

# illuminating the Dark Side of Cosmic Star Formation. III. Building the Largest Homogeneous Sample of Radio-selected Dusty Star-forming Galaxies in COSMOS with PhoEBO

---

Gentile, Fabrizio; Talia, Margherita; Behiri, Meriem; Zamorani, Giovanni;  
Barchiesi, Luigi; Vignali, Cristian; Pozzi, Francesca; Bethermin,  
Matthieu; Enia, Andrea; Faisst, Andreas L.; ...

Source / Izvornik: **The Astrophysical Journal, 2024, 962**

Journal article, Published version

Rad u časopisu, Objavljena verzija rada (izdavačev PDF)

<https://doi.org/10.3847/1538-4357/ad1519>

Permanent link / Trajna poveznica: <https://urn.nsk.hr/urn:nbn:hr:217:425331>

Rights / Prava: [Attribution 4.0 International](#)/[Imenovanje 4.0 međunarodna](#)

Download date / Datum preuzimanja: **2025-03-14**



Repository / Repozitorij:

[Repository of the Faculty of Science - University of Zagreb](#)





# Illuminating the Dark Side of Cosmic Star Formation. III. Building the Largest Homogeneous Sample of Radio-selected Dusty Star-forming Galaxies in COSMOS with PhoEBO

Fabrizio Gentile<sup>1,2</sup> , Margherita Talia<sup>1,2</sup> , Meriem Behiri<sup>3</sup> , Giovanni Zamorani<sup>2</sup> , Luigi Barchiesi<sup>1,2</sup> , Cristian Vignali<sup>1,2</sup> ,  
Francesca Pozzi<sup>1,2</sup> , Matthieu Bethermin<sup>4,5</sup> , Andrea Enia<sup>1,2</sup> , Andreas L. Faissst<sup>6</sup> , Marika Giulietti<sup>2,3</sup> ,  
Carlotta Gruppioni<sup>2</sup> , Andrea Lapi<sup>3,7,8,9</sup> , Marcella Massardi<sup>3,9,10</sup> , Vernesa Smolčić<sup>11</sup> , Mattia Vaccari<sup>9,12,13</sup> , and  
Andrea Cimatti<sup>1,2</sup>

<sup>1</sup> University of Bologna, Department of Physics and Astronomy (DIFA), Via Gobetti 93/2, I-40129, Bologna, Italy; [fabrizio.gentile3@unibo.it](mailto:fabrizio.gentile3@unibo.it)

<sup>2</sup> INAF—Osservatorio di Astrofisica e Scienza dello Spazio, via Gobetti 93/3-I-40129, Bologna, Italy

<sup>3</sup> SISSA, Via Bonomea 265, I-34136 Trieste, Italy

<sup>4</sup> Université de Strasbourg, CNRS, Observatoire astronomique de Strasbourg, UMR 7550, F-67000 Strasbourg, France

<sup>5</sup> Aix Marseille Univ, CNRS, CNES, LAM, Marseille, France

<sup>6</sup> Caltech/IPAC, MS314-6, 1200 E. California Blvd. Pasadena, CA 91125, USA

<sup>7</sup> IFPU—Institute for fundamental physics of the Universe, Via Beirut 2, I-34014 Trieste, Italy

<sup>8</sup> INFN-Sezione di Trieste, via Valerio 2, I-34127 Trieste, Italy

<sup>9</sup> INAF/IRA, Istituto di Radioastronomia, Via Piero Gobetti 101, I-40129 Bologna, Italy

<sup>10</sup> INAF, Istituto di Radioastronomia—Italian ARC, Via Piero Gobetti 101, I-40129 Bologna, Italy

<sup>11</sup> Department of Physics, University of Zagreb, Bijenicka cesta 32, 10002 Zagreb, Croatia

<sup>12</sup> Inter-University Institute for Data Intensive Astronomy, Department of Astronomy, University of Cape Town, 7701 Rondebosch, Cape Town, South Africa

<sup>13</sup> Inter-University Institute for Data Intensive Astronomy, Department of Physics and Astronomy, University of the Western Cape, 7535 Bellville, Cape Town, South Africa

Received 2023 March 27; revised 2023 December 4; accepted 2023 December 7; published 2024 February 2

