43.3	+2:12:47	90.0	6 × 1200	92/56/17
16A-D13/COSMOS-m2n2	N03-D	9:58:43.2	+1:46:15	90.0	3 × 1200	91/77/7
16A-D14/COSMOS-m9n2	N03-D	9:58:43.2	+2:17:00	90.0	6 × 1200	99/64/3
16A-D16/COSMOS-m3n2	N04-D	9:58:43.2	+1:50:24	90.0	3 × 1200	89/43/4
16A-D17/COSMOS-m4n2	N04-D	9:58:43.2	+1:54:36	90.0	3 × 1200	95/70/4
16A-D18/COSMOS-m7n2	N04-D	9:58:43.2	+2:08:15	90.0	6 × 1200	91/52/13
16A-D19/COSMOS-m6n2	N04-D	9:58:43.2	+2:04:16	90.0	4 × 1200	94/72/9
16A-D20/COSMOS-m5n2	N04-D	9:58:43.2	+1:58:48	90.0	3 × 1200	98/80/3
16A-D21/EGS-m1n2	N04-D	14:18:00.0	+52:33:00	90.0	3 × 1200	100/62/10
16A-D22/EGS-m2n2	N04-D	14:18:00.0	+52:41:24	90.0	3 × 1200	104/72/3
16A-L23/COSMOS-m1n5	N05-L	9:59:44.1	+2:36:12	−60.0	4 × 1200	25/3/1
16A-L24/COSMOS-m3n5	N05-L	9:58:58.7	+2:45:56	−110.0	2 × 1200	18/8/0
16A-L25/EGS-m1n5	N05-L	14:19:08.6	+52:28:48	0.0	5 × 1200	28/11/0
16A-L26/EGS-m2n5	N05-L	14:18:04.8	+52:42:01	0.0	5 × 1200	26/4/1

Note. The “Night” column refers to the observing code in the second column of Table 1: night number, followed by a letter indicating the instrument used (D—DEIMOS, L—LRIS, M—MOSFIRE). R.A. and decl. refer to the mask center. The final column gives the total number of slitlets in the mask (ignoring those with failure codes −93/−94 as described in Section 4.2), the total number of high-quality ($Q = 4$) redshifts measured, and the number of serendipitous sources with high-quality redshifts (quality flag $Q = 4$).

Table 3
Spectroscopic Results

ID	R.A.	Decl.	Mask	Slit #	I (AB)	z	Qual.
UDS-3583	02:17:30.65	−05:15:24.4	UDS-m1n1	001	23.4	0.7877	4
UDS-10246	02:17:17.55	−05:13:06.9	UDS-m1n1	002	23.9	0.8028	4
UDS-767	02:17:59.05	−05:16:21.2	UDS-m1n1	003	25.0	0.5558	4
UDS-7109	02:17:00.35	−05:14:15.4	UDS-m1n1	004	23.6	1.0314	3
UDS-2276	02:17:52.83	−05:15:55.2	UDS-m1n1	005	22.9	0.9388	4
UDS-8536	02:17:53.37	−05:13:40.3	UDS-m1n1	006	24.7	0.8619	4
UDS-9784	02:17:56.80	−05:13:15.9	UDS-m1n1	009	23.6	0.8533	4
UDS-10739	02:17:44.88	−05:12:58.7	UDS-m1n1	010	23.2	1.0594	4
UDS-9730	02:17:32.64	−05:13:17.4	UDS-m1n1	012	23.8	1.0949	4
UDS-12725	02:17:14.84	−05:12:19.8	UDS-m1n1	013	23.7	1.0351	4
...							

Note. The full catalog of 1283 sources with quality flags ≥ 3 is given in machine readable format. The first 10 entries are shown here. Note that an “s” appended to the slit number indicates a serendipitous source.

(This table is available in its entirety in machine-readable form.)

lengths of both 6" and 10", with no significant difference in the redshift success rate. In the subsequent DEIMOS observations we settled on a minimum slit width of 8" as a balance between getting the most targets possible on the mask and getting good sky measurements. Data were reduced using a modified version of the DEEP2 pipeline designed to deal with dithered data.

3.2. LRIS

We used the 400 groove mm^{-1} blue grism blazed at 3400 Å and the 400 groove mm^{-1} red grating blazed at 8500 Å, with the D560 dichroic. Our choice of blue grism gives high sensitivity at bluer wavelengths where identifying features are likely to be found for objects with photometric redshifts of $z \sim 1.5\text{--}3$, while the red coverage allows for the detection of [O II] for some sources out to $z \sim 1.6$. The LRIS spectra were reduced using the IRAF-based BOGUS software developed by D. Stern, S. A. Stanford, and A. Bunker, and flux-calibrated using observations of standard stars from Massey & Gronwall (1990) observed on the same night using the same instrument configuration.

