# THE IMPACT OF *JWST* BROADBAND FILTER CHOICE ON PHOTOMETRIC REDSHIFT ESTIMATION

L. Bisigello[1,2], K. I. Caputi[1], L. Colina[3], O. Le Fèvre[4], H. U. Nørgaard-Nielsen[5], P. G. Pérez-González[6,7], J. Pye[8], P. van der Werf[9], O. Ilbert[4], N. Grogin[10], and A. Koekemoer[10]

[1] Kapteyn Astronomical Institute, University of Groningen, P.O. Box 800, 9700 AV, Groningen, The Netherlands
[2] SRON Netherlands Institute for Space Research, 9747 AD, Groningen, The Netherlands; bisigello@astro.rug.nl
[3] Centro de Astrobiología, Departamento de Astrofísica, CSIC-INTA, Cra. de Ajalvir km.4, E-28850—Torrejón de Ardoz, Madrid, Spain
[4] Aix Marseille Université, CNRS, LAM (Laboratoire d'Astrophysique de Marseille), UMR 7326, F-13388, Marseille, France
[5] National Space Institute (DTU Space), Technical University of Denmark, Elektrovej, DK-2800 Kgs. Lyngby, Denmark
[6] Departamento de Astrofísica, Facultad de CC. Físicas, Universidad Complutense de Madrid, E-28040 Madrid, Spain
[7] Associate Astronomer at Steward Observatory, The University of Arizona, USA
[8] University of Leicester, Department of Physics & Astronomy, Leicester, LE1 7RH, UK
[9] Sterrewacht Leiden, Leiden University, P.O. Box 9513, 2300 RA, Leiden, The Netherlands
[10] Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA

## ABSTRACT

The determination of galaxy redshifts in the *James Webb Space Telescope*'s (*JWST*) blank-field surveys will mostly rely on photometric estimates, based on the data provided by *JWST's* Near-Infrared Camera (NIRCam) at 0.6–5.0 $\mu$m and Mid Infrared Instrument (MIRI) at $\lambda > 5.0$ $\mu$m. In this work we analyze the impact of choosing different combinations of NIRCam and MIRI broadband filters (F070W to F770W), as well as having ancillary data at $\lambda < 0.6$ $\mu$m, on the derived photometric redshifts ($z_{phot}$) of a total of 5921 real and simulated galaxies, with known input redshifts $z = 0$–10. We found that observations at $\lambda < 0.6$ $\mu$m are necessary to control the contamination of high-$z$ samples by low-$z$ interlopers. Adding MIRI (F560W and F770W) photometry to the NIRCam data mitigates the absence of ancillary observations at $\lambda < 0.6$ $\mu$m and improves the redshift estimation. At $z = 7$–10, accurate $z_{phot}$ can be obtained with the NIRCam broadbands alone when S/N $\geqslant 10$, but the $z_{phot}$ quality significantly degrades at S/N $\leqslant 5$. Adding MIRI photometry with 1 mag brighter depth than the NIRCam depth allows for a redshift recovery of 83%–99%, depending on spectral energy distribution type, and its effect is particularly noteworthy for galaxies with nebular emission. The vast majority of NIRCam galaxies with [F150W] = 29 AB mag at $z = 7$–10 will be detected with MIRI at [F560W, F770W] < 28 mag if these sources are at least mildly evolved or have spectra with emission lines boosting the mid-infrared fluxes.

*Key words:* galaxies: distances and redshifts – galaxies: high-redshift – galaxies: photometry

## 1. INTRODUCTION

The power of multiwavelength photometric observations to recover galaxy redshifts has been known since the late fifties (Baum 1957), and it has been confirmed in the past decade with deep blank-field imaging surveys (e.g., Ilbert et al. 2009; Cardamone et al. 2010; Dahlen et al. 2013; Rafelski et al. 2015). Although spectroscopic redshifts are much better in precision than photometric estimates, photometry has two important advantages with respect to spectra: the number of sources that can be observed at the same time, and sensitivity. Indeed, with a single photometric pointing it is possible to observe a large number of sources, while in general only a limited number of objects can simultaneously be targeted with spectroscopy. In addition, the spectroscopic observation of faint targets usually requires very long exposure times (e.g., Le Fèvre et al. 2015), and the faintest objects detected in photometric maps are beyond the technical possibilities of contemporary spectrographs (Caputi et al. 2012, 2015).

Photometric redshift determinations usually rely on the identification of strong features, such as the Lyman or 4000 Å break, in a galaxy spectral energy distribution (SED), after they are convolved with the transmission functions of the filters utilized in the observations. This is the reason why it is necessary to do a careful filter selection when planning observations, balancing depth and wavelength coverage, in order to minimize degeneracies and misidentifications when obtaining photometric redshifts.

The *James Webb Space Telescope*[11] (*JWST*; Gardner et al. 2009) is a foremost space mission for the coming years and the awaited successor of the *Hubble Space Telescope* (*HST*) and *Spitzer Space Telescope* at infrared wavelengths. It will have four instruments on board for imaging, spectroscopy, and coronography, covering a wide range of wavelengths from 0.6 through 28.3 $\mu$m with subarcsecond angular resolution.

One of *JWST*'s main scientific aims is to study the formation and evolution of galaxies at early cosmic times. For this purpose, deep blank-field imaging surveys will be carried out with two imaging cameras, namely, the Near Infrared Camera (NIRCam; Rieke et al. 2005) and the Mid Infrared Instrument (MIRI; Rieke et al. 2015; Wright et al. 2015). In the vast majority of cases, galaxy redshift determinations will be done through SED-fitting analysis; therefore, it is crucial to understand the impact of choosing different filter combinations on the ability to recover the right redshifts for all the observed sources.

NIRCam's filter wheels include eight broadband filters, namely, F070W, F090W, F115W, F150W, F200W, F277W, F356W, and F444W, as well as a number of medium- and narrowband filters, covering the wavelength range 0.6–5.0 $\mu$m.[12] The MIRI imager[13] contains nine different broadband filters between 5 and 28 $\mu$m. The latter wavelength range

---

[11] http://www.jwst.nasa.gov
[12] http://www.stsci.edu/jwst/instruments/nircam
[13] http://www.stsci.edu/jwst/instruments/miri

encompasses the redshifted stellar light of high-$z$ galaxies, as well as the hot dust radiation emitted by low-$z$ sources. The two shortest-wavelength bands that we consider here (i.e., F560W and F770W) are the most sensitive in MIRI, so they will be the ones primarily used for high-$z$ galaxy studies.

The complementarity of NIRCam's and MIRI's wavelength coverages confirms that both cameras should ideally be considered for deep galaxy surveys. While surveys with NIRCam alone will be sufficient to encompass the 4000 Å break of galaxies at $0.6 < z < 10$, a secure measurement of the evolved stellar light (around rest 1 $\mu$m) in $z > 4$ galaxies will also require observations with the MIRI filters. However, due to different detector technology in the near- and mid-IR range, MIRI's sensitivity is significantly lower than that of NIRCam, making observers doubt whether it is convenient to invest time on long MIRI observations when planning deep galaxy surveys. At the same time, the lack of *JWST* coverage below 0.6 $\mu$m requires considering whether observing fields with no ancillary shorter-wavelength imaging can be effective to study galaxy evolution. Here we address these issues through the analysis of galaxy redshift recovery in different data availability situations, all treated on an equal basis, in order to allow the reader to decide which camera and filter configuration would be optimal for their scientific interests.

The aim of this work is to assess the impact of different *JWST* broadband filter combinations on recovering reliable photometric redshifts for galaxies at $z = 0$–10. In addition to considering the NIRCam broadbands and MIRI F560W and F770W filters, we tested the need of ancillary data at $\lambda < 0.6$ $\mu$m, such as those provided by *HST* observations. We applied our tests to three different galaxy samples with known redshifts: (1) a spectroscopic galaxy sample at $z = 0$–6, (2) a galaxy sample with consensus photometric redshifts at $z = 4$–7 from the CANDELS survey, and (3) a mock galaxy sample generated with different spectral templates at $z = 7$–10.

The structure of the paper is as follows. In Section 2 we describe our three-sample selection, the obtention of photometry in the relevant NIRCam and MIRI bands, and the photometric redshift test methodology. In Section 3 we analyze our results at different redshifts, and in Section 4 we summarize our main findings and conclusions. Throughout this paper, we adopt a cosmology with $H_0 = 70$ km s$^{-1}$ Mpc$^{-1}$, $\Omega_M = 0.27$, and $\Omega_\Lambda = 0.73$. All magnitudes refer to the AB system (Oke & Gunn 1983).

## 2. SAMPLE SELECTION AND TEST METHODOLOGY

### 2.1. Galaxy Sample Selection

We selected three different galaxy samples spanning rather complementary redshift ranges:

1. a sample of 2422 galaxies with secure spectroscopic redshifts at $z = 0$–6;
2. a sample of 1375 galaxies with consensus photometric redshifts at $z = 4$–7;
3. a sample of 2124 simulated galaxies at $z = 7$–10.

These three samples allowed us to investigate the different problems arising when trying to recover redshifts photometrically using different combinations of *JWST* filters. By no means are these samples trying to emulate a complete, flux-limited galaxy population at $z = 0$–10 as it will be seen in a blank *JWST* field, and this is not necessary given the scope of this paper.

#### 2.1.1. Sample 1

Our first galaxy sample contains 2422 galaxies with secure spectroscopic redshifts ($z_{\rm spec}$) from the ESO public compilation in the Chandra Deep Field South (CDFS; Cristiani et al. 2000; Croom et al. 2001; Bunker et al. 2003; Dickinson et al. 2004; Stanway et al. 2004a, 2004b; Strolger et al. 2004; Szokoly et al. 2004; van der Wel et al. 2004; Doherty et al. 2005; Le Fèvre et al. 2005; Mignoli et al. 2005; Ravikumar et al. 2007; Popesso et al. 2009; Balestra et al. 2010; Silverman et al. 2010; Kurk et al. 2013; Vanzella et al. 2014) and from Morris et al. (2015). In order to obtain multiwavelength photometry for these galaxies, we cross-matched this sample with the public Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS; Grogin et al. 2011; Koekemoer et al. 2011) catalog for the Great Observatory Origins Deep Survey South field (GOODS-S) obtained by Guo et al. (2013). This catalog contains photometry in 17 broadbands, from 0.37 through 8.0 $\mu$m, obtained with ground- and space-based telescopes, including *HST* and *Spitzer* (Ashby et al. 2015). This wavelength sampling allows for a good-quality SED fitting of the galaxy stellar emission and a proper redshift recovery in the vast majority of cases, using standard galaxy spectral templates and photometric redshift codes.

Our aim here is to assess the impact of choosing different *JWST* filter sets on the derived photometric redshifts, rather than the performance of different photometric redshift codes and/or SED template libraries. Therefore, we explicitly excluded from our analysis any galaxy from the parent spectroscopic sample for which we could not correctly recover the redshift photometrically using all the CANDELS photometry up to 4.5 $\mu$m. We did not use the two IRAC band data at longer wavelengths because their signal-to-noise ratio (S/N) is considerably lower than for the other bands and they could be contaminated by polycyclic aromatic hydrocarbon emission at low $z$. The excluded galaxies are only ~8% of all the galaxies with publicly available $z_{\rm spec}$ in the CDFS.

To obtain photometric redshift estimates for all the galaxies in the parent spectroscopic sample, we ran the public code LePhare (Arnouts et al. 1999; Ilbert et al. 2006) using templates from Bruzual & Charlot (2003, hereafter BC03) with solar metallicity, a range of exponentially declining star formation histories with different characteristic times $\tau$ from 0.01 to 10 Gyr, and ages from 0.01 to 10 Gyr. We allowed for emission lines in LePhare and applied the Calzetti et al. reddening law (Calzetti et al. 2000) with $A_V$ extinction values from 0 to 4.

We considered the subsample of 2422 objects with resulting photometric redshifts within $2\sigma$ ($\sigma = 0.13$) of the spectroscopic redshifts (Figure 1), which constitute 92% of the parent spectroscopic sample, and adopted it as our sample 1 for all further analysis. The sample 1 redshift distribution is shown in Figure 2. This distribution peaks around $z \sim 1$ and has a high-$z$ tail up to $z = 6$, although the statistics are rather poor at $z > 4$. So, our main purpose with sample 1 is to study the percentage of low-redshift galaxies that can leak to high $z$ when different filter sets, and thus different wavelength coverages, are available.
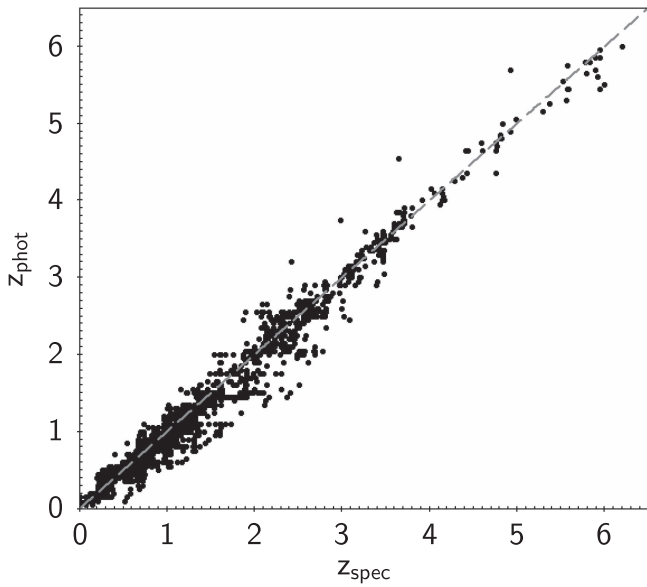
**Figure 1.** Photometric redshifts obtained with CANDELS 15-band photometry vs. spectroscopic redshifts for sample 1.
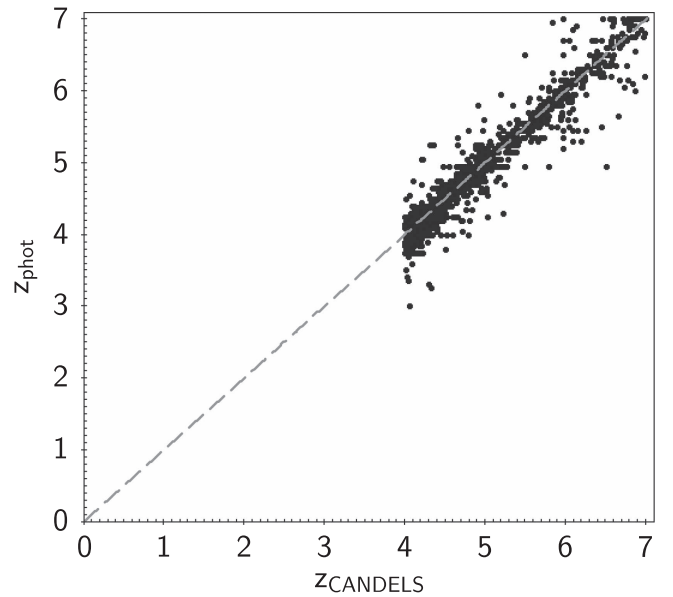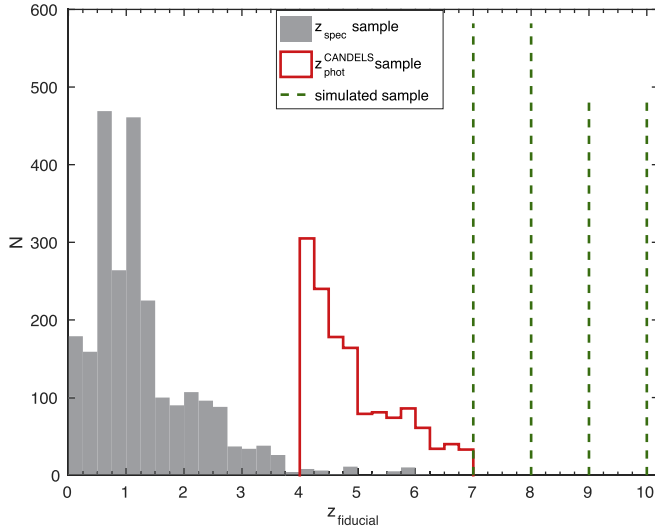


**Figure 2.** Redshift distribution of the three galaxy samples considered in this work: a galaxy sample with spectroscopic redshifts in the GOODS-S field (filled gray; sample 1), a galaxy sample with well-established CANDELS photometric redshifts, also in GOODS-S (red line; sample 2), and a sample of simulated galaxies with known input redshifts (dashed line; sample 3).

### 2.1.2. Sample 2

Our second sample consists of 1375 galaxies at $z = 4$–7 taken from the CANDELS GOODS-S public catalog. For this sample, we considered as fiducial redshifts the consensus photometric redshifts that have been compiled by Dahlen et al. (2013). The CANDELS consensus photometric redshifts result from the comparison of 11 independent SED-fitting runs. We considered only CANDELS galaxies with $z > 4$ here, as lower redshifts are well represented in our sample 1.

As we did for the sample with spectroscopic redshifts (sample 1), we obtained our own photometric redshifts by running LePhare on the CANDELS 15-band photometry and using the same SED templates and parameter values stated before. In this case, our own photometric redshifts agree within $2\sigma$ ($\sigma = 0.10$) with the CANDELS consensus redshifts for



**Figure 3.** Photometric redshifts obtained with 15-band CANDELS photometry in this work vs. the consensus CANDELS photometric redshifts for sample 2.

1375 (95%) of the GOODS-S galaxies with CANDELS redshifts $z_{CANDELS} = 4$–7 (Figure 3), consistently with the typical level of dispersion present in the individual SED-fitting CANDELS runs. We adopted these 1375 galaxies as our sample 2. The resulting redshift distribution is shown in Figure 2.

### 2.1.3. Sample 3

We built our third sample with mock galaxies at fixed redshifts $z = 7$–10 (for an analysis of more redshifts, see Appendix A), simulating their SEDs with two different methods. First, we considered the BC03 synthetic galaxy templates with four metallicities (1, 0.4, 0.2, and 0.02 $Z_{\odot}$), exponentially declining star formation histories, and five different ages, depending on redshift. We applied dust extinction by using the Calzetti et al. reddening law with $A_V = 0$–1 and the effect of absorption by the intergalactic medium (IGM; Madau 1995).

The BC03 models do not include emission lines, so we manually added the Hα, Hβ, [O III] λ5007 and [O II] λ3727 emission lines to those templates for which the age is equal to or lower than the characterisitic $\tau$ of the exponentially declining star formation history, as follows. We started from the observed mean value of the Hα+[N II]+[S II] equivalent width (EW), which is 422–423 Å at $z = 3.8$–5 (Smit et al. 2015). Then, we assumed line ratios from Anders & Fritze-v. Alvensleben (2003) in order to derive the EWs of Hα alone at different metallicities. Fumagalli et al. (2012) obtained a relation between Hα EW and redshift, namely, EW(Hα) $\propto (1 + z)^{1.8}$, which has been confirmed to be consistent with observations up to $z \sim 5$ (Stark et al. 2013).[14] Following this relation, we derived EW(Hα) for all considered redshifts. We also derived EWs for Hβ assuming case B recombination and the corresponding EWs for [O III] and [O II] by using the line ratios tabulated in Anders & Fritze-v. Alvensleben (2003). We

---

[14] This redshift dependence is slightly steeper than that obtained by, e.g., Faisst et al. (2016), but adopting this other prescription would only change our assumed Hα EW by a factor <1.3.

**Table 1**
Parameter Values Used to Create BC03 SED Models for Simulated Galaxies
at $z = 7$–10

| Parameter | Values |
| --- | --- |
| Metallicity ($Z_\odot$) | 1, 0.4, 0.2, 0.02[a] |
| SFH $\tau$ (Gyr) | 0.01, 0.1, 1, 10 |
| $A_V$ | 0, 0.5, 1 |
| Age (Gyr) | 0.01, 0.05, 0.2, 0.4, 0.6[b] |
| $z$ | 7, 8, 9, 10 |
| S/N[c] | 3, 5, 10 |

**Notes.**
[a] For metallicity 0.02 $Z_\odot$ we considered only ages $t < 0.2$ Gyr.
[b] We considered this age only up to redshift $z = 8$.
[c] For the flux corresponding to the normalization magnitude, i.e., [F150W] = 29 mag.

explicitly decided not to include the Ly$\alpha$ line because the importance of its emission in galaxies at $z > 7$ is still unclear (e.g., Stark et al. 2013; Konno et al. 2014; Pentericci et al. 2014). Finally, we normalized every model at 29 AB mag at 1.5 $\mu$m (corresponding to the NIRCam F150W filter pivot wavelength).