## Abstract

In the last decades, an increasing scientific interest has been growing in the elusive population of *dark* (i.e., lacking an optical/near-IR, hereafter NIR, counterpart) dusty star-forming galaxies (DSFGs). Although extremely promising for their likely contribution to the cosmic star formation rate density (SFRD) and for their possible role in the evolution of the first massive and passive galaxies around  $z \sim 3$ , the difficulty in selecting statistically significant samples of dark DSFGs is limiting their scientific potentialities. This work presents the first panchromatic study of a sample of 263 radio-selected NIR-dark (RS-NIRdark) galaxies discovered in the COSMOS field following the procedure by Talia et al. These sources are selected as radio-bright galaxies ( $S_{3\text{ GHz}} > 12.65\ \mu\text{Jy}$ ) with no counterpart in the NIR-selected COSMOS2020 catalog ( $K_s \gtrsim 25.5\ \text{mag}$ ). For these sources, we build a new photometric catalog including accurate photometry from the optical to the radio obtained with a new deblending pipeline (Photometry Extractor for Blended Objects, or PHOEBE). We employ this catalog to estimate the photo- $z$ s and the physical properties of the galaxies through an spectral energy distribution-fitting procedure performed with two different codes (MAGPHYS and CIGALE). Finally, we estimate the active galactic nucleus contamination in our sample by performing a series of complementary tests. The high values of the median extinction ( $A_v \sim 4$ ) and star formation rate ( $\text{SFR} \sim 500\ M_\odot\ \text{yr}^{-1}$ ) confirm the likely DSFG nature of the RS-NIRdark galaxies. The median photo- $z$  ( $z \sim 3$ ) and the presence of a significant tail of high- $z$  candidates ( $z > 4.5$ ) suggest that these sources are important contributors to the cosmic SFRD and the evolutionary path of galaxies at high redshifts.

*Unified Astronomy Thesaurus concepts:* [Extragalactic radio sources \(508\)](#); [Galaxy evolution \(594\)](#); [Galaxy formation \(595\)](#); [High-redshift galaxies \(734\)](#); [Star formation \(1569\)](#)

## 1. Introduction

Our picture of the Universe is essentially the offspring of the instruments we use to observe it. Until a few years ago, the Hubble Space Telescope (HST) was the most powerful facility available to study the high- $z$  Universe. The data from this facility have put several critical constraints on important quantities as a function of cosmic time (see, e.g., the review by Madau & Dickinson 2014 and references therein). One of the main results concerned the reconstruction of the cosmic star formation rate density (SFRD; i.e., the amount of stellar mass

formed in the Universe per each year and  $\text{Mpc}^3$ ) back to the epoch of reionization ( $z \sim 8$ ). The vast majority of these studies agreed in describing an SFRD constantly growing  $\propto (1+z)^{1.7}$  from  $z \sim 0$  up to  $z \sim 2.5$  (the so-called *cosmic noon*) and—then—decreasing  $\propto (1+z)^{-2.9}$  up to  $z \sim 8$  (Madau & Dickinson 2014). Given the spectral coverage of HST, however, our census of galaxies in the first few billion years of cosmic history was forcedly limited to optically bright galaxies. Given the redshift ranges explored, these sources were mainly rest-frame UV/optically bright galaxies.

Even in that time, however, an increasing number of evidences suggested that our picture of galaxy evolution based on these UV-bright sources was quite far from complete. Above all, the discovery of a significant population of massive ( $M_* > 10^{11} M_\odot$ ) and passive ( $\text{sSFR} < 10^{-11}\ \text{yr}^{-1}$ )

galaxies already in place at  $z \sim 3$  (i.e., when the Universe was just 2 Gyr old; see some examples in Straatman et al. 2014; Schreiber et al. 2018b) was—probably—the most outstanding. The number density of these sources ( $n \sim 2 \times 10^{-5} \text{ Mpc}^{-3}$ ) resulted in being more than 2 orders of magnitude higher than that computed for UV-bright galaxies at  $z > 4.5$  (see, e.g., the discussion in Straatman et al. 2014; Schreiber et al. 2018b; Valentino et al. 2020). This suggested that, when not accounting for UV-dark star-forming galaxies, we are missing a significant fraction of the progenitors of high- $z$  passive galaxies.