3.3. MOSFIRE

MOSFIRE was used in its default configuration. For instrumental details we refer the reader to Steidel et al. (2014). We observed four masks in the Y-band and two in the K-band, using integration times of 180 s with ABAB dithering to improve sky subtraction. Reductions were performed with the MOSFIRE Data Reduction Pipeline (DRP) made available by the instrument team.¹⁰

4. Redshift Determination

Each observed source was assessed independently by two co-authors to determine the redshift and associated quality flag. These results were then compared for conflicts in either redshift or quality flag. Conflicts were reconciled through a joint review of the spectra, usually with the help of a third, independent reviewer. As a final step in the process, we investigated all $Q = 4$ (highest-quality, see Section 4.1) sources for which the spectroscopic redshift (z_s) was highly discrepant from the expected photometric redshift (z_p , defined as the median photometric redshift of sources in the relevant SOM cell). Specifically, we investigated all sources with $|z_p - z_s|/(1 + z_s) \geq 0.15$. For most of these outliers, the spectroscopic redshift was deemed solid and we discuss the nature of the discrepancy in more detail in Section 5.2. However, for two cases this step caused us to modify the final redshift assessment. One of these final modifications was due to confusing a target and a very close ($\sim 1''$ separation) serendipitous source, while the other modification was due to a genuine error in line identification aggravated by incomplete sky-line subtraction mimicking a corroborating emission line.

4.1. Quality Flags

The redshift flagging scheme we use is similar to that adopted by the zCOSMOS (Lilly et al. 2007), DEEP2 (Newman et al. 2013), and VUDS surveys (Le Fèvre et al. 2015). The quality flags range from 0 to 4, with 4 indicating the highest confidence redshift and 0 indicating that

no redshift could be found. The interpretation of the flags is roughly as follows.

1. $Q = 4$: a quality flag of 4 indicates an unambiguous redshift identified with multiple features or the presence of the split [O II] $\lambda\lambda 3726, 3729$ doublet.
2. $Q = 3.5$: a quality flag of 3.5 indicates a high-confidence redshift based on a single line, with a very remote possibility of an incorrect identification. An example might be a strong, isolated emission line identified as $H\alpha$, where other identifications of the line are highly improbable due to the lack of associated lines or continuum breaks. This flag is typically only adopted for LRIS and MOSFIRE spectra.
3. $Q = 3$: a quality flag of 3 indicates a high-confidence redshift with a low probability of an incorrect identification. An example might be the low signal-to-noise ratio detection of an emission line, possibly corrupted by telluric emission or absorption, identified as [O II] $\lambda\lambda 3726, 3729$, but where the data quality is insufficient to clearly resolve the doublet.
4. $Q = 2/1$: a quality flag of 2 indicates a reasonable guess, while a quality flag of 1 indicates a highly uncertain guess. Sources with these low-confidence redshifts are not included in the data release.
5. $Q = 0$: a quality flag of 0 indicates that no redshift could be identified. As described next, a code indicating the cause of the redshift failure is assigned in place of the redshift.

Figure 2 shows six C3R2 spectra from 2016A as examples of $Q = 4$, $Q = 3.5$, and $Q = 3$ redshift assignments.

4.2. Failure Codes

It is important for C3R2 to track redshift failures, as well as the reasons for the failures, in order to avoid systematic biases in the sources selected for calibration. Failed targets that were expected to yield a redshift given the instrument and exposure time can be prioritized for additional follow-up. On the other hand, if no spectroscopic redshift was obtained because of a problem with the observing conditions or data (i.e., bad rows, or the target ended up in a region between two detector arrays), no additional prioritization of that source may be needed.

With these considerations in mind, we developed a system to flag different "failure modes" for objects not yielding a redshift. Four categories of failures are used, with the corresponding code assigned in place of a redshift in our catalog. The failure modes we identify are:

1. Code = -91: the object is too faint to identify the redshift. This indicates that a deeper exposure and/or different instrument and/or different wavelength coverage are required to obtain a secure redshift. An example of such a source might be a galaxy expected to have a strong [O II] emission line at 9800 Å, but where the slit placement caused the wavelength coverage to end at 9500 Å, yielding a continuum detection without any strong spectroscopic features. These sources will be further prioritized in future observations.
2. Code = -92: the object is well-detected, but no redshift could be determined. This may require a different instrument for secure redshift determination, due to an incorrect photometric redshift or the wavelength coverage