The complete list of adopted parameter values is summarized in Table 1, while some SED examples at $z = 7$ are shown in Figure 4. Our sample contains a total of 864 galaxies, corresponding to 240 BC03-based SED models at each redshift 7 and 8, and 192 SED models at each redshift 9 and 10. For each model, we considered independently three possible S/N values (S/N = 3, 5, and 10) for the flux corresponding to [F150W] = 29 AB mag. The S/N values of the photometry in all other bands have been scaled, as we explain in Section 2.2.1.

Second, we simulated further $z = 7$–10 galaxy SEDs with the population synthesis code Yggdrasil (Zackrisson et al. 2011), which has been especially conceived to model high-redshift galaxies. We used a set of parameter values similar to the ones used for the BC03 templates, except that the star formation histories are somewhat different, consisting of a step function for which the duration of the constant star formation period is set as a free parameter. And, in addition to the normal stellar templates, we also adopted a template with a metal-free, Population III single stellar population (Schaerer 2002). In all cases, we considered the two possible extreme gas covering factors, namely, $f_{cov} = 0$ and 1. The complete list of adopted parameter values is given in Table 2. We show four SED template examples for $z = 7$ galaxies in Figure 5.

The Yggdrasil models include the effects of nebular line and continuum emission, so we did not need to add any extra line manually as we did for the BC03 templates. We did include the effect of dust reddening and IGM absorption in the same way as for the BC03 models. We removed the Ly$\alpha$ emission line from the Yggdrasil models because its presence and importance are debated at $z > 7$, as we explained before. Once again, we normalized every SED template to [F150W] = 29 mag at 1.5 $\mu$m. Our Yggdrasil simulated sample comprises a total of 1260 galaxies, corresponding to 342 SED models at each redshift 7 and 8 and 288 SED models at each redshift 9 and 10. The complete redshift distribution of our galaxy sample 3, simulated with both the BC03 and Yggdrasil templates, is shown in Figure 2.

## 2.2. Test Methodology

### 2.2.1. Interpolation/Extrapolation of JWST Photometry

Based on the best-fit SEDs obtained running LePhare on the CANDELS photometry for the galaxies in samples 1 and 2 and the simulated SEDs for our mock galaxies in sample 3, we measured the photometry expected at the eight NIRCam broadbands and the two MIRI broadbands F560W and F770W for all our galaxies (see Table 3 for a summary of the different JWST filters considered in this work). Our procedure to interpolate/extrapolate the JWST photometry is similar for our three galaxy samples. The main difference is that in samples 1 and 2, we simply adopted S/N values for the photometry scaled from the CANDELS photometry, while for sample 3 we tested three different S/N values in a controlled manner, in order to assess the effect of S/N on our results.

As a first step, for every galaxy in our real galaxy samples (samples 1 and 2), we derived NIRCam broadband fluxes by convolving the best-fit SED template with the transmission curve of each JWST filter (Meyer et al. 2004). To each flux, we assigned an error bar corresponding to an S/N equal to that of the closest filter in the CANDELS catalog. We adopted as NIRCam flux a random value (selected from a uniform distribution) within the corresponding error bar.

Note that, in practice, NIRCam will reach the depth of the HST CANDELS photometry in very short integration times, so the S/N achieved for CANDELS galaxies in typically deep NIRCam surveys will be higher than what we consider here. However, our assumed S/Ns are already sufficiently high in most cases: virtually all galaxies in sample 1 and more than 90% of galaxies in sample 2 have S/N $\geqslant 5$ in their derived NIRCam F150W photometry. Thus, our results derived here for these real galaxy samples are mostly independent of S/N, as we explain below (the low-S/N effect will be more clearly manifested in our sample 3, where we explicitly set three different S/N values for each template, including a low value of S/N = 3).

As a second step, we measured the photometry expected at the MIRI bands for our sample 1 and 2 galaxies, which results from the extrapolation of the best-fit template redward of the CANDELS IRAC 4.5 $\mu$m coverage. We convolved the best SED model with the two MIRI bands' transmission curves (Bouchet et al. 2015; Glasse et al. 2015). To each MIRI band flux, we assigned an error bar corresponding to an S/N scaled using as reference the S/N in the NIRCam F150W band: we assumed that an MIRI F560W or F770W flux corresponding to a magnitude 28 (AB) would have the same S/N as a NIRCam F150W flux corresponding to 29 mag (whose S/N was, in turn, scaled assuming that the "measured" NIRCam F150W flux S/N for each galaxy was the same as the real CANDELS HST F160W flux S/N). Fixing the S/N in the NIRCam and MIRI photometry implicitly assumes specific integration times with the JWST instruments. In practice, this assumption does not matter, as our results will be equally valid for any galaxies with similar S/N values, independently of the MIRI fluxes at which these S/Ns are reached.[15] Finally, we randomized every flux within its own error bar.

For sample 3 with mock galaxies, we convolved every simulated SED template with the transmission curve of each

---

[15] This statement implicitly assumes that the fainter galaxy SEDs will be represented by those of the currently analyzed galaxies. So far, this has been shown to be the case with progressively deeper galaxy surveys.
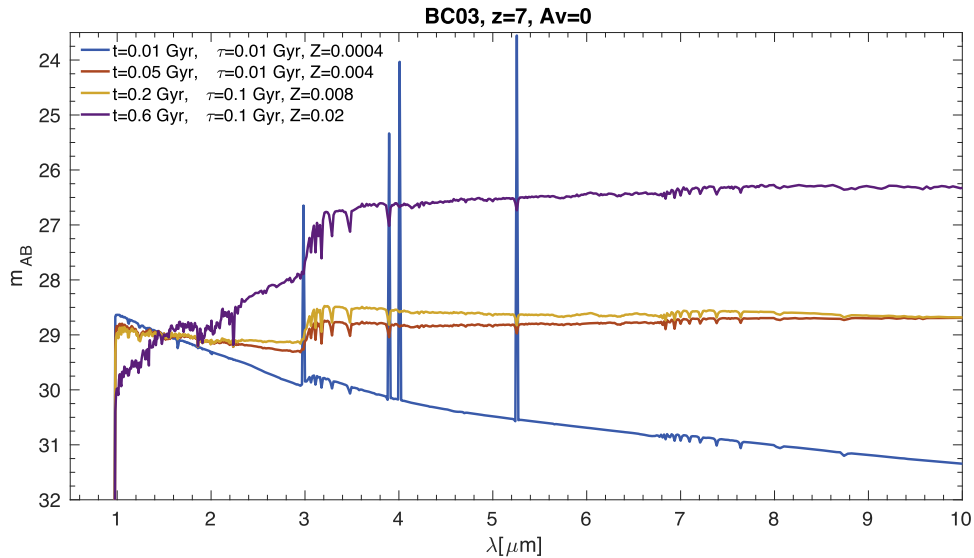
**Figure 4.** Four examples of BC03 SED templates at $z = 7$ with $A_V = 0$.

**Table 2**
Parameter Values Used to Create Yggdrasil SED Models for Simulated Galaxies at $z = 7$–$10$

| Parameter | Values |
|---|---|
| Metallicity ($Z_\odot$) | 1, 0.4, 0.2, 0.02,[a] 0[a] |
| Constant SFH (Myr) | 10, 30, 100 |
| $A_V$ | 0, 0.5, 1 |
| Age (Gyr) | 0.01, 0.05, 0.2, 0.4, 0.6[b] |
| $f_{cov}$ | 0, 1 |
| $z$ | 7, 8, 9, 10 |
| S/N[c] | 3, 5, 10 |

**Notes.**
[a] For this metallicity we considered only ages $t < 0.2$ Gyr.
[b] We considered this age only up to redshift $z = 8$.
[c] For the flux corresponding to the normalization magnitude, i.e., [F150W] = 29 mag.

NIRCam filter. In our sample every SED template is considered three times, in order to apply three fixed, reference S/N values (10, 5, and 3) to the NIRCam F150W fluxes. We assumed the same integration times for all the other NIRCam bands and took into account the differences in filter sensitivities to assign S/N in each band. A different strategy that assumes the same observational depth for all NIRCam bands is presented in Appendix B. We considered every flux below the $2\sigma$ level as a nondetection. The $2\sigma$ detection limits in each band for each fixed reference S/N value at F150W are listed in Table 4. Note that, as a conservative approach, we adopted $3\sigma$ flux upper limits (rather than $2\sigma$) in all cases of nondetections for running the $z_{phot}$ code, as is explained below.

Finally, we obtained 10 random realizations of each NIRCam band flux within the corresponding error bar. Note that, in contrast, for samples 1 and 2 we considered a single realization of each flux, simply because in the case of real galaxies the NIRCam fluxes are well constrained by interpolation of the CANDELS fluxes, so their error bars are typically smaller than in sample 3.

In a similar way, we derived fluxes in the MIRI bands by convolving all SED templates in the sample with the two MIRI filter transmission curves. Then, we scaled each MIRI flux S/

N, such that a $\text{mag}_{AB} = 28$ in MIRI has the same S/N as a NIRCam magnitude [F150W] = 29. This is a realistic assumption of how the NIRCam and MIRI data will typically be matched. As for the NIRCam bands, we randomized every MIRI flux 10 times within its own error bar.

### 2.3. Photometric Redshift Determinations

With the expected *JWST* photometry measured on the galaxy templates for the three samples, we could test how accurately it is possible to derive photometric redshifts by using different *JWST* band combinations. The band combinations that we analyzed in this work are as follows:

1. Eight NIRCam broadbands.
2. *HST* F435W and F606W, and eight NIRCam broadbands.
3. VLT *U* band, *HST* F435W and F606W, and eight NIRCam broadbands.
4. Eight NIRCam broadbands and two MIRI bands (F560W and F770W).
5. Eight NIRCam broadbands and MIRI F560W only.
6. Eight NIRCam broadbands and MIRI F770W only.
7. *HST* F435W and F606W, eight NIRCam broadbands, and two MIRI bands.
8. *HST* F435W and F606W, eight NIRCam broadbands, and MIRI F560W.
9. *HST* F435W and F606W, eight NIRCam broadbands, and MIRI F770W.

These different filter combinations have been chosen to test in which cases the NIRCam broadbands alone are sufficient to recover good photometric redshifts and in which cases the MIRI bands can improve these estimates. Because MIRI is considerably less sensitive than NIRCam, much longer integration times are necessary to reach comparable depths. Thus, many observing programs may opt for using a single MIRI filter (F560W or F770W) rather than both of them. This is why we tested separately the cases in which a single and both MIRI bands are used.

In addition, we analyzed the cases in which two short-wavelength *HST* bands and the *U* band are included. All these bands are beyond *JWST*'s wavelength coverage, so they cannot
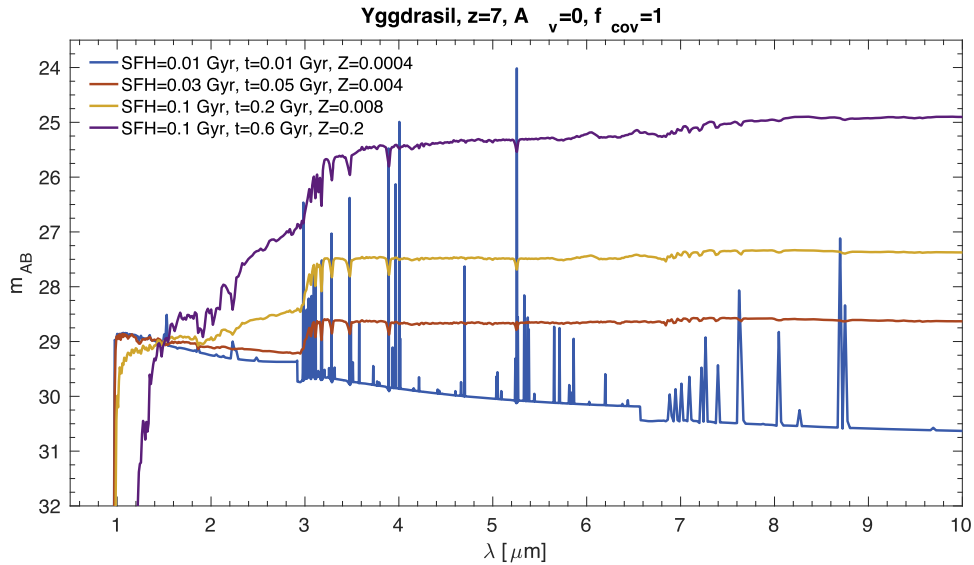
**Figure 5.** Four examples of Yggdrasil SED templates at $z = 7$ with $A_V = 0$. Note that templates with gas covering factor $f_{cov} = 1$ have multiple emission lines, but only when a young stellar population, due to very recent star formation, is still present.

**Table 3**
List of *JWST* Filters Considered in This Work, Their Pivot Wavelengths, and Their Bandwidths

| Filter Name | $\lambda$ ($\mu$m) | Bandwidth ($\mu$m)[a] |
|---|---|---|
| NIRCam/F070W | 0.70 | 0.18 |
| NIRCam/F090W | 0.90 | 0.23 |
| NIRCam/F115W | 1.15 | 0.30 |
| NIRCam/F150W | 1.50 | 0.38 |
| NIRCam/F200W | 1.99 | 0.53 |
| NIRCam/F277W | 2.77 | 0.82 |
| NIRCam/F356W | 3.56 | 0.99 |
| NIRCam/F444W | 4.41 | 1.29 |
| MIRI/F560W | 5.60 | 1.41 |
| MIRI/F770W | 7.70 | 2.30 |

**Note.**
[a] This bandwidth corresponds to the entire wavelength range in which the filter transmission is >1%.

be substituted with any *JWST* band. We used the CANDELS catalog to recover observations in these bands for galaxies in samples 1 and 2. We ignored these bands for sample 3: they are irrelevant, as the Lyman break is completely contained within the NIRCam wavelength range at $z \geqslant 7$.

We note that LePhare has been run using the native BC03 models, allowing the code to incorporate emission lines with its own prescription in the cases of samples 1 and 2 (as is described in Section 2.1.1). For our mock galaxies in sample 3, we ran LePhare using directly our own customized templates, i.e., the BC03 templates with emission lines incorporated as explained in Section 2.1.3 and the Yggdrasil templates. In addition, in the LePhare runs for sample 3 we also included some older galaxy models (as in samples 1 and 2) and allowed for extinction values from $A_V = 0$–4 in all cases, in order to test whether there could be degeneracies between redshift and dust/age produced in the photometric fitting. This consideration of a wide range of redshifts, ages, and extinction values emulates the real situation, in which one simply has a photometric input catalog without knowing a priori which sources are at high or low redshifts. No extra consideration of emission lines has been

allowed for sample 3 in LePhare, as our young-galaxy models already account for them.

For all the filter combinations we ran LePhare on the three galaxy samples, in order to recover the galaxy photometric redshifts and compare them with the fiducial redshift available for each galaxy. In the filter combinations with MIRI bands in samples 1 and 2, and all cases for sample 3, we chose the median photometric redshift of the 10 iterations performed for each object (corresponding to different variations in the photometry, as explained above).

For each sample and filter set combination, in Section 3 we quote the resulting mean value of the normalized redshift difference distribution $\delta z = (z_{phot} - z_{fiduc.})/(1 + z_{fiduc.})$ and the rms ($\sigma$) of the *normalized* absolute redshift difference distribution $|z_{phot} - z_{fiduc.}|/(1 + z_{fiduc.})$. Note that, although we computed $\sigma$ independently for each sample and broadband combination, we considered as reference to define outliers a single $\sigma$ value per sample (the minimum obtained with all the filter sets). This allows us to directly compare how the outlier fractions change among the different filter combinations.

We warn the reader that the specific values of the statistical quantities ($\sigma$, outlier percentages) quoted in this paper refer to the samples analyzed here and may have variations for other galaxy samples with different redshifts and SED-type distributions. Even so, as all our tests are internally consistent, our results constitute a useful reference to assess the ability to recover galaxy redshifts using different *JWST* filter combinations and ancillary data.

## 3. RESULTS

In this section we present our results for the three galaxy samples analyzed in this work.

### 3.1. Results for Sample 1

In Figures 6–8 we show the comparison between the spectroscopic and our derived photometric redshifts obtained with different filter combinations for sample 1. For each case, we quote the mean of the $(z_{phot} - z_{spec})/(1 + z_{spec})$ distribution and the rms of $|z_{phot} - z_{spec}|/(1 + z_{spec})$, computed considering
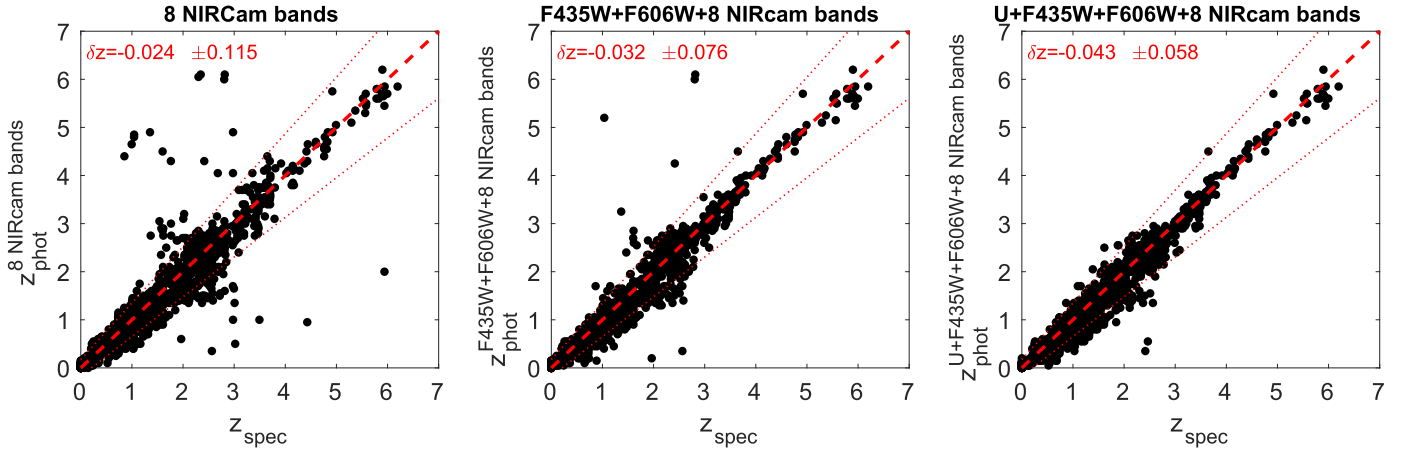
**Figure 6.** Photometric vs. spectroscopic redshifts for galaxies in sample 1. Left: photometric redshifts obtained with eight NIRCam broadbands. Middle: photometric redshifts obtained with *HST* F435W, *HST* F606W, and eight NIRCam broadbands. Right: same as the middle panel, but considering also the VLT *U* band for computing photometric redshifts. The dotted lines in each panel delimit the $3\sigma$ limit of the $|z_{\rm phot} - z_{\rm spec}|/(1 + z_{\rm spec})$ distribution, with $\sigma = 0.058$, which is the minimum value of $\sigma$ that we get for sample 1 with all the considered filter combinations.