¹⁰ <https://keck-datareductionpipelines.github.io/MosfireDRP/>

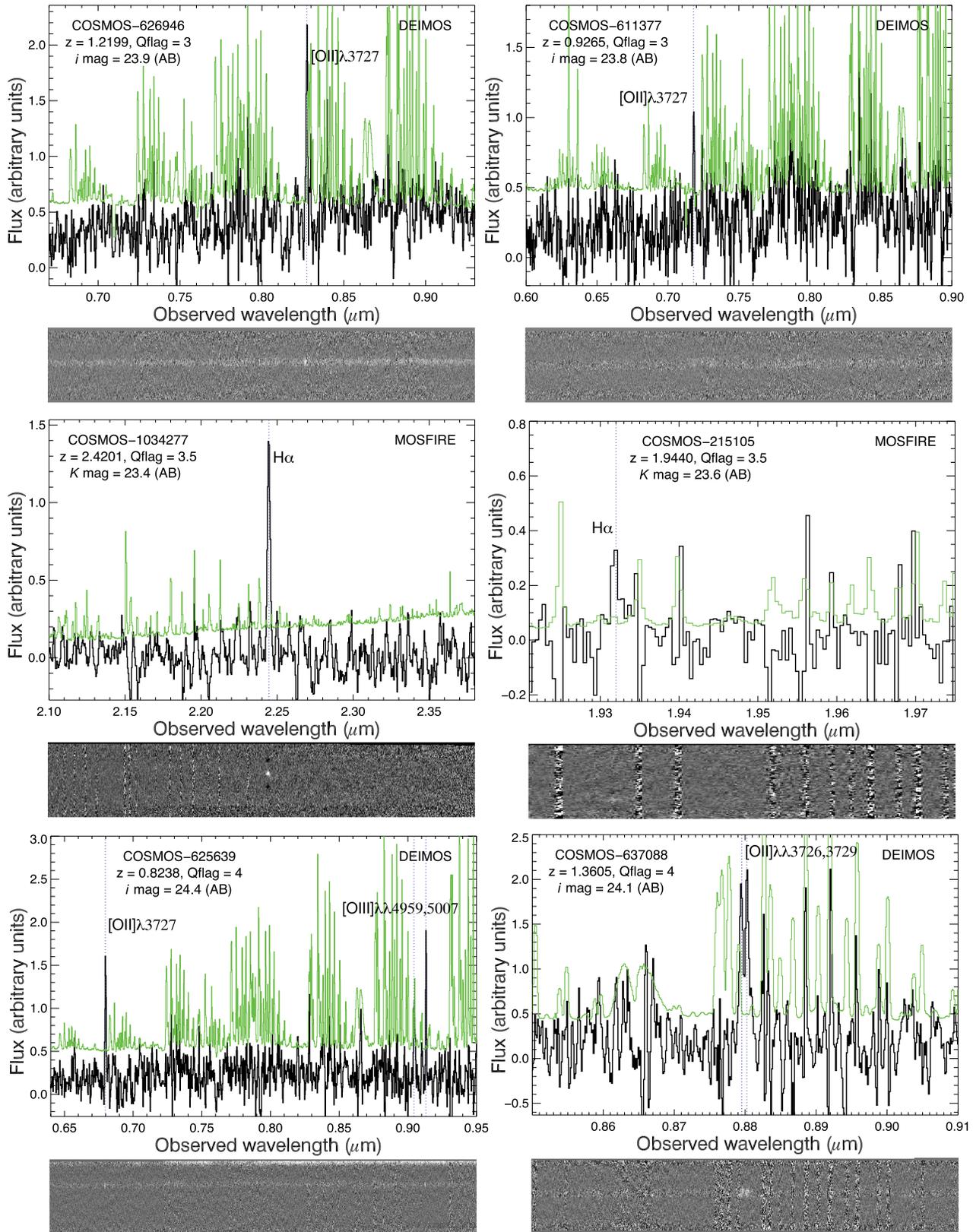


Figure 2. Example spectra from the 2016A semester. Error spectra are overlotted in green. Top left: DEIMOS spectrum with a confidence flag $Q = 3$ redshift based on a line identified as [O II] with an associated continuum break. The redshift is not class $Q = 4$, as the line is not well-resolved into the [O II] $\lambda\lambda$ 3726, 3729 doublet. Top right: Another similar example of a $Q = 3$ redshift based on a weak line identified as [O II], together with a continuum break. Middle left: a MOSFIRE K -band spectrum of a source at $z = 2.4201$. The single strong line is identified as $H\alpha$, with a confidence flag $Q = 3.5$ indicating high confidence (but not certainty, due to the lack of corroborating features). Middle right: a similar example of a $Q = 3.5$ MOSFIRE redshift based on a weak line identified as $H\alpha$. Bottom left: DEIMOS spectrum illustrating the assignment of a confidence flag $Q = 4$ redshift based on the detection of multiple features ([O II] λ 3727, [O III] $\lambda\lambda$ 4959, 5007). Bottom right: DEIMOS spectrum illustrating a $Q = 4$ redshift based on the clear, resolved presence of the [O II] $\lambda\lambda$ 3726, 3729 doublet.

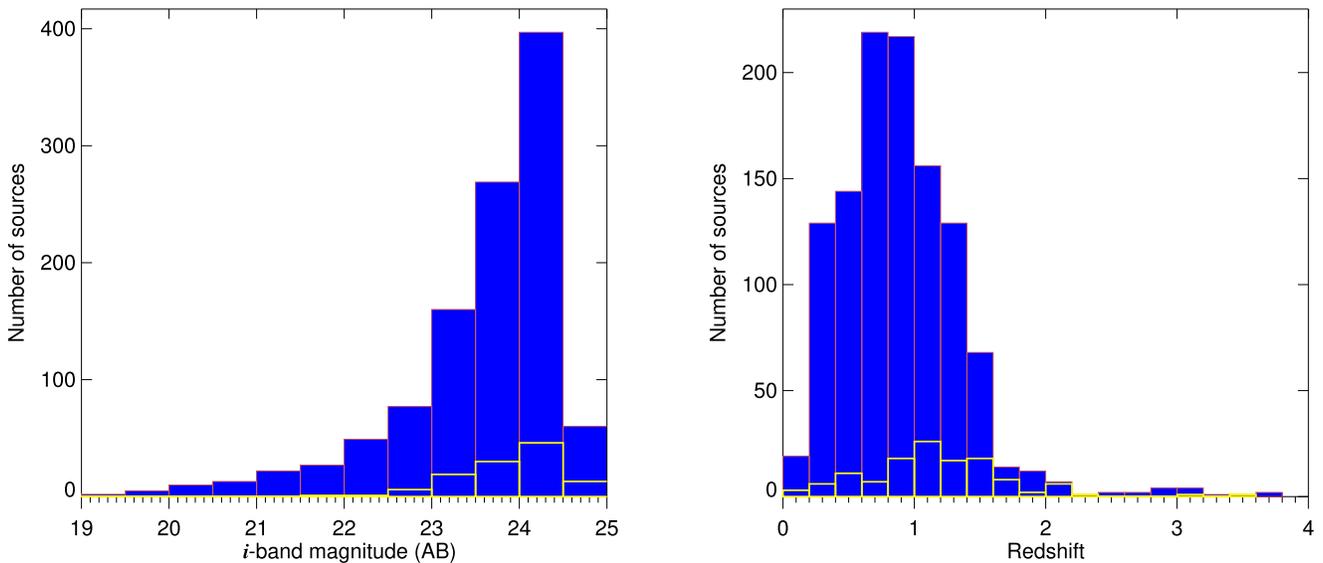


Figure 3. Left: the i -band magnitude distribution of C3R2-targeted sources yielding $Q = 4$ redshifts (blue) and $Q = 3/3.5$ redshifts (yellow) in the 2016A semester (1283 sources total). The majority are fainter than $i = 23$ (AB). Right: the redshift distribution of the same samples.

obtained for a given observation. We emphasize that the dividing line between -91 and -92 failure codes is imprecise, and no strong effort was made to homogenize the classification. Fundamentally, both codes can be considered two aspects of the same issue. Again, these sources will increase in priority going forward.

3. Code = -93 : this flag indicates a corrupted slit, typically due to bad rows/columns in the data or the source falling on or near detector chip gaps. This does not affect the object priority in future observations.
4. Code = -94 : this flag indicates a missing slit, as an extreme case of code -93 . This does not affect the object priority in future observations.

Failure codes -91 and -92 essentially correspond to spectral quality issues (i.e., signal-to-noise ratio, wavelength range), while codes -93 and -94 correspond to data quality issues (i.e., slitmask design issues, detector issues). While in DR1 we distinguished between these four failure modes, considering just the two general categories will be sufficient for the purposes of most analyses.

Failure code -91 , the most common failure code, generally indicates that the signal-to-noise ratio of the data was insufficient for redshift determination. Indeed, considering the 131 DEIMOS-observed sources in COSMOS with this failure code, 122 (93%) were *anticipated* to fail based on our estimated exposure time needed to get a redshift. As with low-confidence redshifts, sources for which we failed to find a redshift are not included in this data release.