**Table 4**
List of $2\sigma$ Detection Magnitude Limits for Different *JWST* Filters Corresponding to Our Fiducial Source[a] and Each Reference S/N Value Considered for Sample 3

| Reference S/N[a] | F070W | F090W | F115W | F150W | F200W | F277W | F356W | F444W | F560W | F770W |
|---|---|---|---|---|---|---|---|---|---|---|
| 10 | 30.2 | 30.5 | 30.6 | 30.7 | 30.9 | 30.9 | 30.7 | 30.2 | 29.7 | 29.7 |
| 5 | 29.4 | 29.7 | 29.9 | 30.0 | 30.2 | 30.2 | 30.0 | 29.4 | 29.0 | 29.0 |
| 3 | 28.9 | 29.2 | 29.3 | 29.4 | 29.6 | 29.6 | 29.4 | 28.9 | 28.4 | 28.4 |

**Note.**
[a] S/N value for [F150W] = 29 mag.

**Table 5**
Statistical Properties of Sample 1 for All the Different *JWST* and Ancillary Data Broadband Combinations

| Bands | $z = 1$–4 | $\langle \delta z \rangle$ | $\sigma_{|\delta z|}$ |
|---|---|---|---|
| Eight NIRCam broadbands | 134 (10.3%) | −0.024 | 0.115 |
| *HST* F435W + *HST* F606W + eight NIRCam bands | 93 (7.1%) | −0.032 | 0.076 |
| VLT *U* + *HST* F435W, F606W + eight NIRCam | 89 (6.8%) | −0.043 | 0.058 |
| Eight NIRCam bands + MIRI F560W, F770W | 116 (8.9%) | −0.037 | 0.066 |
| Eight NIRCam bands + MIRI F560W | 120 (9.2%) | −0.029 | 0.077 |
| Eight NIRCam bands + MIRI F770W | 128 (9.8%) | −0.037 | 0.067 |
| *HST* F435W, F606W + eight NIRCam bands + MIRI F560W, F770W | 109 (8.3%) | −0.042 | 0.062 |
| *HST* F435W, F606W + eight NIRCam bands + MIRI F560W | 93 (7.1%) | −0.034 | 0.060 |
| *HST* F435W, F606W + eight NIRCam bands + MIRI F770W | 106 (8.1%) | −0.042 | 0.064 |

**Note.** The last two columns indicate the mean of $(z_{\rm phot} - z_{\rm spec})/(1 + z_{\rm spec})$ and the rms ($\sigma$) of $|z_{\rm phot} - z_{\rm spec}|/(1 + z_{\rm spec})$, respectively. The second column provides the number and percentages of $z = 1$–4 sources that become outliers, i.e., objects beyond $3\sigma$ in the $z_{\rm phot}-z_{\rm spec}$ relation, with $\sigma = 0.058$, which is the minimum $\sigma$ obtained with all the filter combinations considered for this galaxy sample. The total original number of $z = 1$–4 galaxies in sample 1 is 1306.

all galaxies in the sample. We identified $3\sigma$ outliers using as a reference $\sigma = 0.058$, which is the minimum of all the $\sigma$ values that we get for sample 1. This galaxy sample is particularly useful to test the percentage of low-$z$ galaxies ($z < 4$) that leak to wrong redshifts due to failures in the redshift recovery. The percentages of these low-$z$ sources that are leaking outliers for each filter combination are listed in Table 5. For this analysis, we considered only the redshift bin $z = 1$–4, avoiding lower-$z$ galaxies, as the MIRI filters are not very suitable for tracing the direct stellar emission at such low $z$. Note that, although these obtained outlier percentages apply particularly to this galaxy sample, the changes obtained with different passbands are of

general validity and show the impact of incorporating different filters on the redshift recovery.

First, we analyzed the case of using the eight NIRCam broadbands with and without ancillary *HST* and ground-based *U*-band data (Figure 6). We found that when using only the eight NIRCam broadbands, the proportion of $z = 1$–4 sources that are $3\sigma$ outliers is 10.3%, including several extreme cases that appear as $z > 4$ contaminants. Besides, the rms of the $|z_{\rm phot} - z_{\rm spec}|/(1 + z_{\rm spec})$ distribution is high ($\sigma = 0.115$) compared with the typical precision that can currently be achieved for low-$z$ galaxies in deep surveys. When adding shorter-wavelength *HST* data, the normalized redshift difference distribution rms is significanlty lower, the outlier
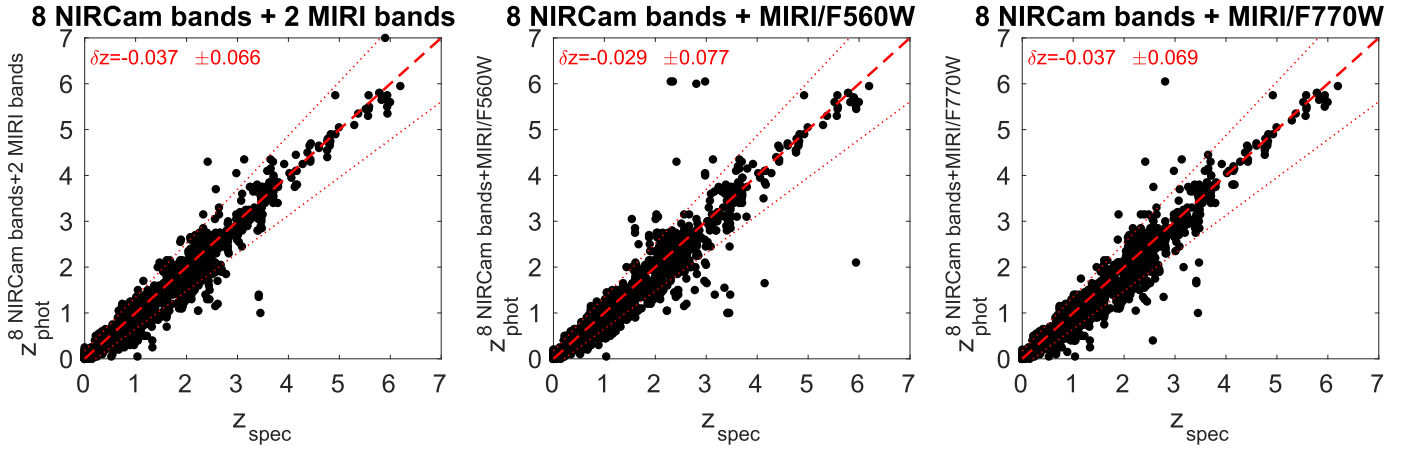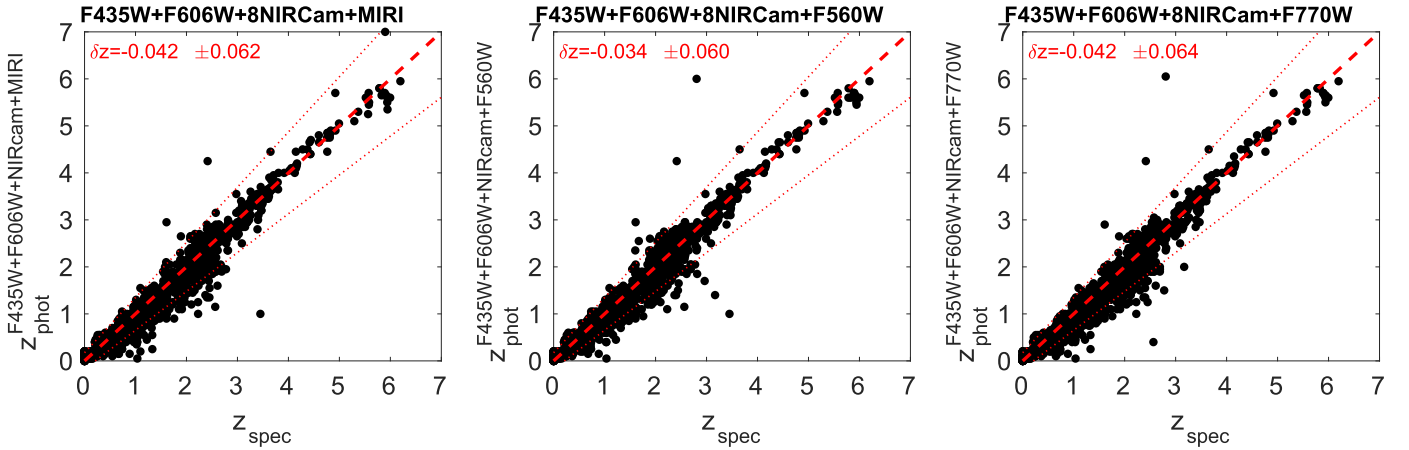
**Figure 7.** Photometric vs. spectroscopic redshifts for galaxies in sample 1. Left: photometric redshifts obtained with eight NIRCam broadbands and MIRI F560W and F770W. Middle: photometric redshifts obtained with eight NIRCam broadbands and MIRI F560W only. Right: photometric redshift obtained with eight NIRCam broadbands and MIRI F770W only. The dotted lines in each panel delimit the $3\sigma$ limit of the $|z_{\rm phot} - z_{\rm spec}|/(1 + z_{\rm spec})$ distribution, with $\sigma = 0.058$, which is the minimum value of $\sigma$ that we get for sample 1 with all the considered filter combinations.



**Figure 8.** Photometric vs. spectroscopic redshifts for galaxies in sample 1. Left: photometric redshifts obtained with *HST* F345W and F606W, eight NIRCam broadbands, and MIRI F560W and F770W. Middle: photometric redshifts obtained with *HST* F345W and F606W, eight NIRCam broadbands, and MIRI F560W. Right: photometric redshifts obtained with *HST* F345W and F606W, eight NIRCam broadbands, and MIRI F770W. The dotted lines in each panel delimit the $3\sigma$ limit of the $|z_{\rm phot} - z_{\rm spec}|/(1 + z_{\rm spec})$ distribution, with $\sigma = 0.058$, which is the minimum value of $\sigma$ that we get for sample 1 with all the considered filter combinations.

percentage decreases to 7.1%, and the majority of the most extreme redshift outliers are corrected. When adding also the *U* band, the rms reaches a minimum ($\sigma = 0.058$), and so does the outlier percentage (6.8%), leaving basically no catastrophic outlier cases. This is because it is necessary to cover wavelengths shorter than 0.7 $\mu$m in order to distinguish between the Lyman and 4000 Å breaks at $z < 6$. The improvement on the photometric redshift estimation is also due to the mere increase in the wavelength coverage given by the larger number of bands. Besides, the incorporation of the *HST* and *U*-band data into the NIRCam photometry reduces the photometric redshift dispersion ($\sigma$) to about one-half (Figure 6).

Second, we tested the photometric redshift estimation when complementing the NIRCam photometry with MIRI photometry, instead of shorter-wavelength data (Figure 7). When we incorporate both MIRI F560W and F770W bands, the resulting normalized redshift difference distribution rms is significantly lower ($\sigma = 0.066$) than in the case with NIRCam data alone, indicating that the presence of MIRI photometry can mitigate the absence of short-wavelength data (Table 5).

Lastly, we considered the situation in which *HST* ancillary data along with NIRCam and MIRI data are available (Figure 8). We explicitly tested a combination without *U*-band photometry, as ground-based *U*-band observations with matching depth to *JWST* observations will be time-consuming and difficult to achieve. The results are rather similar to the case with NIRCam and MIRI bands considered together: the obtained rms is $\sigma = 0.062$, with the fraction of redshift outliers being slightly reduced. Interestingly, when only NIRCam and MIRI photometry is considered, the F770W data improve the overall redshift statistics. Instead, when *HST* data are also available, the incorporation of F560W alone provides slightly better results. (This is related to how the $\chi^2$-fitting procedure works, typically giving more weight to longer-wavelength data where the sources are brighter.)

To conclude, the eight NIRCam broadbands alone are not adequate to obtain sufficiently good photometric redshifts at $z < 4$. Although for most galaxies the derived redshifts are correct, the overall obtained statistics is poor compared with the typical $z_{\rm phot}$ quality that can currently be achieved in deep
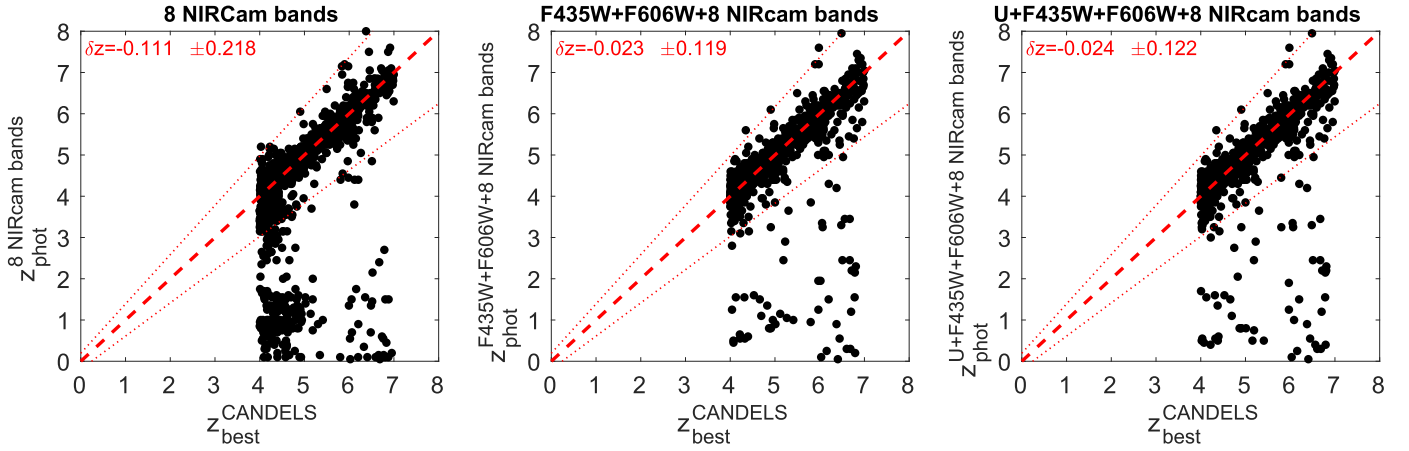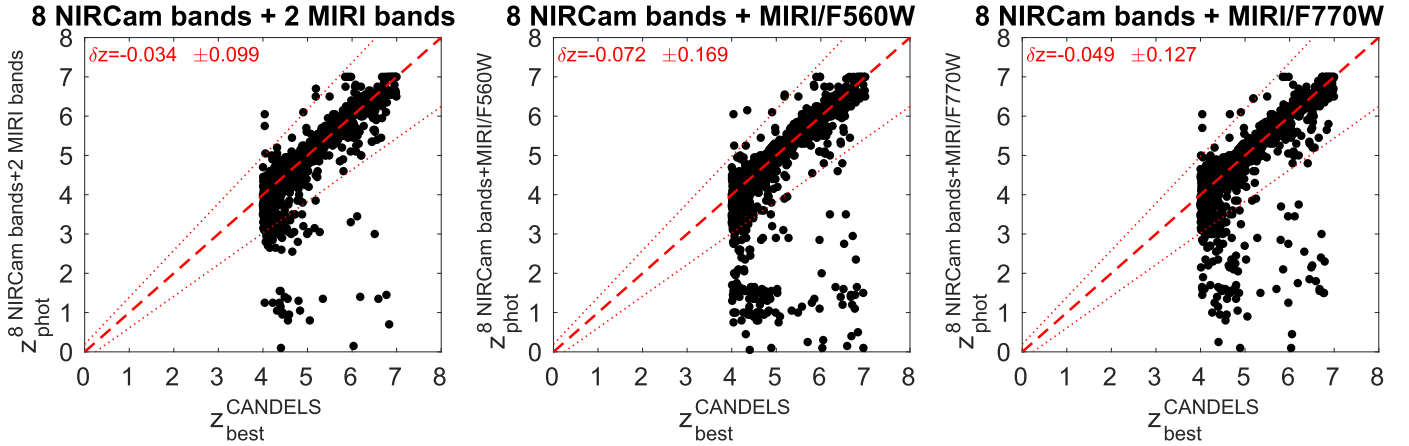
**Figure 9.** Our derived photometric redshifts vs. CANDELS consensus photometric redshifts for galaxies in sample 2. Left: photometric redshifts obtained with eight NIRCam broadbands. Middle: photometric redshifts obtained with *HST* F435W and F606W and eight NIRCam broadbands. Right: photometric redshifts obtained with the same bands as in the middle panel plus the VLT *U* band. The dotted lines in each panel delimit the $3\sigma$ limit of the $|z_{\mathrm{phot}} - z_{\mathrm{fiduc.}}|/(1 + z_{\mathrm{fiduc.}})$ distribution, with $\sigma = 0.065$, which is the minimum value of $\sigma$ that we get for sample 2 with all the considered filter combinations.



**Figure 10.** Our derived photometric redshifts vs. CANDELS consensus photometric redshifts for galaxies in sample 2. Left: photometric redshifts obtained with eight NIRCam broadbands and MIRI F560W and F770W. Middle: photometric redshifts obtained with eight NIRCam broadbands and MIRI F560W only. Right: photometric redshifts obtained with eight NIRCam broadbands and MIRI F770W only. The dotted lines in each panel delimit the $3\sigma$ limit of the $|z_{\mathrm{phot}} - z_{\mathrm{fiduc.}}|/(1 + z_{\mathrm{fiduc.}})$ distribution, with $\sigma = 0.065$, which is the minimum value of $\sigma$ that we get for sample 2 with all the considered filter combinations.

**Table 6**
Statistical Properties for Sample 2 for All the Different *JWST* and Ancillary Data Broadband Combinations

| Bands | $z = 4\text{--}5$ | $z = 5\text{--}7$ | $<\delta z>$ | $\sigma_{<|\delta z|>}$ |
|---|---|---|---|---|
| Eight NIRCam broadbands | 190 (21.4%) | 46 (9.4%) | −0.111 | 0.218 |
| *HST* F435W, F606W + eight NIRCam bands | 21 (2.4%) | 36 (8.0%) | −0.023 | 0.119 |
| VLT *U* + *HST* F435W, F606W + eight NIRCam bands | 23 (2.6%) | 37 (8.2%) | −0.024 | 0.122 |
| Eight NIRCam bands + MIRI F560W, F770W | 68 (7.7%) | 14 (3.1%) | −0.034 | 0.099 |
| Eight NIRCam bands + MIRI F560W | 120 (13.5%) | 41 (9.1%) | −0.072 | 0.169 |
| Eight NIRCam bands + MIRI F770W | 90 (10.1%) | 31 (6.9%) | −0.049 | 0.127 |
| *HST* F435W, F606W + eight NIRCam bands + MIRI F560W, F770W | 11 (1.2%) | 7 (1.4%) | −0.004 | 0.065 |
| *HST* F435W, F606W + eight NIRCam bands + MIRI F560W | 20 (2.2%) | 32 (6.6%) | −0.019 | 0.113 |
| *HST* F435W, F606W + eight NIRCam bands + MIRI F770W | 17 (1.9%) | 18 (4.0%) | −0.011 | 0.083 |

**Note.** The last two columns indicate the mean of $(z_{\mathrm{phot}} - z_{\mathrm{spec}})/(1 + z_{\mathrm{spec}})$ and the rms ($\sigma$) of $|z_{\mathrm{phot}} - z_{\mathrm{spec}}|/(1 + z_{\mathrm{spec}})$, respectively. Columns (2) and (3) provide the numbers and percentages of galaxies leaking toward lower redshifts. These outliers are defined as galaxies beyond $3\sigma$, with $\sigma = 0.065$, which is the minimum $\sigma$ obtained with the different filter combinations considered for this galaxy sample. The total original number of $z = 4\text{--}5$ ($z = 5\text{--}7$) galaxies in sample 2 is 887 (488).

surveys at these redshifts. Observations at wavelengths shorter than 0.7 $\mu$m are necessary to remove the most extreme redshift outliers. Incorporating MIRI longer-wavelength photometry instead can mitigate the absence of short-wavelength observations. Having a complete optical and infrared coverage up to 7.7 $\mu$m will allow us to keep the $z < 4$ redshift statistics well under control.

### 3.2. Results for Sample 2

In Figures 9–11 we show the comparison between our recovered photometric redshifts and fiducial CANDELS consensus redshifts for galaxies in sample 2 (all of which have fiducial redshifts $z > 4$). As we did before, we derived $|z_{phot} - z_{fiduc.}|/(1 + z_{fiduc.})$ for all our galaxies and identified $3\sigma$ outliers, where we adopted $\sigma = 0.065$, which is the smallest rms value that we obtained with the different filter combinations. The percentages of outliers in each case are listed in Table 6.

As for sample 1, our first step consisted of analyzing the resulting photometric redshifts obtained using the eight NIRCam broadbands with and without the *HST* and *U*-band data (Figure 9). We found that when using only the eight NIRCam broadbands, the percentage of outliers for sample 2 is high. We did not show the percentage of low-$z$ contaminants because of the sample selection, which by construction only has $z > 4$ objects. The most evident effect in sample 2 is the leakage of high-$z$ sources to low-$z$ values produced by the insufficient photometric coverage. This is particularly important for galaxies with input redshifts $z = 4$–5: a bit more than 20% of them are found $>3\sigma$ away at lower redshifts (Table 6).