4.3. Serendipitous Sources

We measured the redshifts of 134 serendipitously detected sources that happened to fall in slits with primary C3R2 targets. The coordinates of these sources were identified and they were matched back to the survey catalogs. The redshifts for these sources are included in our published catalog.

4.4. Literature Sources

Some (unintentional) overlap with literature redshifts allows a check on our results. In COSMOS and EGS we observed 38 sources that have previously existing high-quality redshifts. Most (24) were serendipitous detections. We find an rms discrepancy between our redshifts and the literature values of 4×10^{-4} . C3R2 redshifts are often higher precision than the literature values, which likely explains this small difference. There is no systematic difference between the C3R2 and literature redshifts.

5. Redshift Results and Calibration Progress

A total of 1825 sources were targeted in the 2016A observations. We identified 1131 $Q = 4$ redshifts, 27 $Q = 3.5$ redshifts, and 125 $Q = 3$ redshifts. In principle, only the highest confidence redshifts should be used for calibration for cosmology; whether this restricts usable sources to those with $Q = 4$ is worth investigating. Another 99 spectra yielded redshifts of low confidence ($Q = 1/2$), while there were 443 failures. Of these, 409 were failure codes -91 or -92 , indicating that the source was too faint or lacking in identifying features, while 34 were codes -93 or -94 , indicating a corrupted or missing slit.

In terms of the SOM presented in M15, and using only the C3R2 sources observed in COSMOS with $Q \geq 3$ for this analysis (911 redshifts), we have increased color space coverage by 5.4%. Figure 3 shows the i -band magnitude distribution and redshift distribution of the 2016A “gold” sample of $Q = 4$ sources, as well as the $Q = 3/3.5$ sources. The distributions are very similar to the overall distribution of the unsampled cells of galaxy color space identified in M15, indicating that we are targeting the correct sources.

5.1. SOM-based Photo- z Performance

The SOM colored by the median photo- z of sources per cell (the left panel of Figure 1) effectively defines a photometric redshift estimate for each galaxy based on its position in the $ugrizYJH$ color space of *Euclid*/*WFIRST*. Figure 4 compares

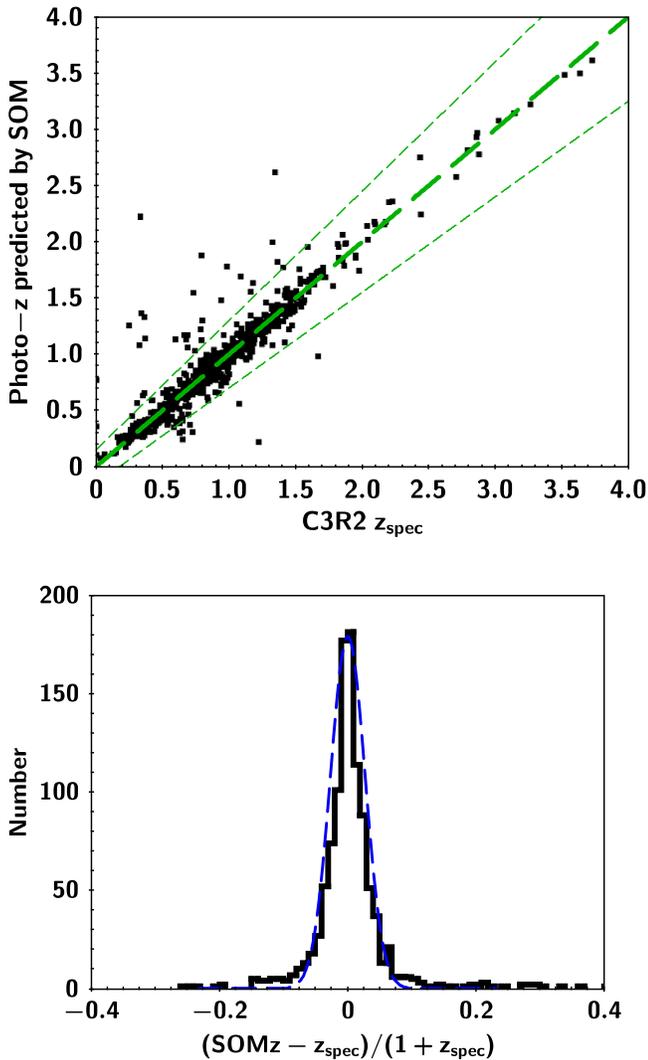


Figure 4. Top: comparison of the photometric redshift predicted by the SOM with the measured $Q = 4$ C3R2 redshifts. We define the SOM photo- z of an object to be the median of the photo- z 's for all objects that best associate with its particular SOM color cell. As can be seen, the SOM photo- z estimates are mostly unbiased, with a small outlier fraction ($\sim 3.8\%$, where outliers are defined as redshift errors $\geq 15\%$, indicated by the thin dashed lines). Bottom: the distribution of $(z_p - z_s)/(1 + z_s)$ for the $Q = 4$ sample. Overplotted is a Gaussian with $\sigma = 0.027$, equal to the measured σ_{NMAD} .