The redshift estimation is better at higher redshifts ($z = 5$–7). The proportion of sources leaking to lower $z$ is only ~9%. As we discussed before, the Lyman break enters the shortest-wavelength NIRCam band at $z \sim 5$–6; therefore, the degeneracy between the 4000 Å and Lyman breaks occurs mainly at lower redshifts. For this reason, adding the *HST* F435W and F606W photometry does not have an important effect on the redshift estimation at these higher redshifts, but it does at $z = 4$–5, where it reduces the percentage of leaking sources to only ~2%. The overall redshift distribution dispersion reduces to one-half when the *HST* data are considered along the NIRCam photometry (Figure 9).

The addition of the *U* band does not produce any obvious effect in the redshift estimation for sample 2, once the *HST* photometry is also considered. The role of the *U* band is more important for identifying low-redshift contaminants rather than preventing the leakage of high-$z$ sources to low $z$, because galaxies at $z > 4$ are very faint or not detected in this band. As we will discuss later, the *HST* optical bands have a similar effect at $z > 5$.

As a second step, we obtained photometric redshifts with the eight NIRCam broadbands and two MIRI bands F560W and F770W, considered together and one at a time (Figure 10). Adding both MIRI bands significantly improves the photometric redshift estimates by decreasing the percentage of leaking sources to <8% at $z = 4$–5 and <4% at $z = 5$–7. Among the remaining $z = 4$–5 leaking sources, some are placed at $z \sim 1$ and a few at higher redshifts, i.e., $z \sim 6.5$. When considering one MIRI band at a time, F770W appears to be more effective at improving the photometric redshift estimation than F560W, but it is the combined effect of both

MIRI bands that produces the substantial reduction in the fraction of leaking galaxies.

Adding MIRI photometry appears to be more important than adding *HST* photometry to minimize the dispersion in the obtained redshift distribution: the resulting rms in the case of NIRCam and MIRI data is $\sigma = 0.099$, which is slightly lower than for the *HST* and NIRCam filter combination ($\sigma = 0.119$).

Lastly, we considered *HST*, NIRCam, and MIRI photometry altogether to obtain photometric redshifts (Figure 11). As expected, using all these bands jointly provides the best results in the identification of $z > 4$ galaxies. The presence of *HST* photometry at $\lambda < 0.6$–0.7 $\mu$m helps to correctly identify the Lyman break at $z = 4$–5, while the MIRI mid-infrared bands provide a better wavelength coverage to correctly identify sources at $z = 5$–7. The resulting redshift dispersion gets a minimum value of $\sigma = 0.065$, and the fraction of $z = 4$–5 sources leaking beyond $3\sigma$ in redshift decreases to ~1%.

As before, the MIRI F770W filter appears to have a more important role in improving the redshift estimation than MIRI F560W, but both bands are necessary to properly identify all galaxies at $z = 4$–7. To sum up, the short-wavelength *HST* data and long-wavelength MIRI data have complementary roles in the correct identification of $z > 4$ galaxies.

We remind the reader that the photometry in our samples 1 and 2 is of good quality (S/N > 5 in F150W for the vast majority of galaxies; see Section 2.2.1). The redshift outlier percentages that we quote here correspond to this high-quality photometry, so they do not depend on the data quality, but are rather produced by the intrinsic limitations of the SED fitting with a limited number of photometric bands. We checked the S/N of the photometry for the $z_{phot}$ outliers and found no significant difference with respect to the photometry of all other sources, confirming that the results for samples 1 and 2 are not affected by the photometric S/N.

### 3.3. Results for Sample 3

With sample 3 we aim to test the ability to recover photometric redshifts for simulated galaxies at $z = 7$–10, using the NIRCam broadbands alone and in conjunction with the MIRI F560W and F770W bands. For this sample, we did not consider the incorporation of the *HST* short-wavelength and *U* filters because these bands map rest wavelengths far blueward of the Lyman break at these redshifts, so they basically correspond to nondetections in all cases.

#### 3.3.1. Galaxies Simulated with BC03 Templates at $z = 7$–10

Figures 12–15 show the distributions of the recovered photometric redshifts for each input redshift ($z = 7$–10) for the sample 3 galaxies simulated with the BC03 templates, obtained by running LePhare with different *JWST* filter combinations. Since in this case the input redshifts are four fixed values, it is clearer to show the output $z_{phot}$ distributions for each input redshift rather than $z_{phot} - z_{phot}$ plots, which would contain four vertical columns each.

As before, in each plot we quote the mean value of the normalized redshift difference distribution $\delta z = (z_{phot} - z_{fiduc.})/(1 + z_{fiduc.})$, taking into account all galaxies. However, for sample 3 galaxies, we adapted our method for the rms calculation. As can be seen in Figures 12–15, the sample 3 redshift histograms are characterized by quite a narrow distribution around the main peak and the presence of some catastrophic failures producing
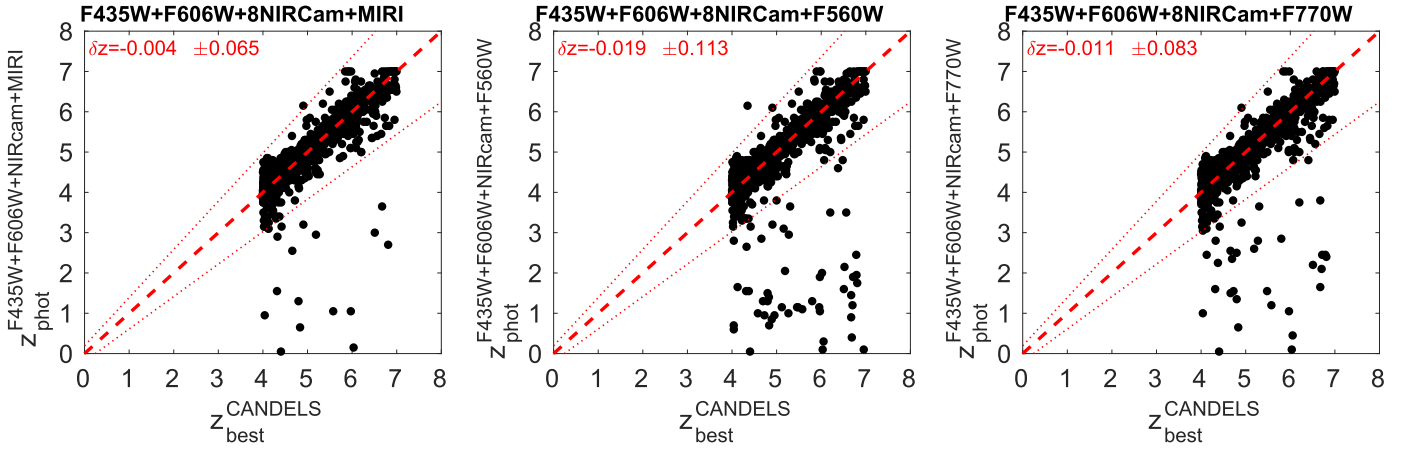
**Figure 11.** Our derived photometric redshifts vs. CANDELS consensus photometric redshifts for galaxies in sample 2. Left: photometric redshifts obtained with *HST* F345W and F606W, eight NIRCam broadbands, and MIRI F560W and F770W. Middle: photometric redshifts obtained with *HST* F345W and F606W, eight NIRCam broadbands, and MIRI F560W. Right: photometric redshifts obtained with *HST* F345W and F606W, eight NIRCam broadbands, and MIRI F770W. The dotted lines in each panel delimit the $3\sigma$ limit of the $|z_{\rm phot} - z_{\rm fiduc.}|/(1 + z_{\rm fiduc.})$ distribution, with $\sigma = 0.065$, which is the minimum value of $\sigma$ that we get for sample 2 with all the considered filter combinations.



**Figure 12.** Photometric redshifts obtained for the BC03 simulated galaxies with (F150W) S/N = 10 at different fixed redshifts. From top to bottom: redshifts $z = 7$–$10$. Photometric redshifts in each column are obtained with different combinations of bands. From left to right: eight NIRCam broadbands; eight NIRCam broadbands, MIRI F560W, and MIRI F770W; eight NIRCam broadbands and MIRI F560W only; eight NIRCam broadbands and MIRI F770W only. Each row corresponds to one of the four specific input redshifts. The vertical lines indicate the $3\sigma$ interval around the mean normalized redshift difference.
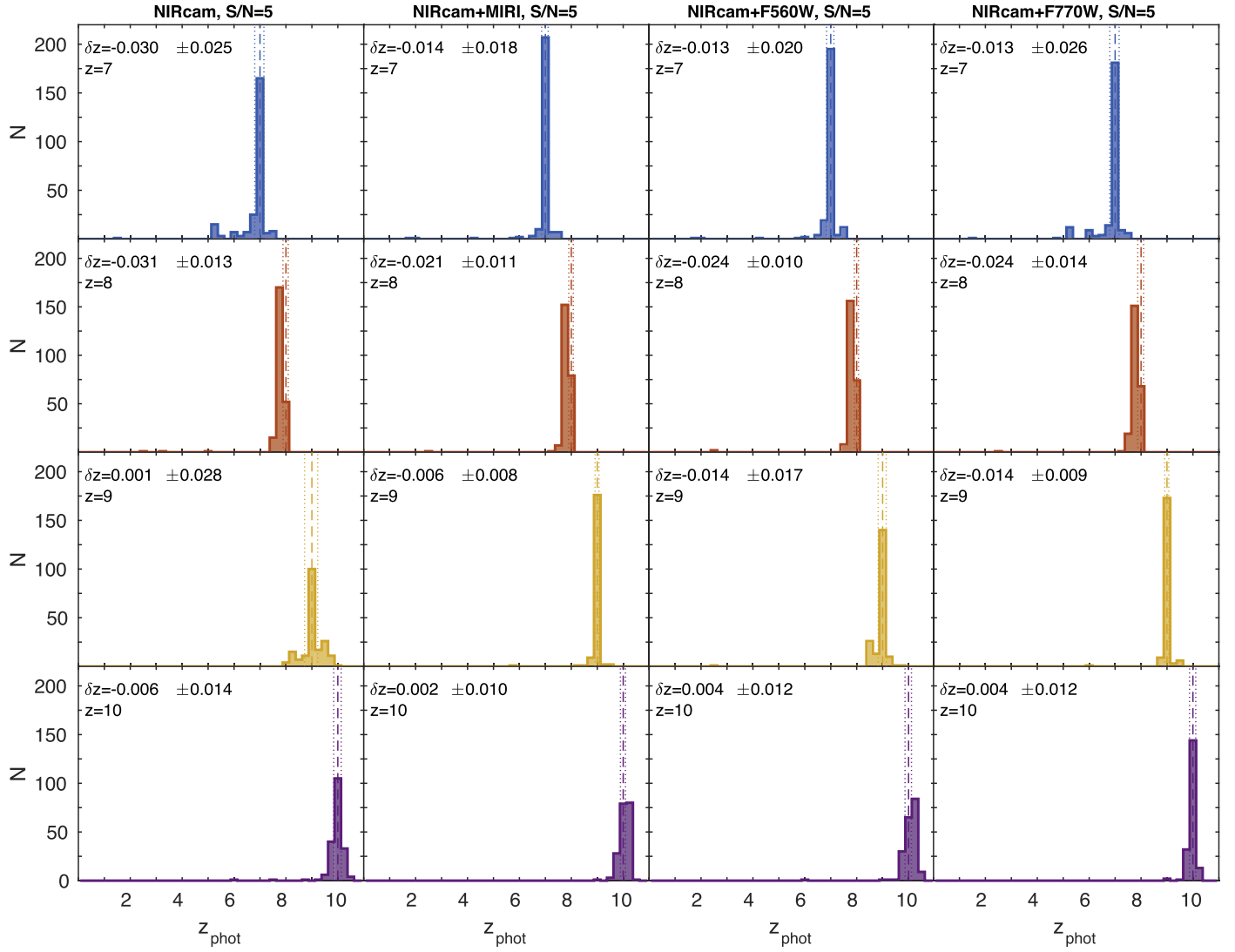
11

**Figure 13.** Same as Figure 12, but for (F150W) S/N = 5.

**Table 7**
Number of Outliers among Our 864 Sample 3 Galaxies Simulated with BC03 Models and Manual Addition of Emission Lines for Different *JWST* Filter Combinations and Assumed S/N Values at F150W

| Bands | S/N = 10 | S/N = 5 | S/N = 5 (Only EL)[a] | S/N = 3 |
|---|---|---|---|---|
| Eight NIRCam broadbands | 1 (0.1%) | 24 (2.8%) | 20[b] (3.5%) | 87 (10.1%) |
| Eight NIRCam bands + MIRI F560W, F770W | 1 (0.1%) | 5 (0.6%) | 3[b] (0.5%) | 26 (3.0%) |
| Eight NIRCam bands + MIRI F560W | 1 (0.1%) | 7 (0.8%) | 3[b] (0.5%) | 29 (3.4%) |
| Eight NIRCam bands + MIRI F770W | 2 (0.2%) | 16 (1.9%) | 14[b] (2.4%) | 45 (5.2%) |

**Notes.** Outliers are defined as galaxies with $|z_{phot} - z_{fiduc.}|/(1 + z_{fiduc.}) > 0.15$.
[a] Only templates with emission lines (EL).
[b] Out of a total of 576 galaxies with emission lines. The percentages refer to this denominator.

secondary peaks at much lower redshifts. These catastrophic redshift failures would significantly bias the computed $\sigma$ of $|z_{phot} - z_{fiduc.}|/(1 + z_{fiduc.})$ if they were considered in the statistics. So, for sample 3 galaxies (both based on the BC03 and Yggdrasil models) we explicitly excluded all galaxies with $|z_{phot} - z_{fiduc.}|/(1 + z_{fiduc.}) > 0.15$ for the $\sigma$ computation, and we considered those excluded galaxies to be the outliers of our

sample. We provide the statistics on the outlier fractions for the BC03 SED galaxies in Table 7.

For S/N = 10, the eight NIRCam broadbands alone allowed us to derive accurate redshift estimates, with mean values for the redshift distributions close to 0 and $\sigma \leqslant 0.011$ at all redshifts (Figure 12). The percentage of outliers quoted in Table 7 is very small (0.1%). Adding MIRI photometry to the
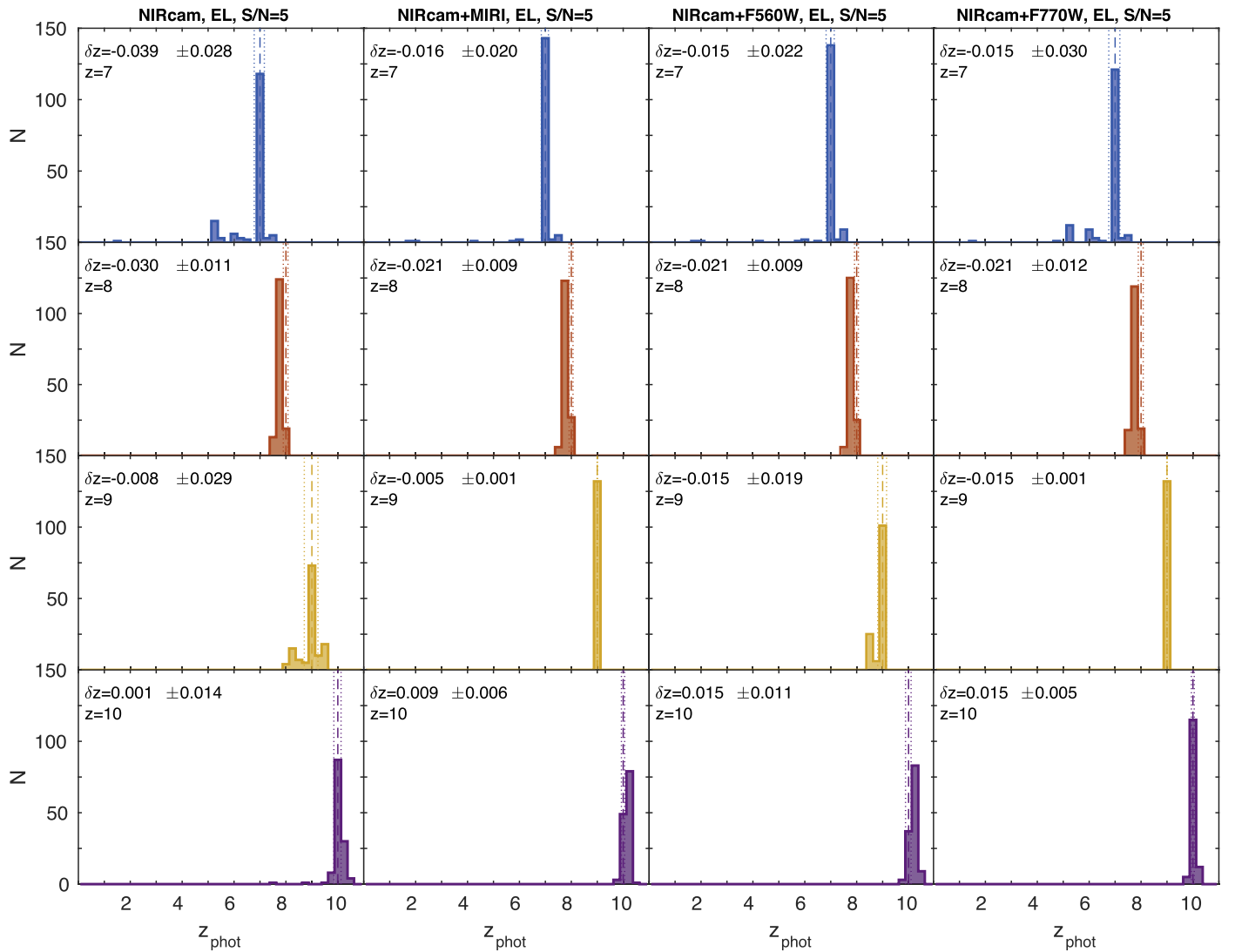
**Figure 14.** Same as Figure 13, but only for galaxies with emission lines.

NIRCam photometry only slightly enhances the already excellent quality of the output redshifts, resulting in $\sigma < 0.01$ at all redshifts. These results indicate that the incorporation of the MIRI photometry into the NIRCam photometry has in practice little effect on deriving photometric redshifts for $z = 7$–10 galaxies with BC03-like SEDs when the NIRCam S/N is very high.

The MIRI effect on the photometric redshift estimation is more evident at S/N = 5 (Figure 13). The NIRCam data alone are sufficient for a reasonably good redshift recovery ($\sigma \approx 0.01$–0.03), but the addition of MIRI data improves the photometric redshift quality, taking the outlier percentage to <1% (when F560W photometry is included). Inspection of Figure 4 suggests that this is a $k$-correction effect: galaxies are brighter in the MIRI bands either because they are old or because of the presence of emission lines. Indeed, both effects are important. In Figure 14 we show separately our results only for galaxies with emission lines: it can be seen that the effect of incorporating the MIRI photometry is similar to the general case with S/N = 5. Note that, on the one hand, the presence of emission lines helps in the redshift identification, but on the other hand having prominent emission lines in contiguous broadbands (Figure 16) makes the observed photometry mimic a less featured SED, resulting in a more difficult redshift recovery.