our $Q = 4$ spectroscopic redshifts with the redshift that would be inferred based on the SOM, with encouraging results. The normalized median absolute deviation (a dispersion measure that is not sensitive to catastrophic outliers Ilbert et al. 2009; Dahlen et al. 2013) defined as,

$$\sigma_{\text{NMAD}} = 1.48 \times \text{median} \left(\frac{|z_p - z_s|}{1 + z_s} \right), \quad (1)$$

is 0.027 (2.7%) for the sample, which is quite low.

Using the standard definition of catastrophic photo- z outliers as those with $|z_p - z_s|/(1 + z_s) \geq 0.15$, we measure a low outlier fraction of 3.8%. The measured bias, defined as

$$\text{mean} \left(\frac{z_p - z_s}{1 + z_s} \right), \quad (2)$$

is $\lesssim 0.1\%$ after removing the catastrophic outliers. Further improvements to these results will result from folding in all

spectroscopic information from C3R2 and other surveys to the $P(z|C)$ relation encoded by the SOM. Notably, these results are already competitive with or better than the photo- z results of codes tested in Dahlen et al. (2013), where the photometry used comprised 14 bands including full depth CANDELS and *Spitzer* data.

While the performance we find is quite good, and may be representative of what can be achieved with a survey such as *Euclid* or *WFIRST*, the results depend on the depth and stability of the photometry. The photometry used to place objects on the SOM in order to estimate a photo- z in the above analysis is quite deep (*i*-band depth ~ 25.4 AB). The results will degrade as the photometry gets shallower or bands are lost in a manner that can be directly characterized via the SOM. A detailed study of the expected performance from the SOM-based photo- z approach will be the subject of a future paper.

5.2. Outliers

Out of 1079 sources with $Q = 4$ redshifts and reliable SOM-based photo- z estimates, only 41 (3.8%) are outliers according to the standard definition, $|z_p - z_s|/(1 + z_s) \geq 0.15$. If, instead of the SOM-based photo- z , we use the photo- z for each object based on deep multiband data (e.g., the 30-band COSMOS data), we find an outlier fraction for the same sources of $\sim 3.1\%$. Thus the SOM photo- z (effectively based only on the seven color *Euclid*-like SEDs) performs nearly as well in terms of outlier fraction.

We have analyzed all of the outliers on a case-by-case basis. The majority (24/41; 59%) have *individual* (rather than SOM-based) photo- z estimates more in line with the measured redshift, indicating that the color cells they belong to have real redshift scatter. For nearly all of these sources, the measured dispersion in the 30-band photo- z 's within the relevant color cell is significantly larger than the median redshift dispersion per cell; in other words, these are sources that fall in more degenerate regions of the color space. The SOM can be used to identify these regions in a consistent way in order to either reject them in weak lensing analysis or direct extra spectroscopy at them to characterize the redshift distribution in those cells.

In addition, there are several other examples easily understood as Galactic stars (3) or obvious quasars/active galaxies (2), which are known not to have typical galaxy colors (total = 5/41; 12%). This process caught one mistaken line identification where our initial assessment of a MOSFIRE spectrum identified an isolated, narrow, strong line as [O III] $\lambda 5007$ with corroborating [O III] $\lambda 4959$ emission. Subsequent analysis reveals the latter emission line to be due to poorly subtracted telluric emission, and we now identify the strong emission line as $H\alpha$ ($Q = 3.5$). The remaining cases seem to be genuine mismatches between the spectroscopic redshift and the photometric redshifts, for both the individual photometric redshift of the galaxy and the SOM-based photometric redshift. Consideration of *Hubble* imaging reveals at least some of these as likely being due to two close-separation galaxies where the ground-based imaging used for the photometry was unable to separate the sources.