In the case of S/N = 3 at F150W (Figure 15), assuming identical integration times in all NIRCam filters implies that some galaxies are undetected at the shortest-wavelength NIRCam bands. In this case, having complementary MIRI photometry becomes very important, even with the assumption that we made here, i.e., that the MIRI depth will be 1 mag shallower than the NIRCam depth at 1.5 $\mu$m. This assumption is reasonable, as achieving the same depth as in NIRCam will require too long exposure times.[16]

The result of this poorer-quality photometry is that the output redshift distributions are broader and secondary peaks at lower redshifts become more significant (Figure 15). The incorporation of MIRI photometry has a significant effect in reducing the overall outlier fraction: it goes down from 10.1% with NIRCam data only to 3.0% after adding the two MIRI bands (Table 7). Note that, among the two MIRI filters, F560W appears to have
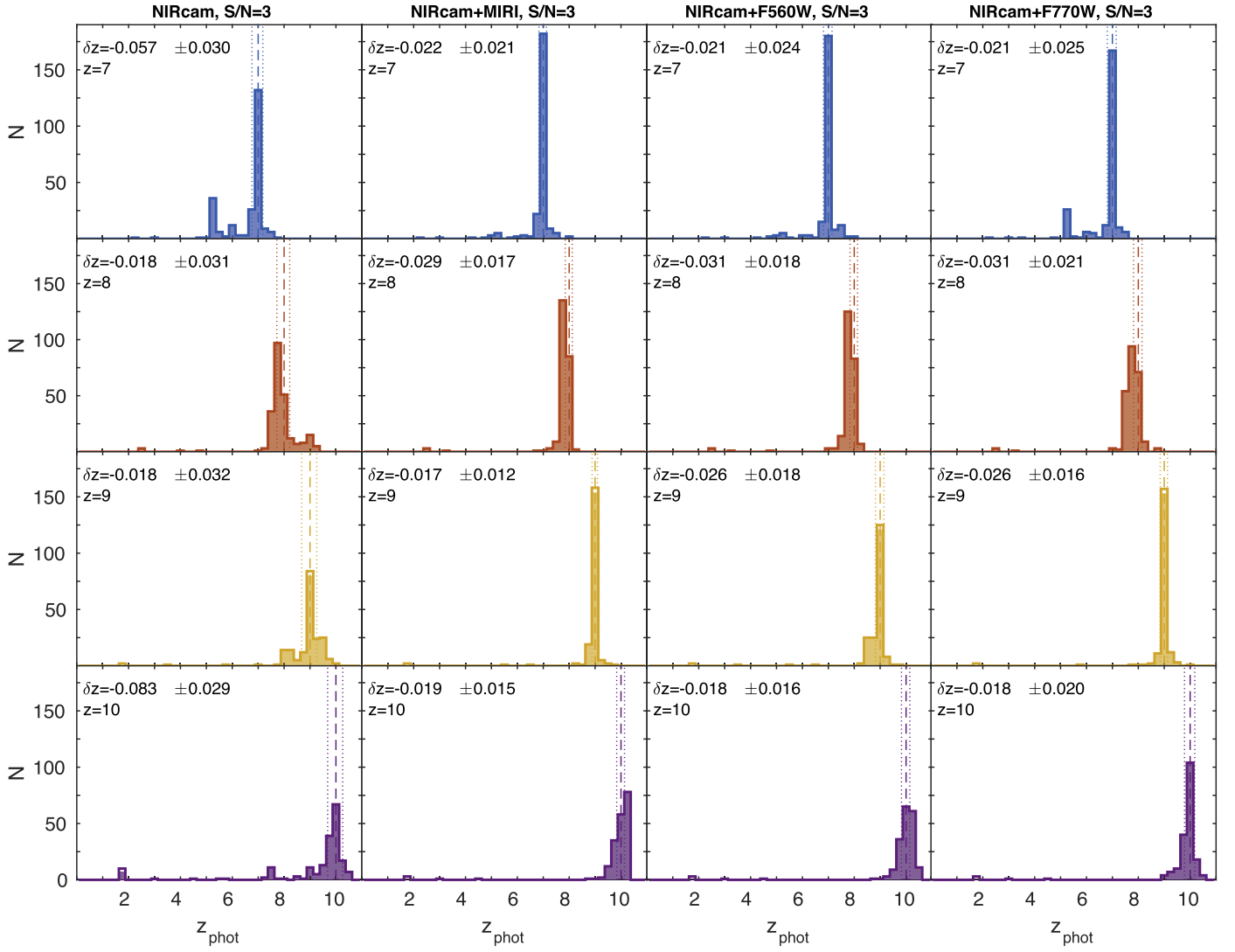
**Figure 15.** Same as Figure 12, but for (F150W) S/N = 3.

a more important effect in improving the $z_{phot}$ determination of the BC03 SED sources at $z = 7$–10.

As a summary, we conclude that considering MIRI photometry of 1 mag shallower depth than the NIRCam photometry keeps the fraction of $z_{phot}$ outliers low for the $z \geqslant 7$ galaxies that have a moderate or low S/N in the NIRCam images. This effect is evident for different SED types (with or without emission lines).

### 3.3.2. Galaxies Simulated with Yggdrasil Templates at z = 7–10

Figures 17–20 show the distributions of the recovered photometric redshifts for the sample 3 galaxies simulated with the Yggdrasil templates at each fixed redshift, obtained by running LePhare with different *JWST* filter combinations. The outlier percentages for each analyzed S/N value are listed in Table 8. Yggdrasil templates are more complex than the BC03 ones, given that Yggdrasil automatically incorporates nebular emission lines and continuum emission when galaxy ages are young and the gas covering factor is $f_{cov} > 0$, which may be the case of many high-$z$ galaxies.

When considering objects with S/N = 10 at F150W (Figure 17), we see that the NIRCam data alone provide a

good photometric redshift recovery ($\sigma = 0.01$–0.02). The incorporation of the MIRI bands has little effect on the resulting $\sigma$ values in most cases, and the fraction of catastrophic outliers is only slightly reduced (see Table 8).

As in the case of BC03 SED galaxies, the impact of incorporating MIRI photometry is more evident at S/N = 5 in F150W (Figure 18). The overall percentage of catastrophic outliers reduces from 20.4% to 12.1% by adding MIRI photometry in this case (Table 8).

If we consider only those galaxies with emission lines in their SEDs (i.e., young and with gas covering factor $f_{cov} = 1$), also at S/N = 5, we see two interesting effects (Figure 19 and Table 8). First, the degeneracies in redshift space are more important than for the general galaxy sample with S/N = 5. This is manifested in a much larger outlier percentage, namely, ~49% when considering the NIRCam data alone. Second, the effect of incorporating MIRI photometry becomes more noteworthy: including both MIRI bands reduces the percentage of outliers to ~17%. These effects are produced by two factors: on the one hand, the nebular line emission boosts the MIRI fluxes and, thus, the S/N of the MIRI data; on the other hand, the multiple emission lines in the Yggdrasil templates with nebular emission result in similarly bright observed fluxes in
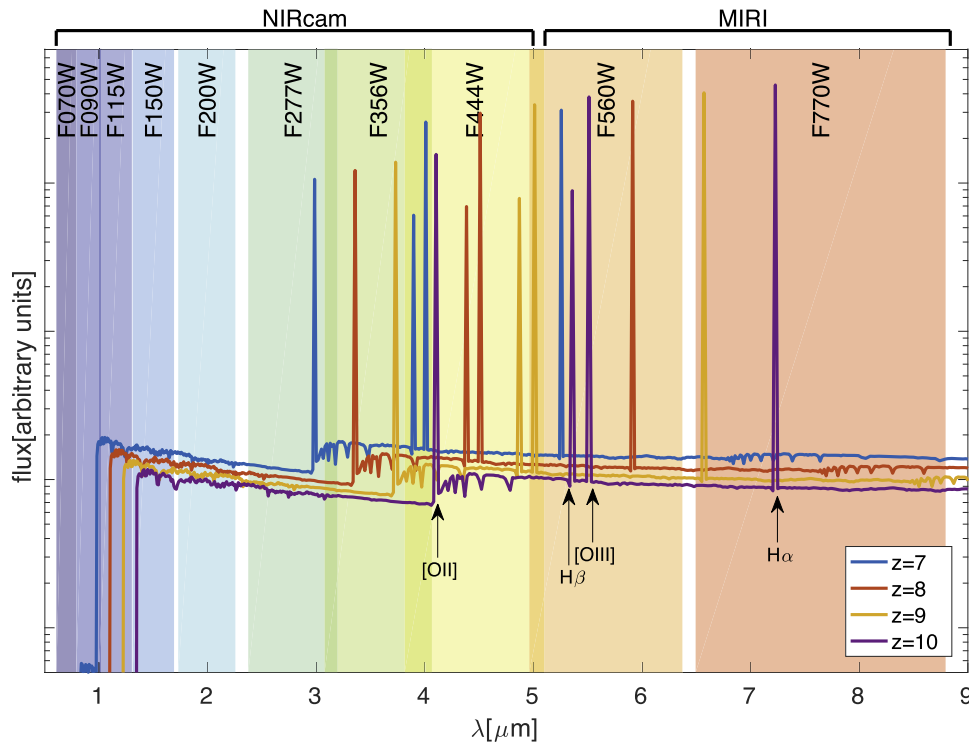
**Figure 16.** Four examples of SED templates from BC03 with manual emission-line addition, considered for simulating galaxies for sample 3. The colored vertical strips indicate the wavelength coverage of the eight NIRCam broadbands and MIRI F560W and F770W. This figure shows in which *JWST* filters the main emission lines and most important SED features are observed at high redshifts.

contiguous filters, masking the continuum SED shape and making the SED fitting more difficult (as explained in Section 3.3.1), so having a longer baseline is necessary for a proper redshift recovery.

For the lowest-quality photometry (S/N = 3; Figure 20), the percentage of catastrophic outliers in the entire sample is 44.4% using only NIRCam photometry. The advantage of incorporating MIRI data, which also have lower S/N, is less obvious in this case, as the percentage of $z_{phot}$ outliers still remains high even with the two MIRI bands (36.7%). A significant fraction of galaxies are detected in only a few (NIRCam and MIRI) bands, so the photometry provides poor constraints for the SED fitting, making degeneracies in redshift space become evident.

Indeed, as is the case for the BC03 SED galaxies, the catastrophic outliers correspond mainly to sources with detections in four or fewer *JWST* bands (open regions in the histograms of Figure 20). However, in the case of the Yggdrasil SED galaxies, the secondary peaks appearing in the output redshift distributions are much more pronounced than for the BC03 galaxies. These secondary peaks are mainly produced by two kinds of templates that are not represented in our BC03 SED sample: (1) young galaxies with nebular emission, where the combination of multiple emission lines and low S/N produces an enhancement in the number of catastrophic outliers; and (2) young galaxies with $f_{cov} = 0$ (i.e., no emission lines) and the highest extinction adopted to build our mock sample (i.e., $A_V = 1$). These galaxies have no analogs either in the BC03-based sample, as we explicitly incorporated the main emission lines in all galaxies with ages lower than the characteristic $\tau$ in the star formation history.

As a summary, we conclude that the benefit of considering MIRI photometry along with the NIRCam data for Yggdrasil SED galaxies is mainly obvious for NIRCam sources with

moderate S/N values, and particularly helpful in the case of SEDs with nebular emission. At low S/N the advantage is less obvious. This is in contrast to our conclusion for the BC03 SED galaxies, indicating that at low S/N the importance of having MIRI data depends significantly on the galaxy spectral type.

### 3.3.3. MIRI Magnitudes and NIRCam/MIRI Colors

In this section we show the MIRI magnitudes and NIRCam/MIRI colors of our simulated galaxies in sample 3, as this can be a useful reference when planning *JWST* observations. In particular, we analyze how the MIRI detections vary with the galaxy SED type. Given our chosen normalization [F150W] = 29 mag, all these galaxies will easily be observed in NIRCam deep surveys (as it takes only ~30 minutes of integration per pointing to reach [F150W] = 29 mag, 3σ with NIRCam, according to the public *JWST*/NIRCam exposure time calculator[17]). So our aim here is to discuss the fractions of such galaxies that will be detected in typically deep MIRI observations. Of course, these same results only need to be rescaled to remain valid for other NIRCam magnitudes.

Figures 21–24 show the expected MIRI magnitude distributions for our sample 3 galaxies generated with the BC03 and Yggdrasil models. As the NIRCam F150W magnitude for all these galaxies has been fixed to [F150W] = 29 mag, there is a direct correspondence between the MIRI F560W magnitudes and the F150W–F560W colors, which are also shown in each plot. For the BC03 galaxies (Figures 21 and 22), we analyzed separately the cases with and without emission lines, which here correspond to galaxies with ages lower and higher than their characteristic SFH decay parameter $\tau$, respectively. For

---
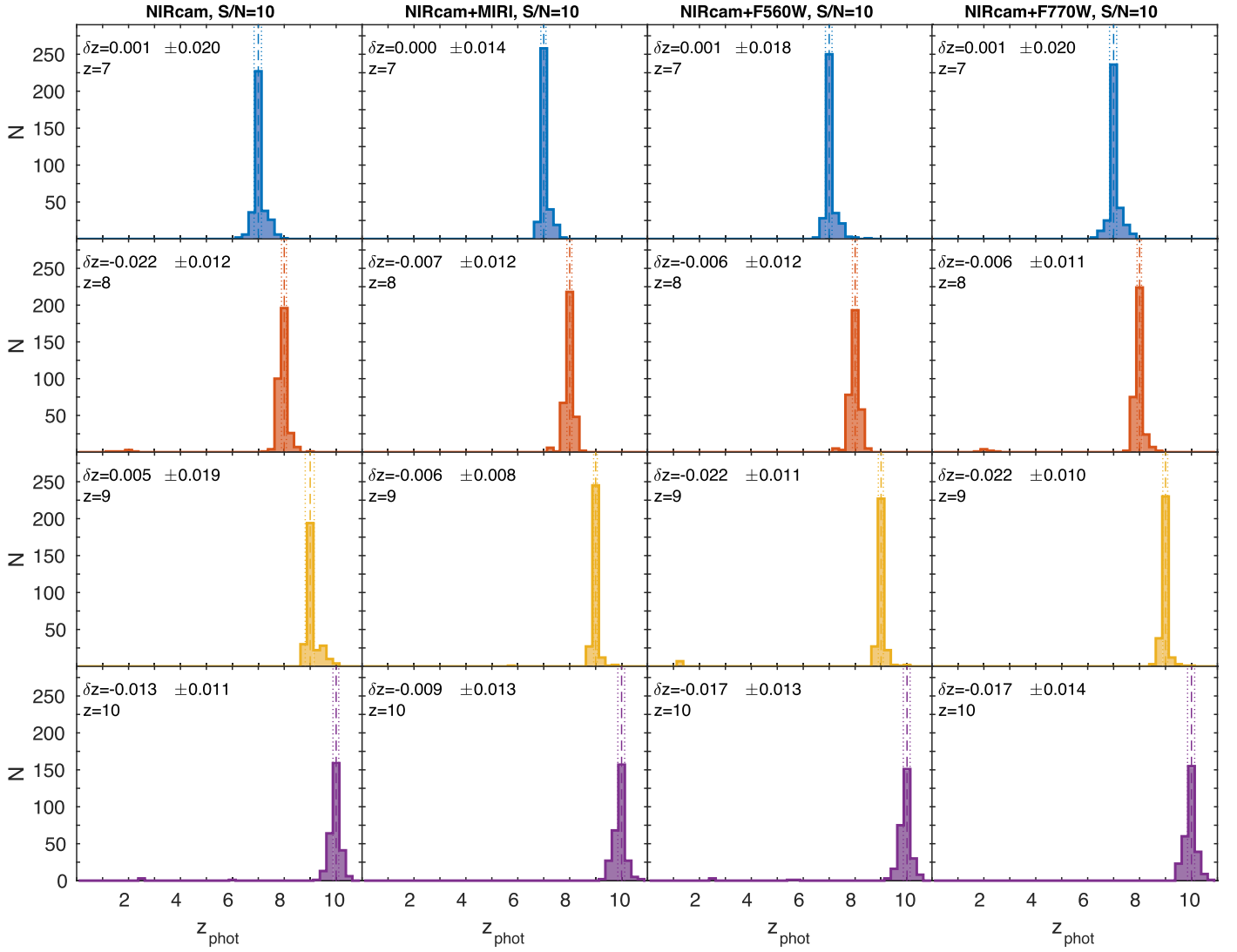
[17] http://jwstetc.stsci.edu/etc/input/nircam/imaging/

**Figure 17.** Photometric redshifts for the sample of Yggdrasil simulated galaxies with S/N = 10. From top to bottom: redshifts $z = 7$–10. Photometric redshifts in each column are obtained with different band combinations. From left to right: eight NIRCam broadbands; eight NIRCam broadbands and MIRI F560W, F770W; eight NIRCam broadbands and MIRI F560W only; eight NIRCam broadbands and MIRI F770W only. Each row corresponds to one of the four specific input redshifts. The vertical lines indicate the $3\sigma$ interval around the mean normalized redshift difference.

the Yggdrasil galaxies (Figures 23 and 24), we analyzed three cases: galaxies with ages larger that their constant star formation period, which are galaxies in passive evolution; and galaxies that are still forming stars with two different gas covering factors, $f_{cov} = 0$ (implying no line or nebular continuum emission) and 1 (with nebular emission). The median values of all distributions are provided in Tables 9 and 10.

The magnitude distributions in Figure 21 show that the vast majority of BC03 galaxies with [F150W] = 29 mag at $z = 7$, 8, and 10, and more than half of those at $z = 9$, will be detected by an MIRI survey with a detection limit [F560W] = 28 mag. This result is valid for galaxies with or without emission lines. All the latter are cases in which a 4000 Å break is at least partly developed, so the SEDs are brighter at the MIRI wavelengths than at 1.5 $\mu$m, particularly at $z = 7$ (see Figure 4). This $k$-correction effect causes the median [F560W] value at this redshift to be brighter for galaxies without emission lines, as those with emission lines correspond to younger galaxies with fainter continua that partly cancel the emission-line flux enhancement

(Table 9). At $z = 8$ and 10, instead, the combination of the continuum $k$-correction and the nebular emission results in a net flux boosting in the case of emission-line galaxies.

For F770W (Figure 22), the flux enhancement due to the emission-line presence is only evident at $z = 9$ and 10, as the H$\alpha$ emission line enters the F770W wavelength window at these redshifts (Table 10). Most young galaxies at $z = 7$ and 8, instead, remain below our reference [F770W] = 28 mag detection limit, as their continua are blue and no emission line is present in this filter to help with the detection. So, the F770W filter will only allow us to detect evolved galaxies at $z \geqslant 7$ or young line emitters at $z = 9$ and 10 to an [F770W] = 28 mag cut.

The vast majority of the Yggdrasil SED galaxies with [F150W] = 29 mag that finished forming stars at $z \geqslant 7$ will have mag <28 in both MIRI bands (Figures 23 and 24, and Tables 9 and 10). This is the result of the same $k$-correction effect discussed above.

For the Yggdrasil star-forming galaxies, there is a clear enhancement of the MIRI fluxes produced by the nebular lines
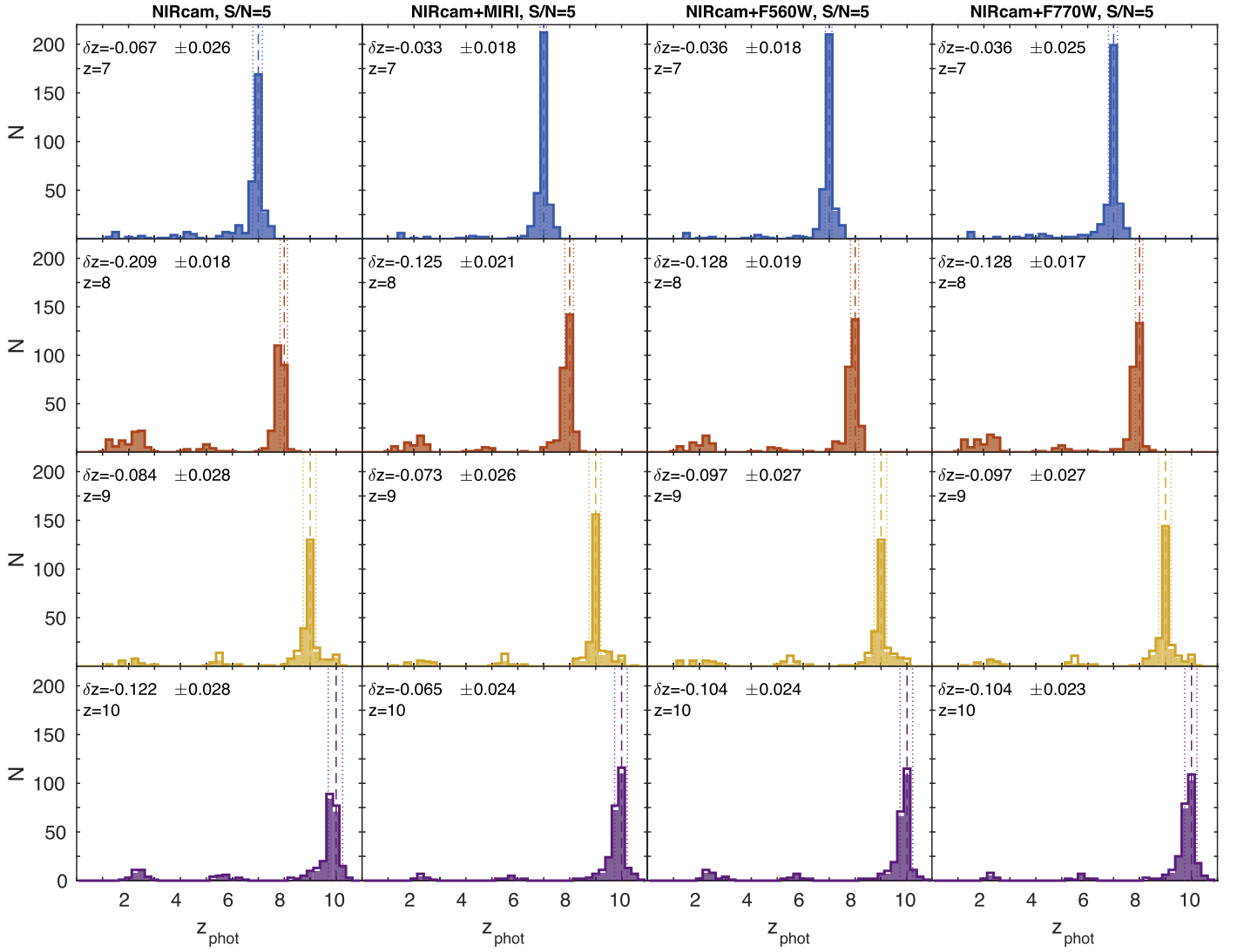
**Figure 18.** Same as Figure 17, but for S/N = 5. The shaded area within each histogram corresponds to objects with more than four detections in total in the NIRCam and MIRI bands, while open areas correspond to sources with less than four band detections.

**Table 8**
Number of Outliers among Our 1260 Sample 3 Galaxies Simulated with Yggdrasil Models for Different *JWST* Filter Combinations and Assumed S/N Values at F150W

| Bands | S/N = 10 | S/N = 5 | S/N = 5 (Only EL)[a] | S/N = 3 |
|---|---|---|---|---|
| Eight NIRCam bands | 11 (0.9%) | 257 (20.4%) | 118[b] (49.2%) | 560 (44.4%) |
| Eight NIRCam bands + MIRI F560W, F770W | 1 (0.1%) | 152 (12.1%) | 41[b] (17.1%) | 462 (36.7%) |
| Eight NIRCam bands + MIRI F560W | 13 (1.0%) | 188 (14.9%) | 68[b] (28.3%) | 490 (38.9%) |
| Eight NIRCam bands + MIRI F770W | 8 (0.6%) | 196 (15.6%) | 79[b] (32.9%) | 508 (40.3%) |

**Notes.** Outliers are defined as galaxies with $|z_{phot} - z_{fiduc.}|/(1 + z_{fiduc.}) > 0.15$.
[a] Only templates with emission lines (EL).
[b] Out of a total of 240 galaxies with emission lines. The percentages refer to this denominator.

(compare the cases with $f_{cov} = 0$ and 1 in Figures 23 and 24 and the corresponding tables). In this case, the flux enhacement is more obvious than for the BC03 templates because here we are comparing only young galaxies with and without lines (segregated by their different gas covering factors), so all their continua are similarly blue. In both cases, however, the median MIRI magnitude values are >28 mag. This implies that an MIRI galaxy survey to our reference depth will only detect a minority of star-forming galaxies with [F150W] = 29 mag at $z \geqslant 7$ if their SEDs are similar to the Yggdrasil templates. Note
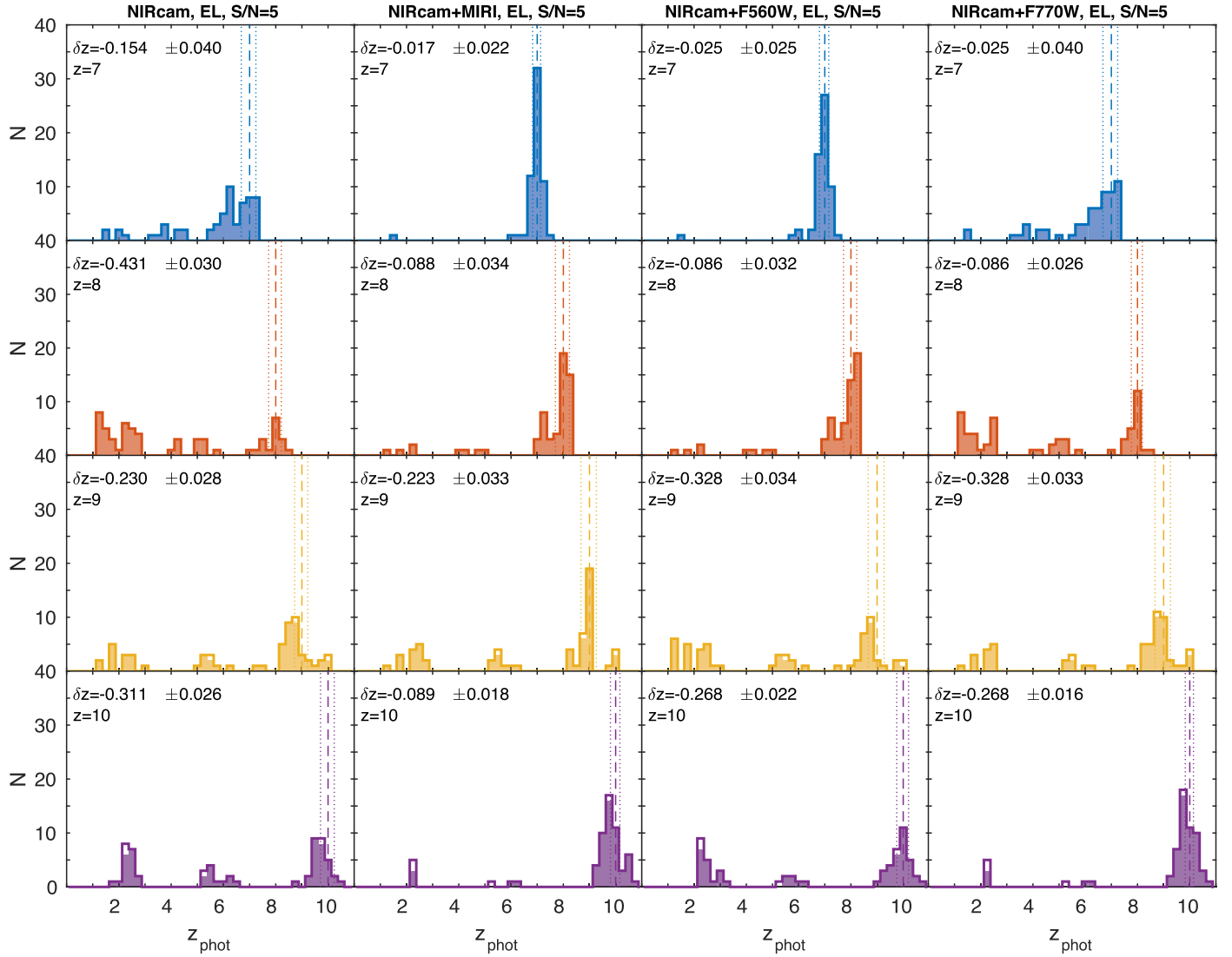
**Figure 19.** Same as Figure 18, but only for galaxies with emission lines.

that these galaxies are considerably brighter in F560W if they are at $z = 7$ or 8, while slightly brighter in F770W if at $z = 9$ or 10.

Note that, conversely, our color analysis shows that detecting more than one-half of red MIRI galaxies at $z \geqslant 7$ will require NIRCam data at least 3 mag deeper. This is the case both for "passive galaxies" (i.e., galaxies that finished forming stars) and for galaxies that are red simply because of the presence of prominent emission lines.

So, as a summary, "old" galaxies at $z \geqslant 7$ will easily be detected with MIRI in reasonably deep surveys, independently of their galaxy type. For young galaxies, the situation is more complex, as it depends on to what extent line emission can compensate for the very faint continua that are characteristic in these sources. F560W will provide more detections than F770W if maps of the same depth are obtained in both passbands.

## 4. SUMMARY AND CONCLUSIONS

In this work we have tested the impact of having data in different *JWST* filter combinations on deriving photometric

redshifts for galaxies at different redshifts. We considered the eight NIRCam broadbands and the two shortest-wavelength MIRI bands (F560W and F770W), which are the most sensitive ones and, thus, those that will be preferred for the study of high-$z$ galaxies. In addition, we also investigated the effect of having (or not) ancillary photometry from *HST* and ground-based telescopes at wavelengths shorter than those to be observed with *JWST* (i.e., $\lambda < 0.6 \ \mu$m).

We performed our tests on three galaxy samples with known input redshifts: a sample of 2422 galaxies with spectroscopic redshifts $z = 0$–6 (sample 1), a sample of 1375 galaxies with CANDELS consensus photometric redshifts at $z = 4$–7 (sample 2), and a sample of 2124 mock galaxies at $z = 7$–10 (sample 3), whose SEDs have been simulated with two template libraries, namely, BC03 with manual addition of the main emission lines for star-forming galaxies (considered as those with age lower than the characteristic $\tau$ of the star formation history), and the Yggdrasil library. Besides, for sample 3 we also explicitly investigated the effect of having photometry with different S/N values on the ability to recover the galaxy redshifts. In samples 1 and 2, more than 90% of
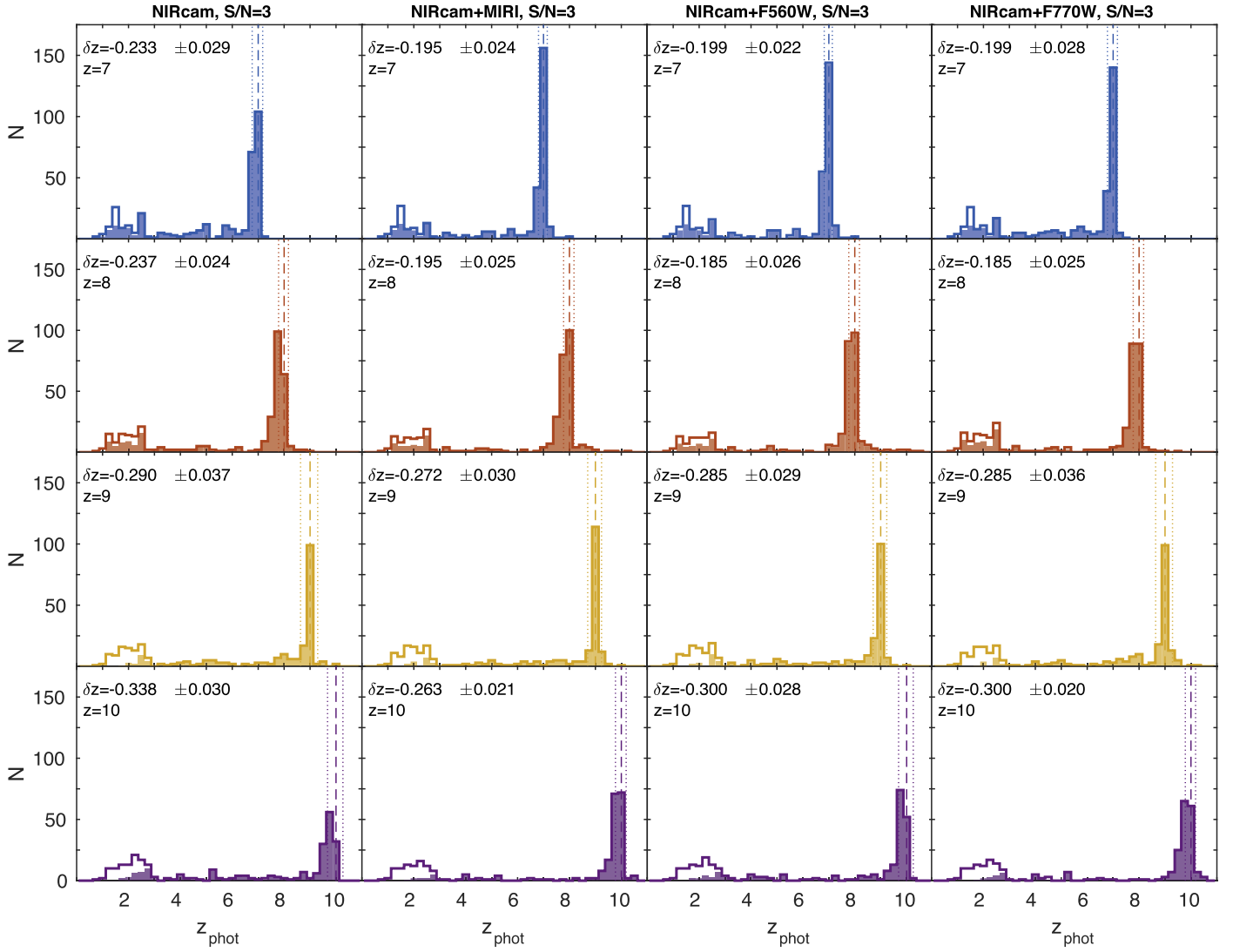
**Figure 20.** Same as Figure 17, but for S/N = 3. The shaded area within each histogram corresponds to objects with more than four detections in total in the NIRCam and MIRI bands, while open areas correspond to sources with less than four band detections.

galaxies have F150W photometry with S/N > 5, so the test results based on these samples are not affected by the photometric quality.

These three samples altogether have allowed us to assess different effects in the photometric redshift estimation, such as the presence of high-$z$ contaminants and the leakage of high-$z$ sources toward low $z$. Note that our three samples are not meant to emulate a real galaxy population selected from a *JWST* NIRCam or MIRI blank field, but rather sample the typical redshifts and SED types of the galaxies that will be observed in these surveys, in order to test potential problems in the $z_{\rm phot}$ derivation.

Our main results are as follows:

1. The NIRCam broadbands alone are not sufficient to obtain good-quality photometric redshift estimates at low and intermediate redshifts ($z < 7$). For sample 1, which is optimized to study the redshift range $z = 0$–4, the use of NIRCam data alone results in a high photometric redshift dispersion and outlier percentage (>10%). In sample 2, which is particularly suited to study the leakage of $z = 4$–7 sources to low $z$ due to redshift failure, we found

that ~20% (~9%) of the input sources with $z = 4$–5 ($z = 5$–7) are misidentified as lower-$z$ galaxies. These effects are smaller at $z > 6$, as the Lyman break is shifted into the NIRCam bands.

2. Having photometry in the *HST* F435W and F606W bands, and if possible also in the ground-based $U$ band, is very important for keeping the fraction of low-$z$ contaminants under control. This photometry is important for constraining the Lyman break shift and, thus, the redshifts of low-$z$ sources. Indeed, in sample 1 the percentage of low-$z$ outliers changes from >10% to ~7%, and the normalized redshift difference distribution rms reduces to 40%. In sample 2, the percentage of sources leaking from $z = 4$–5 ($z = 5$–7) to lower $z$ reduces from ~20% (~9%) to ~2% (~8%). At $z = 5$–7, the reduction in the percentage of leaking sources is exclusively due to the incorporation of the *HST* bands. The additional incorporation of $U$-band data has no effect because the Lyman break shifts inside the *HST* wavelength range at these redshifts.

3. MIRI F560W and F770W data can help mitigate the absence of photometry at $\lambda < 0.6\ \mu$m. In sample 1,
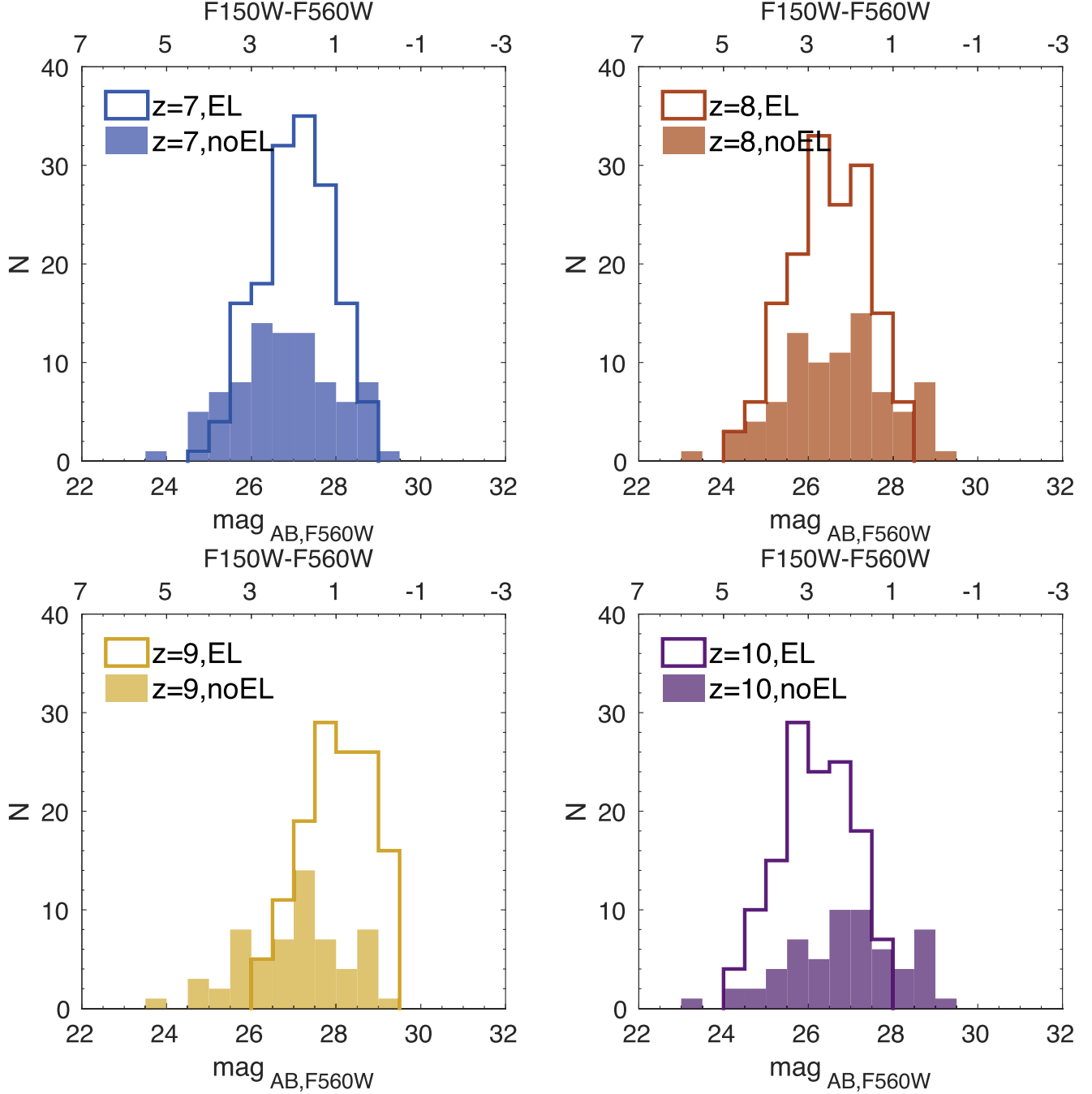
**Figure 21.** Expected MIRI F560W magnitude distribution for the sample 3 galaxies simulated with the BC03 models at different redshifts. The cases of galaxies with emission lines (i.e., star-forming galaxies with ages lower than their characteristic SFH $\tau$) and without emission lines are shown in separate histograms. As the NIRCam F150W magnitude for all these galaxies has been fixed to [F150W] = 29 mag, there is a direct correspondence between the MIRI F560W magnitudes and the F150W–F560W colors (top $x$-axis).

we found that the percentage of outliers reduces from ∼10% to <9%, by complementing the NIRCam data only with MIRI data in the photometric input catalog. For the leaking sources in sample 2, the addition of MIRI photometry to the NIRCam data reduces the percentage of outliers from ∼20% (∼9%) to <8% (<4%) at the same redshifts. When both the *HST* photometry and MIRI photometry are

considered along with the NIRCam photometry, the percentage of leaking sources from $z = 4$–5 goes down to ∼1%.