5.3. Increased Color Space Coverage

The five nights of observing in 2016A filled in $\sim 6\%$ of the map, in addition to existing spectroscopy that already filled

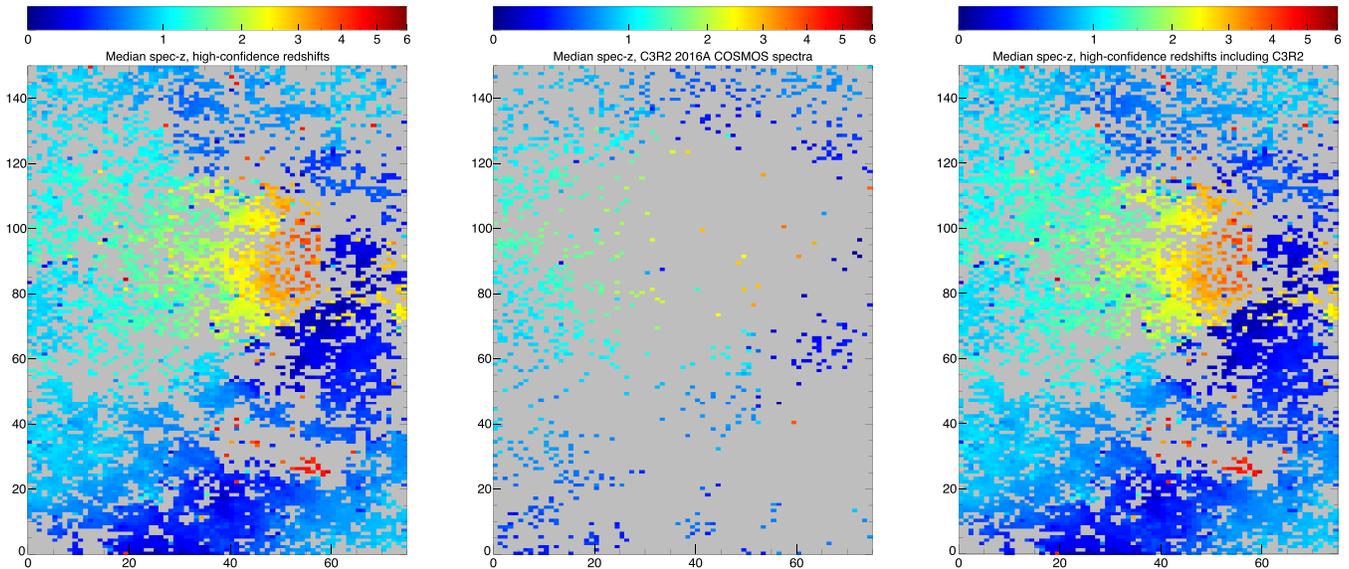


Figure 5. SOM used as the basis for the C3R2 targeting strategy in the 2016A semester, illustrating the inclusion of the new C3R2 redshifts. Left: the SOM colored by the median spec- z in each color cell from preexisting spectroscopy from COSMOS. Gray cells have no preexisting high-quality redshifts. Middle: distribution in color space of the 911 high-quality redshifts obtained by C3R2 in the 2016A semester in COSMOS. Right: the map including the high-quality redshifts added in the COSMOS field by the 2016A C3R2 nights. An additional 5.4% of the map is measured, as well as additional constraints placed on other regions of color space. The increased density of coverage (fewer gray cells) is apparent on the right compared with the left panel. Note that this map of the relation of color to redshift is not a model, but is entirely empirical.

$\sim 50\%$ (see Figure 5). Thus, we completed $\gtrsim 10\%$ of the required calibration. However, some of the remaining observations may prove more challenging. Given the recent progress toward bringing multiple deep fields (and their spectroscopy) onto a consistent color system, the requirements may also change to some extent, in the sense that somewhat fewer spectra are required due to the inclusion of other spectroscopic surveys.

It should also be noted that a certain fraction of the remaining cells represent faint, red sources for which spectroscopic redshifts are prohibitively difficult to obtain with current instruments. These constitute a small ($\sim 3\%$) fraction of the unsampled cells. If needed, the SOM provides a consistent method for identifying such objects and removing them from the weak lensing sample.

6. Conclusion

We have presented initial results of the C3R2 survey based on five nights of Keck spectroscopy in the 2016A semester. C3R2 is designed to supplement extensive existing spectroscopy in order to provide a spectroscopic sample spanning the observed colors of galaxies to the *Euclid* weak lensing photometric depth. The ultimate aim of the survey is to calibrate the color–redshift relation sufficiently to meet the requirements set by weak lensing cosmology. We estimate that the survey would require ~ 40 Keck nights (or their equivalent) in total to meet the requirements set by *Euclid*.

In future papers we will present the updated survey strategy based on bringing multiple *Euclid* calibration fields onto a consistent color system, as well as realistic tests of the performance of the method. Initial tests show that the empirical color mapping technique is quite good at successfully reproducing $N(z)$ distributions with low bias.