4. Among the different photometric configurations tested here, the MIRI F560W and F770W bands are the only ones that can meaningfully complement the NIRCam data in galaxy studies at $z \geqslant 7$. The importance of the *HST* data at $\lambda < 0.6\,\mu$m becomes negligible once the
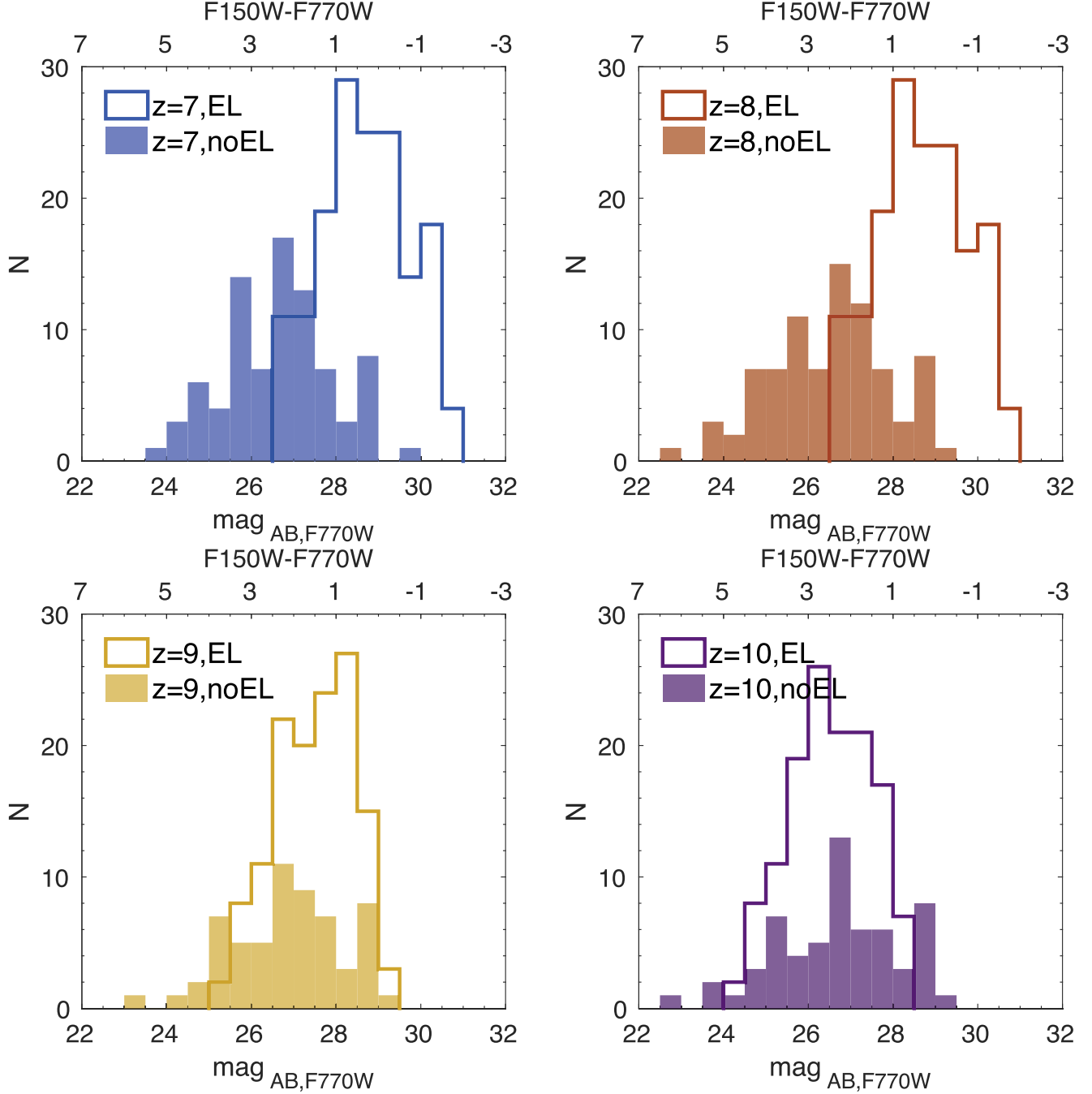
## BC03 templates



**Figure 22.** Same as Figure 21, but for the MIRI F770W magnitudes and F150W–F770W colors.

Lyman break shifts within the NIRCam bands, except to confirm the expected nondetections.

5. For photometry with F150W S/N = 10, the NIRCam data alone allow for accurate redshift estimates for galaxies at $z \geqslant 7$ in most cases. We found that the rms of the $|z_{\mathrm{phot}} - z_{\mathrm{input}}|/(1 + z_{\mathrm{input}})$ distributions for sample 3 range between $\sigma < 0.01$ and 0.02 with NIRCam data alone at all high $z$. The addition of MIRI photometry has a mild effect, being useful to enhance the accuracy of the derived $z_{\mathrm{phot}}$.

6. For sources with F150W S/N = 5, the benefit of incorporating MIRI photometry becomes more evident. In our reference case, in which the MIRI depth is 1 mag brighter than the NIRCam F150W depth, the incorporation of MIRI data significantly reduces the fraction of outliers with respect to the case with NIRCam data alone. This improvement is particularly noteworthy in the case of Yggdrasil SED galaxies with emission lines, as the multiple nebular lines coupled with a moderate-quality S/N make the SED fitting more challenging. The MIRI
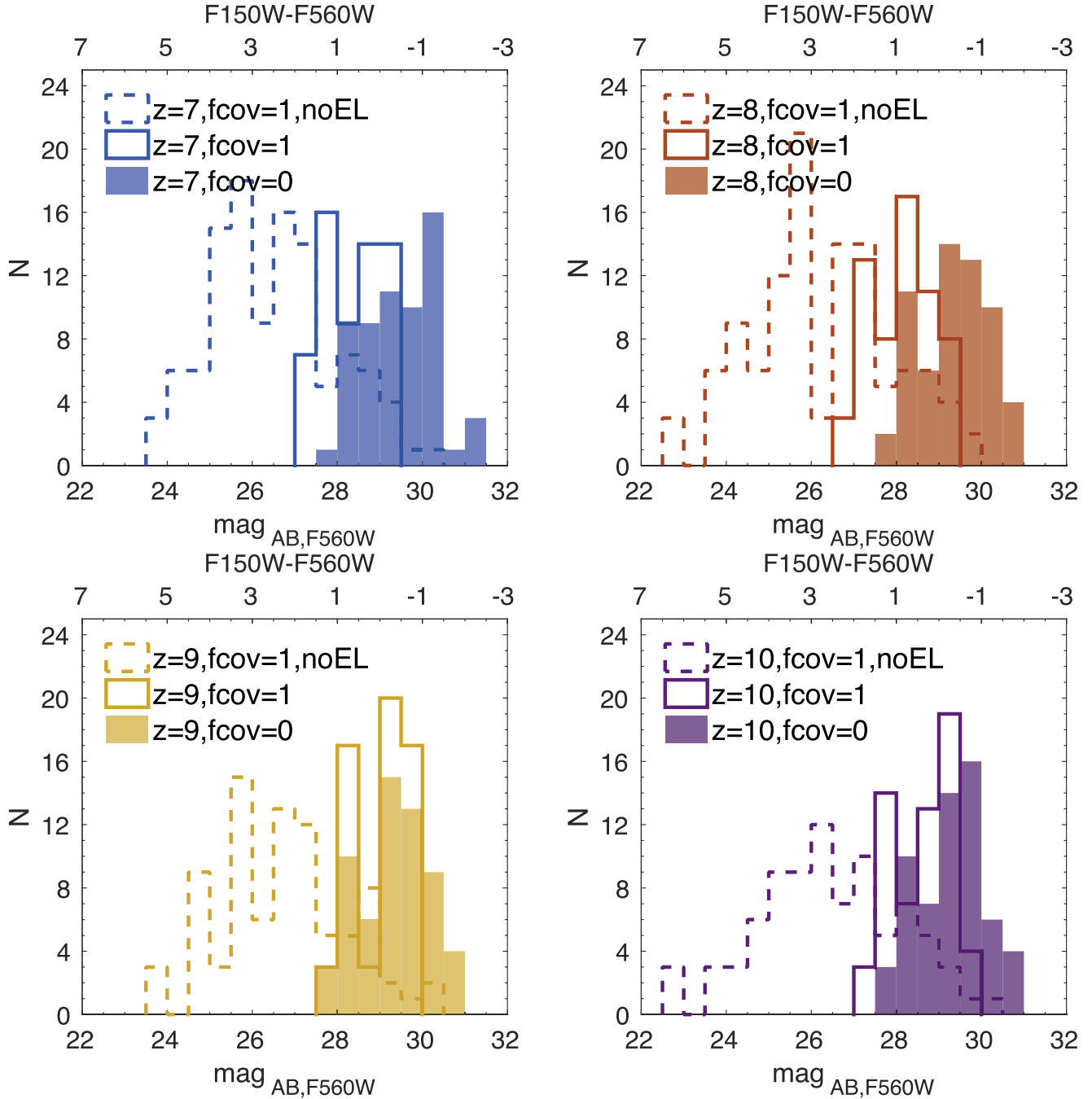
## Yggdrasil templates



**Figure 23.** Expected MIRI F560W magnitude distribution for the sample 3 galaxies simulated with the Yggdrasil models for galaxies in passive evolution (i.e., ages larger than their constant star formation period; dashed line) and galaxies that are still forming stars. The star-forming galaxies are also divided into models with covering factors $f_{cov} = 0$ (filled area) and 1 (solid line), as only the latter have emission lines and continuum nebular emission. Note that the gas covering factor is irrelevant in the cases of galaxies in passive evolution, as no nebular emission is present any more in these galaxies. The top x-axis shows the F150W–F560W colors.

bands help extend the wavelength baseline for the SED fitting, and thus using NIRCam and MIRI data altogether results in a clear redshift recovery improvement.

7. At F150W S/N = 3, the role of MIRI data in improving the redshift determination depends very much on the SED type. As "old" galaxies are red, MIRI fluxes are brighter that NIRCam fluxes, so the incorporation of the MIRI data is of clear benefit for these sources. For star-forming galaxies, the presence of emission lines in the MIRI

bands boosts the MIRI fluxes, but at the same time these galaxies have bluer continua, so the net effect is more variable in these cases.

8. At fixed S/N, the presence of multiple emission lines, like in the case of nebular emission, makes the correct redshift identification much more difficult. Although the emission lines enhance the fluxes, the presence of emission lines in all passbands makes the filter-convolved SEDs rather featureless, and the SED fitting becomes more complicated.

**Yggdrasil templates**
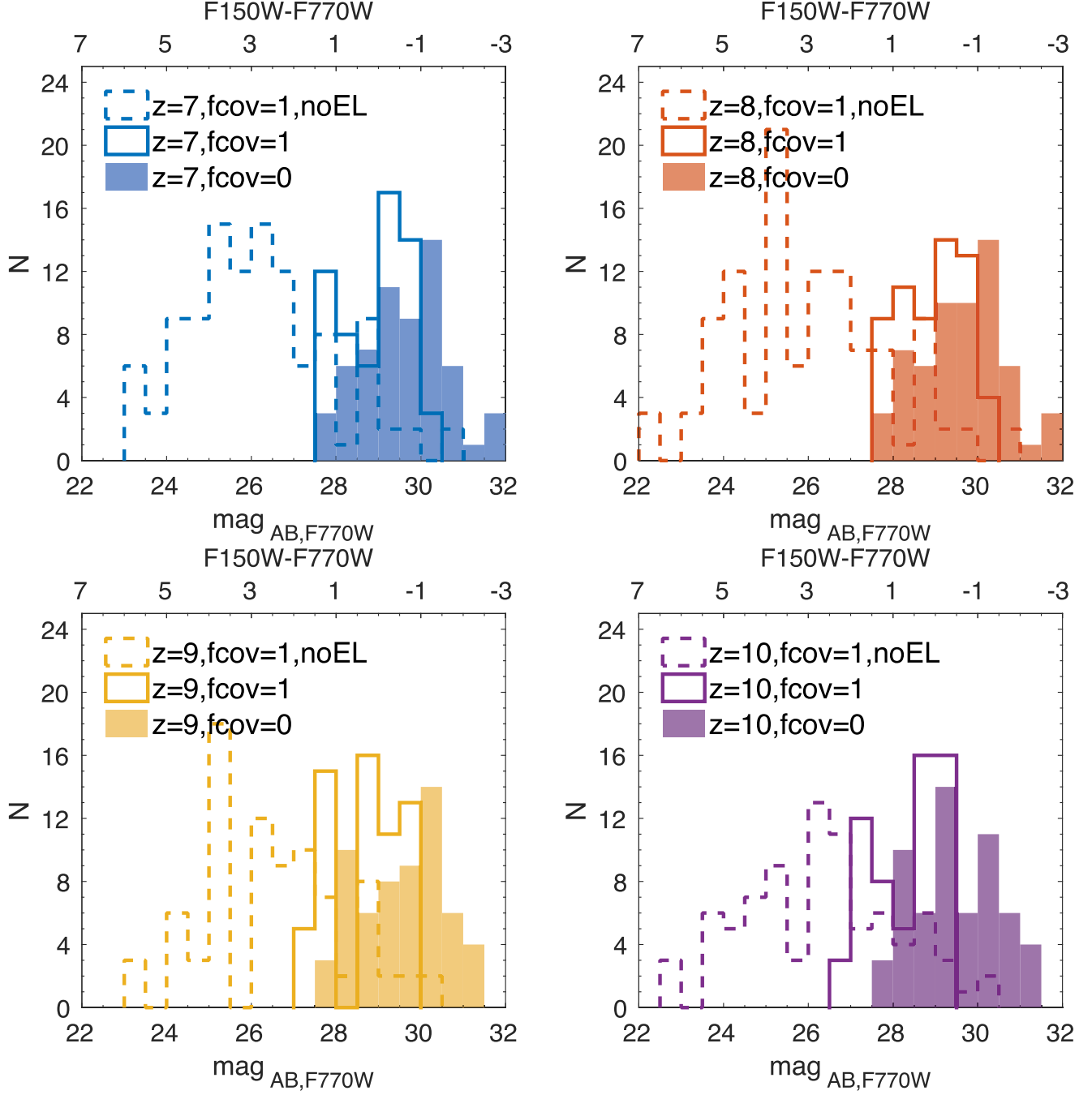


**Figure 24.** Same as Figure 23, but for the MIRI F770W magnitudes and F150W–F770W colors.

<table>
<tr><td colspan="6" align="center"><b>Table 9</b><br>Median MIRI [F560W] Magnitudes for Our Sample 3 Galaxies at $z = 7$–10<br>(All Normalized at [F150W] = 29 mag)</td></tr>
</table>

| Redshift | BC03 | | Yggdrasil Passive Evolution | Yggdrasil Star Formation | |
| --- | --- | --- | --- | --- | --- |
| | No Emiss. Lines | With Emiss. Lines | | $f_{cov} = 0$ | $f_{cov} = 1$ |
| 7 | 26.86 | 27.08 | 26.28 | 29.50 | 28.47 |
| 8 | 26.77 | 26.49 | 25.96 | 29.43 | 28.32 |
| 9 | 27.12 | 28.08 | 26.66 | 29.35 | 29.15 |
| 10 | 26.96 | 26.17 | 26.44 | 29.27 | 28.72 |

<table>
<tr><td colspan="6" align="center"><b>Table 10</b><br>Median MIRI [F770W] Magnitudes for Our Sample 3 Galaxies at $z = 7$–10<br>(All Normalized at [F150W] = 29 mag)</td></tr>
</table>

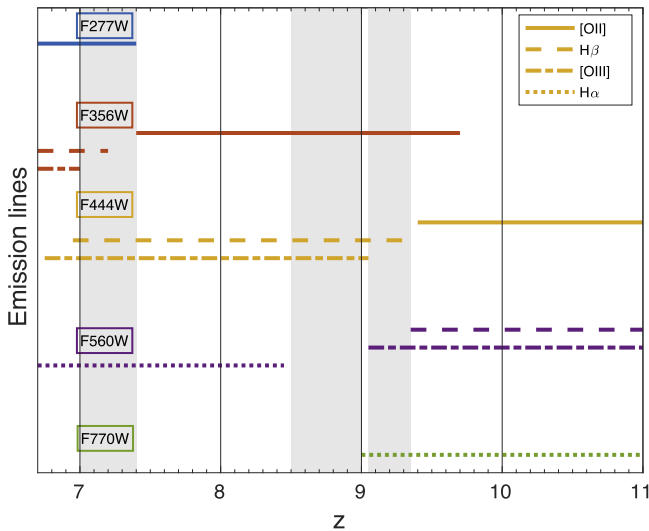| Redshift | BC03 | | Yggdrasil Passive Evolution | Yggdrasil Star Formation | |
| --- | --- | --- | --- | --- | --- |
| | No Emiss. Lines | With Emiss. Lines | | $f_{cov} = 0$ | $f_{cov} = 1$ |
| 7 | 26.70 | 28.64 | 26.10 | 29.64 | 29.05 |
| 8 | 26.59 | 28.68 | 25.80 | 29.62 | 29.13 |
| 9 | 26.98 | 27.52 | 26.44 | 29.58 | 28.87 |
| 10 | 26.79 | 26.51 | 26.24 | 29.47 | 28.53 |

**Figure 25.** Emission lines present inside NIRCam and MIRI bands at high redshifts. Different colors indicate different bands. Blue: F277W; red: F356W; yellow: F444W; purple: F5602; green: F770W. The shortest-wavelength NIRCam bands are not shown because they do not contain emission lines in this redshift range. Different line styles represent different emission lines: [O II] (solid line); Hβ (dashed line); [O III] (dot-dashed line); and Hα (dotted line). Vertical black lines indicate the redshifts considered in sample 3 ($z = 7$–10). These specific redshifts are representative of all $z = 7$–10 redshifts, except those in the gray areas, which are instead analyzed in this appendix.

In addition, we analyzed the MIRI magnitudes of our sample 3 simulated galaxies at $z \geqslant 7$. As we have normalized all the SEDs to [F150W] = 29 mag, the F150W–MIRI colors directly correspond to MIRI magnitude values. Analyzing this expected MIRI photometry has allowed us to assess what types of high-$z$ galaxies will likely be detected in MIRI deep surveys. For a different F150W normalization, one simply has to scale these results, as the galaxy colors are fixed for each SED model.

We found that an MIRI survey with a detection limit of [F560W] = 28 mag should enable the detection of the vast majority of the NIRCam sources with [F150W] = 29 mag at $z = 7$–10, provided that these sources are at least mildly evolved. The youngest galaxies that are still forming stars (Yggdrasil SED galaxies with ages <100 Myr) will remain mostly undetected. Although in the case of gas covering factor $f_{cov} = 1$ the nebular lines make the MIRI fluxes much brighter than in the case of $f_{cov} = 0$, the continua of these young galaxies are very blue, so the resulting MIRI fluxes are not very high even in the case of nebular emission.

At the same MIRI reference depth, the F770W filter will provide fewer detections than the F560W filter at $z \geqslant 7$. Among the NIRCam sources with [F150W] = 29 mag, evolved galaxies at $z \geqslant 7$ and some line emitters at $z = 9$ and 10 will be detected in F770W, but very few line emitters at $z = 7$ and 8.

Conversely, our color analysis shows that detecting the majority of red MIRI galaxies at $z \geqslant 7$ will require NIRCam data at least 3 mag deeper. This is the case both for "passive galaxies" (i.e., galaxies that finished forming stars) and for galaxies that are red simply because of the presence of main emission lines.

As an overall summary, we argue that the results presented here constitute a useful reference for designing deep imaging surveys with *JWST*. We conclude that NIRCam and MIRI will have complementary roles and the optimal observing strategy

and filter combination to be adopted will depend on each observing program's science goals. As adding MIRI imaging to the NIRCam observations would significantly enlarge the requested observing times, observers must clearly assess the need and importance of the former for their specific studies.

MIRI observations with both the F560W and F770W filters are necessary along with the NIRCam data if a correct identification of low- and moderate-redshift ($z < 7$) sources is desired in fields without $\lambda < 0.6\ \mu$m coverage. Alternatively, a less expensive strategy would be limiting galaxy surveys to fields with ancillary $\lambda < 0.6\ \mu$m data of matching depth and obtaining only NIRCam data, but the lack of any MIRI imaging would still result in a slightly higher percentage of redshift outliers and a lower overall recovered-redshift precision.

If the main science goal is simply identifying the bulk of the galaxy population at $z = 7$–10, proposing an observing program based on ultradeep NIRCam data alone will be adequate. However, adding MIRI imaging will help improve the identification of red sources and is of utmost importance to recover stellar ages and stellar masses at $z > 7$ (L. Bisigello et al. 2016, in preparation). Therefore, observing with both cameras will be necessary for a comprehensive study of galaxy evolution since early cosmic times.