Additional data, including results from 16.5 nights allocated in 2016B, as well as time allocated in 2017A and 2017B, will be presented in follow-up papers. Combined with data from

VLT, GTC, and MMT, we expect the calibration samples will be sufficient to meet the needs of *Euclid*. Work is ongoing to understand the needs for *WFIRST* calibration, but these spectra will form part of the foundation of that survey as well. Further tests and refinements of the calibration method, as well as studies to determine the optimal way to incorporate all existing spectroscopic and photometric information from deep fields into photo- z estimation using a limited set of broadband observations, are avenues of continuing research.

The authors thank the referee, Marcin Sawicki, for a helpful report that improved this paper. D.M. would like to thank Giuseppe Longo, Audrey Galametz, Sotiria Fotopoulou, and George Helou for helpful conversations. D.M., P.C., D.S., and J.R. acknowledge support by NASA ROSES grant 12-EUCLID12-0004. J.R. and D.S. are supported by JPL, run by Caltech for NASA. F.C. acknowledges support by MINECO grant ESP2015-88861. The authors wish to recognize and acknowledge the very significant cultural role and reverence that the summit of Mauna Kea has always had within the indigenous Hawaiian community. We are most fortunate to have the opportunity to conduct observations from this mountain.

References

- Benítez, N. 2000, *ApJ*, 536, 571
 Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2008, *ApJ*, 686, 1503
 Brown, M. J. I., Moustakas, J., Smith, J.-D. T., et al. 2014, *ApJS*, 212, 18
 Capak, P., Aussel, H., Ajiki, M., et al. 2007, *ApJS*, 172, 99
 Carrasco Kind, M., & Brunner, R. J. 2013, *MNRAS*, 432, 1483
 Dahlen, T., Mobasher, B., Faber, S. M., et al. 2013, *ApJ*, 775, 93
 Doré, O., Bock, J., Ashby, M., et al. 2014, arXiv:1412.4872
 Faber, S. M., Phillips, A. C., Kibrick, R. I., et al. 2003, *Proc. SPIE*, 4841, 1657
 Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, *ApJS*, 197, 35
 Huterer, D., Takada, M., Bernstein, G., & Jain, B. 2006, *MNRAS*, 366, 101
 Ilbert, O., Capak, P., Salvato, M., et al. 2009, *ApJ*, 690, 1236
 Ivezić, Z., Tyson, J. A., Abel, B., et al. 2008, arXiv:0805.2366
 Kartaltepe, J. S., Sanders, D. B., Le Floch, E., et al. 2010, *ApJ*, 709, 572
 Kohonen, T. 1990, *IEEEEP*, 78, 1464

- Kriek, M., Shapley, A. E., Reddy, N. A., et al. 2015, *ApJS*, 218, 15
- Laigle, C., McCracken, H. J., Ilbert, O., et al. 2016, *ApJS*, 224, 24
- Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, arXiv:1110.3193
- Le Fèvre, O., Tasca, L. A. M., Cassata, P., et al. 2015, *A&A*, 576, A79
- Lilly, S. J., Le Fèvre, O., Renzini, A., et al. 2007, *ApJS*, 172, 70
- Massey, P., & Gronwall, C. 1990, *ApJ*, 358, 344
- Masters, D., Capak, P., Stern, D., et al. 2015, *ApJ*, 813, 53
- McLean, I. S., Steidel, C. C., Epps, H. W., et al. 2012, *Proc. SPIE*, 8446, 84460J
- Newman, J. A. 2008, *ApJ*, 684, 88
- Newman, J. A., Cooper, M. C., Davis, M., et al. 2013, *ApJS*, 208, 5
- Oke, J. B., Cohen, J. G., Carr, M., et al. 1995, *PASP*, 107, 375
- Rahman, M., Ménard, B., Scranton, R., Schmidt, S. J., & Morrison, C. B. 2015, *MNRAS*, 447, 3500
- Salvato, M., Hasinger, G., Ilbert, O., et al. 2009, *ApJ*, 690, 1250
- Scoville, N., Aussel, H., Brusa, M., et al. 2007, *ApJS*, 172, 1
- Speagle, J. S., Capak, P. L., Eisenstein, D. J., Masters, D. C., & Steinhardt, C. L. 2016, *MNRAS*, 461, 3432
- Spergel, D., Gehrels, N., Baltay, C., et al. 2015, arXiv:1503.03757
- Steidel, C. C., Rudie, G. C., Strom, A. L., et al. 2014, *ApJ*, 795, 165
- Stickley, N. R., Capak, P., Masters, D., et al. 2016, arXiv:1606.06374
- Trump, J. R., Impey, C. D., McCarthy, P. J., et al. 2007, *ApJS*, 172, 383