### APPENDIX A
### SAMPLE 3: ANALYSIS OF ADDITIONAL REDSHIFTS

In Section 2.1.3 we considered four redshifts for our sample 3 analysis: $z = 7$–10. These four redshifts are representative of most of the entire redshift range $z = 7$–10, except for a few exceptions that we study here. For this analysis, we considered only BC03 templates, to which we manually added the main emission lines. We analyzed how these emission lines move in and out of NIRCam and MIRI filters to identify all possible configurations (Figure 25). For each filter, we only took into account the wavelength range where its transmission is above 20%. To cover the redshift ranges that are unrepresented by our
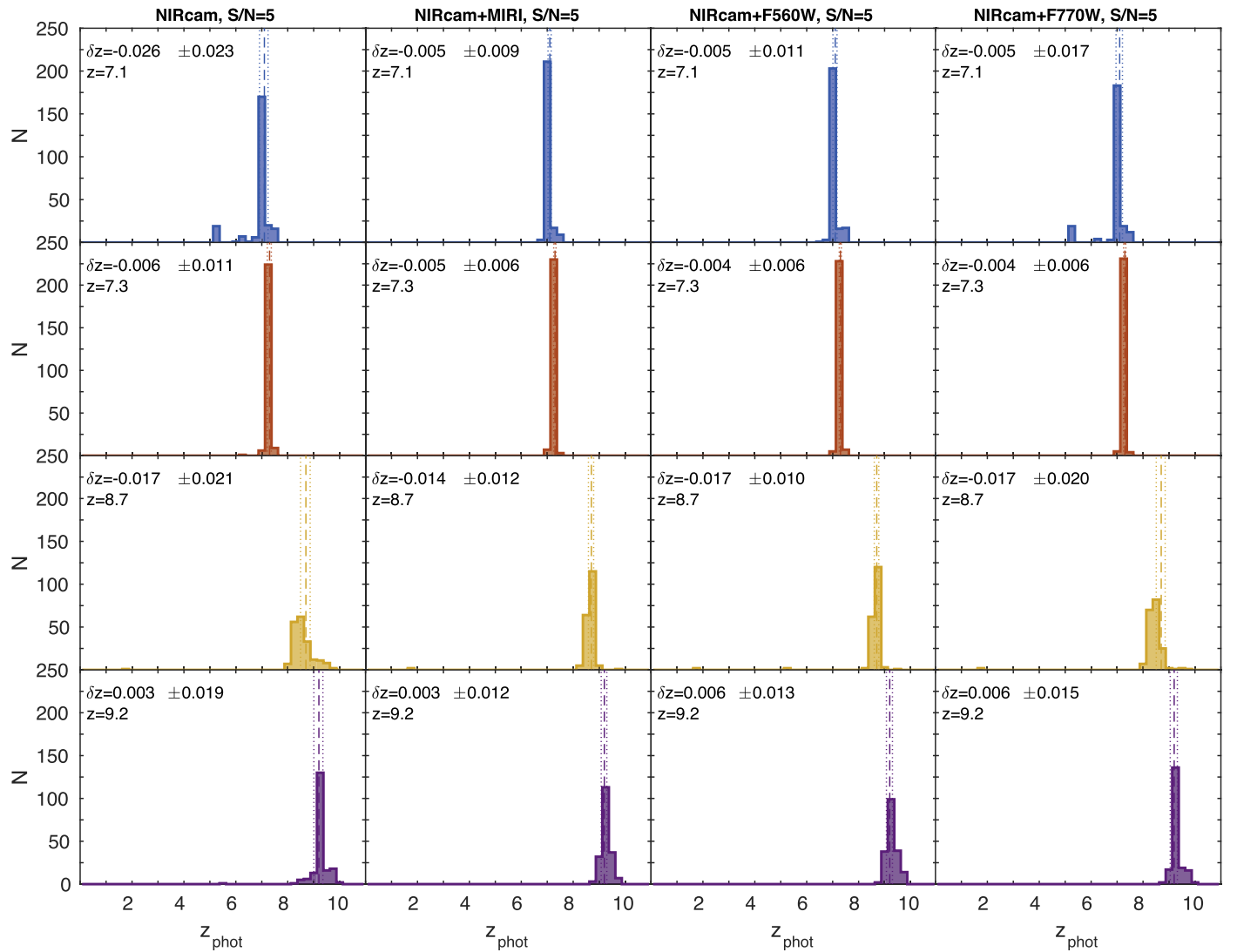
**Figure 26.** Photometric redshifts obtained for the BC03 simulated galaxies with (F150W) S/N = 5 at different fixed redshifts. From left to right: input redshifts $z = 7.1, 7.3, 8.7,$ and 9.2. Photometric redshifts in each column are obtained with different combinations of bands. From left to right: eight NIRCam broadbands; eight NIRCam broadbands, MIRI F560W, and MIRI F770W; eight NIRCam broadbands and MIRI F560W only; eight NIRCam broadbands and MIRI F770W only. The vertical lines indicate the $3\sigma$ interval around the mean normalized redshift difference. On the top left of each panel we quote the mean of $(z_{\rm phot} - z_{\rm input})/(1 + z_{\rm input})$ and the rms of $|z_{\rm phot} - z_{\rm input}|/(1 + z_{\rm input})$.

original analysis (corresponding to the shaded areas in Figure 25), we retested the sample 3 redshift recovery at four additional redshifts, namely, $z = 7.1, 7.3, 8.7,$ and 9.2.

The results of our new tests are shown in Figure 26 (for S/N = 5). At $z = 7.1$, the main difference with respect to the $z = 7$ configuration is that the [O III] line moves out of the F356W NIRCam filter while H$\beta$ is still inside. The results when using NIRCam broadbands alone are very similar to the $z = 7$ case, with similar rms values, 0.023 in this case (compared with 0.025 at $z = 7$), and similar distribution shape. The number of outliers decreases from ~8% to 0 when adding the F560W band, but it does not change when adding the F770W band. Also, the rms improves more when adding the F560W filter rather than the F770W one.

At $z = 7.3$ the main difference with respect to the $z = 7$ configuration and the previous one is that there are no emission lines inside the F356W band. When using NIRCam bands alone, it is already possible to obtain a good redshift estimation with no outliers and a small rms. This is because the absence of

emission lines inside the F356W band creates a strong contrast with the closest bands, allowing the code to properly identify the main SED features. Moreover, this contrast happens in a narrow redshift range, confining the range of possible redshift solutions and producing a very small rms value. This is the most favorable configuration for the NIRCam bands among all considered redshifts. By adding the MIRI bands, there is only a slight improvement to the already good redshift estimation.

The case of $z = 8.7$ is representative of the redshift range where no emission lines at all are present in the MIRI bands, while [O III], H$\beta$, and [O II] are inside the NIRCam wavelength range. The main difference between this case and the $z = 8$ one is that the flux observed in the F115W band, where the Lyman break falls, is fainter and more often is below the $2\sigma$ detection limit. When using the NIRCam bands only, the derived distribution is broad, but there are almost no outliers (0.5%). When adding the MIRI bands, and in particular the F560W band, the rms becomes about one-half.
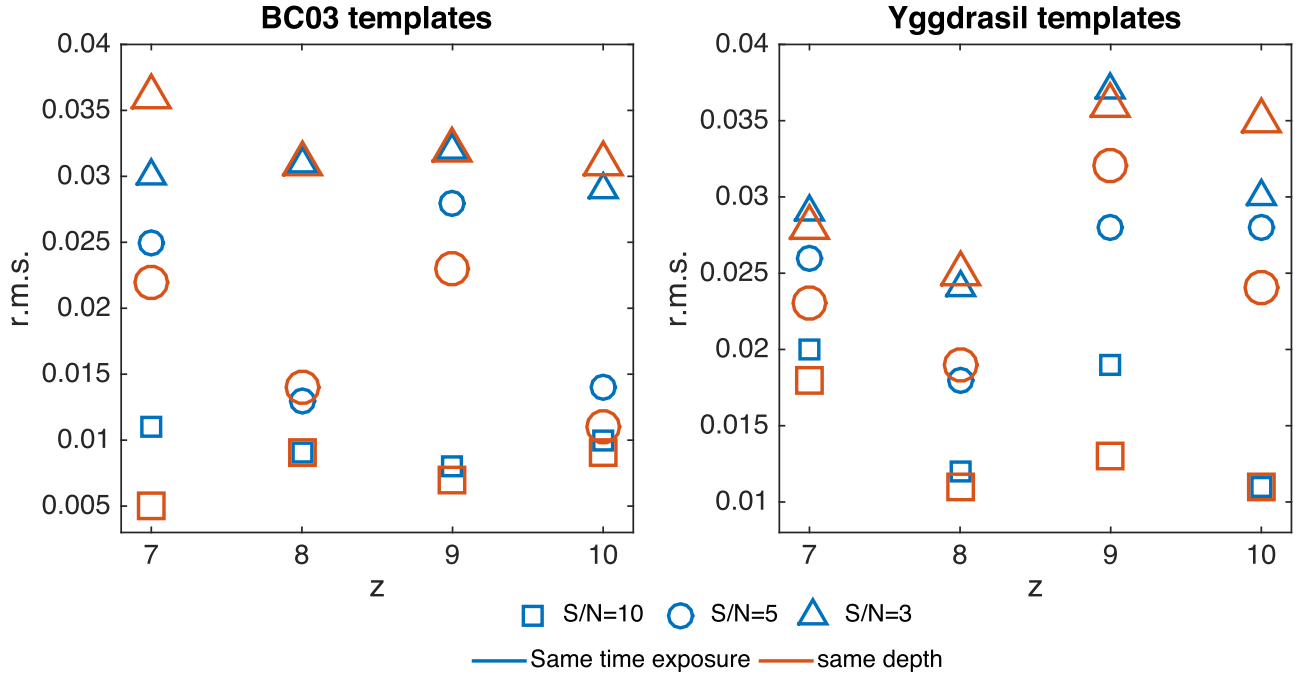
**Figure 27.** Redshifts vs. rms values of $|z_{phot} - z_{input}|/(1 + z_{input})$ derived with the NIRCam broadbands considering the same integration time (blue symbols) or the same observational depth (red symbols). Different symbols correspond to different S/N values: S/N = 10 (squares), 5 (circles), and 3 (triangles). Left: BC03 templates; right: Yggdrasil templates.

**Table 11**
Mean of $(z_{phot} - z_{input})/(1 + z_{input})$ and rms of $|z_{phot} - z_{input}|/(1 + z_{input})$ for Two Different Observational Strategies: The Same Integration Time and the Same Depth in All NIRCam Filters

| Redshift | Same Integration Time | | | Same Observational Depth | | |
|---|---|---|---|---|---|---|
| | S/N = 10 | S/N = 5 | S/N = 3 | S/N = 10 | S/N = 5 | S/N = 3 |
| $z = 7$ (BC03) | $-0.009 \pm 0.011$ | $-0.030 \pm 0.025$ | $-0.057 \pm 0.030$ | $-0.007 \pm 0.005$ | $-0.030 \pm 0.022$ | $-0.070 \pm 0.036$ |
| $z = 8$ (BC03) | $-0.021 \pm 0.009$ | $-0.031 \pm 0.013$ | $-0.018 \pm 0.031$ | $-0.017 \pm 0.009$ | $-0.023 \pm 0.014$ | $-0.025 \pm 0.031$ |
| $z = 9$ (BC03) | $-0.003 \pm 0.008$ | $0.001 \pm 0.028$ | $-0.018 \pm 0.032$ | $-0.003 \pm 0.007$ | $0.003 \pm 0.023$ | $-0.011 \pm 0.032$ |
| $z = 10$ (BC03) | $-0.002 \pm 0.010$ | $-0.006 \pm 0.014$ | $-0.083 \pm 0.029$ | $-0.002 \pm 0.009$ | $-0.080 \pm 0.011$ | $-0.042 \pm 0.031$ |
| $z = 7$ (Ygg) | $0.001 \pm 0.020$ | $-0.067 \pm 0.026$ | $-0.233 \pm 0.029$ | $-0.005 \pm 0.018$ | $-0.036 \pm 0.023$ | $-0.209 \pm 0.028$ |
| $z = 8$ (Ygg) | $-0.022 \pm 0.012$ | $-0.209 \pm 0.018$ | $-0.237 \pm 0.024$ | $-0.018 \pm 0.011$ | $-0.164 \pm 0.019$ | $-0.246 \pm 0.025$ |
| $z = 9$ (Ygg) | $0.005 \pm 0.019$ | $-0.084 \pm 0.028$ | $-0.290 \pm 0.037$ | $-0.001 \pm 0.013$ | $-0.031 \pm 0.032$ | $-0.244 \pm 0.036$ |
| $z = 10$ (Ygg) | $-0.013 \pm 0.011$ | $-0.122 \pm 0.028$ | $-0.338 \pm 0.030$ | $-0.015 \pm 0.011$ | $-0.070 \pm 0.024$ | $-0.307 \pm 0.035$ |

**Note.** Photometric redshifts are derived considering only NIRCam bands for four redshifts ($z = 7$–10) and both BC03 and Yggdrasil templates.

Finally, the $z = 9.2$ case corresponds to a redshift in which [O III] is just outside the F444W band, but H$\beta$ is still inside, while the opposite situation happens in the F560W MIRI band. The redshift estimation with NIRCam bands alone is generally good, with an rms of 0.019 and only one outlier. When adding the MIRI bands, there is a slight improvement in the redshift estimation (rms $\sigma = 0.012$) and no outliers.

The cases presented here are slightly different from the cases presented in the main body of the paper, but do not lead to any significantly different conclusion. Considering all eight analyzed redshifts and S/N = 5, the redshift estimation by using only NIRCam broadbands alone is particularly good at $z \sim 7.3$ and 8, but is the worst at $z = 8.7$ and 9. Adding MIRI bands, the redshift estimation improves at all considered redshifts, and, in the majority of cases, F560W decreases the rms a bit more than F770W.

## APPENDIX B
## SAMPLE 3: AN ALTERNATIVE OBSERVATIONAL STRATEGY

In sample 3 we derived fluxes for all NIRCam broadbands assuming the same integration time for all of them. However, the sensitivity of NIRCam peaks around 2 $\mu$m, so different integration times could be considered in order to compensate for different sensitivities. Therefore, we tested this alternative scenario by considering observations of the same depth for all NIRCam bands, i.e., the F150W band depth, and we rederived photometric redshifts for all galaxies in sample 3. In particular, this method implies deeper observations for the less sensitive bands, i.e., F070W, F090W, F115W, and F444W, with respect to the method considered before, but shallower observations for the most sensitive bands, i.e., F200W and F277W. We did this

test for both BC03 and Yggdrasil templates and for all three S/N values considered before (3, 5, and 10).

We compared the mean and rms of the normalized redshift difference $|z_{phot} - z_{fiduc.}|/(1 + z_{fiduc.})$ for the current case and that where we assumed the same integration time in all NIRCam filters. Our results are shown in Table 11, while the rms values are also shown in Figure 27. At $S/N = 10$, the derived results considering the same depth are slightly better than or similar to the one derived with the same integration times. On the other hand, at $S/N = 3$, results are similar or slightly worse. At $S/N = 5$ the results are mixed, depending on the type of template and redshift considered. These results reveal a different importance of deep observations with the most sensitive and less sensitive NIRCam bands. In particular, deep observations with the less sensitive bands seem to improve results generally when the S/N is high, while deep observations with the most sensitive bands are more crucial at low S/N values. It is worth noticing that, despite the fact that reaching the same depth in all NIRCam filters improves the results derived using NIRCam bands alone, these results are still worse than those derived using NIRCam and MIRI bands together.

## REFERENCES

Anders, P., & Fritze-v. Alvensleben, U. 2003, A&A, 401, 1063
Arnouts, S., Cristiani, S., Moscardini, L., et al. 1999, MNRAS, 310, 540
Ashby, M. L. N., Willner, S. P., Fazio, G. G., et al. 2015, ApJS, 218, 33
Balestra, I., Mainieri, V., Popesso, P., et al. 2010, A&A, 512, A12
Baum, W. A. 1957, AJ, 62, 6
Bouchet, P., García-Marín, M., Lagage, P.-O., et al. 2015, PASP, 127, 612
Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000
Bunker, A. J., Stanway, E. R., Ellis, R. S., McMahon, R. G., & McCarthy, P. J. 2003, MNRAS, 342, L47
Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, ApJ, 533, 682
Caputi, K. I., Dunlop, J. S., McLure, R. J., et al. 2012, ApJL, 750, L20
Caputi, K. I., Ilbert, O., Laigle, C., et al. 2015, ApJ, 810, 73
Cardamone, C. N., van Dokkum, P. G., Urry, C. M., et al. 2010, ApJS, 189, 270
Cristiani, S., Appenzeller, I., Arnouts, S., et al. 2000, A&A, 359, 489
Croom, S. M., Warren, S. J., & Glazebrook, K. 2001, MNRAS, 328, 150
Dahlen, T., Mobasher, B., Faber, S. M., et al. 2013, ApJ, 775, 93
Dickinson, M., Stern, D., Giavalisco, M., et al. 2004, ApJL, 600, L99
Doherty, M., Bunker, A. J., Ellis, R. S., & McCarthy, P. J. 2005, MNRAS, 361, 525
Faisst, A. L., Capak, P., Hsieh, B. C., et al. 2016, ApJ, 821, 122
Fumagalli, M., Patel, S. G., Franx, M., et al. 2012, ApJL, 757, L22
Gardner, J. P., Mather, J. C., Clampin, M., et al. 2009, ITASS, 10, 1
Glasse, A., Rieke, G. H., Bauwens, E., et al. 2015, PASP, 127, 686
Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, ApJS, 197, 35
Guo, Y., Ferguson, H. C., Giavalisco, M., et al. 2013, ApJS, 207, 24
Ilbert, O., Arnouts, S., McCracken, H. J., et al. 2006, A&A, 457, 841
Ilbert, O., Capak, P., Salvato, M., et al. 2009, ApJ, 690, 1236
Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, ApJS, 197, 36
Konno, A., Ouchi, M., Ono, Y., et al. 2014, ApJ, 797, 16
Kurk, J., Cimatti, A., Daddi, E., et al. 2013, A&A, 549, A63
Le Fèvre, O., Tasca, L. A. M., Cassata, P., et al. 2015, A&A, 576A, 79
Le Fèvre, O., Vettolani, G., Garilli, B., et al. 2005, A&A, 439, 845
Madau, P. 1995, ApJ, 441, 18
Meyer, M. R., Rieke, M., Eisenstein, D., et al. 2004, BAAS, 37, 180.09
Mignoli, M., Cimatti, A., Zamorani, G., et al. 2005, A&A, 437, 883
Morris, A. M., Kocevski, D. D., Trump, J. R., et al. 2015, AJ, 149, 178
Oke, J. B., & Gunn, J. E. 1983, ApJ, 266, 713
Pentericci, L., Vanzella, E., Fontana, A., et al. 2014, ApJ, 793, 113
Popesso, P., Dickinson, M., Nonino, M., et al. 2009, A&A, 494, 443
Rafelski, M., Teplitz, H. I., Gardner, J. P., et al. 2015, AJ, 150, 31
Ravikumar, C. D., Puech, M., Flores, H., et al. 2007, A&A, 465, 1099
Rieke, G. H., Wright, G. S., Böker, T., et al. 2015, PASP, 127, 584
Rieke, M. J., Kelly, D., & Horner, S. 2005, Proc. SPIE, 5904, 1
Schaerer, D. 2002, A&A, 382, 28
Silverman, J. D., Mainieri, V., Salvato, M., et al. 2010, ApJS, 191, 124
Smit, R., Bouwens, R. J., Labbé, I., et al. 2015, arXiv:1511.08808
Stanway, E. R., Bunker, A. J., McMahon, R. G., et al. 2004a, ApJ, 607, 704
Stanway, E. R., Glazebrook, K., Bunker, A. J., et al. 2004b, ApJL, 604, L13
Stark, D. P., Schenker, M. A., Ellis, R., et al. 2013, ApJ, 763, 129
Strolger, L.-G., Riess, A. G., Dahlen, T., et al. 2004, ApJ, 613, 200
Szokoly, G. P., Bergeron, J., Hasinger, G., et al. 2004, ApJS, 155, 271
van der Wel, A., Franx, M., van Dokkum, P. G., & Rix, H.-W. 2004, ApJL, 601, L5
Vanzella, E., Fontana, A., Pentericci, L., et al. 2014, A&A, 569, A78
Wright, G. S., Wright, D., Goodson, G. B., et al. 2015, PASP, 127, 595
Zackrisson, E., Rydberg, C.-E., Schaerer, D., Östlin, G., & Tuli, M. 2011, ApJ, 740, 13