Observations carried out with facilities able to explore other regions of the electromagnetic spectrum such as the sub-millimeter array camera SCUBA on the James Clerk Maxwell Telescope (e.g., Smail et al. 1997; Hughes et al. 1998) or the Herschel Space Observatory (e.g., Burgarella et al. 2013; Gruppioni et al. 2013) confirmed that selections based only on the optical regime actually miss a significant fraction of high- $z$  galaxies (see, e.g., Blain & Phillips 2002; Weiß et al. 2013). Among the main missing population of galaxies, the so-called “Dusty Star-forming Galaxies” (DSFGs; see, e.g., Lagache et al. 2005; Casey et al. 2014 for reviews) are worth a deeper discussion. These sources are characterized by high values of the star formation rate (SFR) and high stellar extinction due to the presence of a significant amount of dust. Hence, these sources are extremely faint or even undetected in the optical/near-IR (NIR) regime observed by HST. Despite their elusiveness, several studies targeting these sources (e.g., Wang et al. 2019; Gruppioni et al. 2020; Talia et al. 2021; Zavala et al. 2021; Enia et al. 2022; van der Vlugt et al. 2023) are unanimous in finding that the contribution of these galaxies to the cosmic SFRD in the high- $z$  regime is definitely significant (up to 40% the contribution of optically bright galaxies at  $z > 4.5$ ; see, e.g., Talia et al. 2021; Enia et al. 2022; Behiri et al. 2023). The inclusion of these new sources in the cosmic census of the high- $z$  galaxies could even change the behavior of the SFRD at  $z > 3$  (favoring a flatter behavior before the cosmic noon; see, e.g., Gruppioni et al. 2020) and solve the puzzle of the missing progenitors of massive galaxies (e.g., Toft et al. 2014).

It is crucial to underline how all the aforementioned studies on DSFGs are generally based on limited samples of galaxies. The main reason for this is the extreme faintness of these sources in the optical/NIR regime, making their identification feasible only at longer wavelengths. For this reason, these galaxies are generally selected in the IR or (sub)millimeter regimes (e.g., Dudzevičiūtė et al. 2021; Smail et al. 2021; Giulietti et al. 2022; Manning et al. 2022), taking advantage of the bright thermal emission of the dust and the intensely negative  $k$ -correction that makes these sources bright even at high redshifts.

IR/(sub)millimeter selections, however, are affected by—at least—two significant issues. The first one is the low sensitivity of single-dish (sub)millimeter telescopes. This, combined with their large beam sizes, makes it hard to identify the multi-wavelength counterparts of each galaxy. In principle, this problem could be overcome by employing state-of-the-art facilities such as the Atacama Large Millimeter/submillimeter Array (ALMA) with higher sensitivities and smaller beams. However, the deep surveys conducted with these telescopes (e.g., Franco et al. 2018; Casey et al. 2021) are currently mapping small volumes, hence are affected by low statistics

and are prone to cosmic variance. The second issue concerning IR/(sub)millimeter selections resides in the still-debated properties of the dust in the high- $z$  regime. For instance, a correlation between the dust temperature and the redshift (e.g., the one suggested by Béthermin et al. 2015; Faisst et al. 2017; Schreiber et al. 2018a) could produce a significant selection bias.

A solution to both the aforementioned issues can reside in the radio selection of DSFGs (e.g., Talia et al. 2021; Enia et al. 2022; Behiri et al. 2023). In *normal* (i.e., nonactive) galaxies, radio photons can be generated by free-free emission in H II regions and synchrotron emission from relativistic electrons accelerated in supernovae remnants. In addition, they are not affected by the presence of dust. For these reasons, after excluding the nuclear origin of these photons, they could represent an excellent unbiased tracer of star formation. Therefore, as noted by Talia et al. (2021), the selection of radio-bright galaxies with a faint or undetected counterpart in the optical/NIR regime could provide a significant sample of likely DSFGs.

This paper expands the studies by Talia et al. (2021), Behiri et al. (2023), which aimed to assemble the widest homogeneously selected sample of candidate DSFGs available in the current literature. We take advantage of the multiwavelength coverage in the Cosmic Evolution Survey (COSMOS; Scoville et al. 2007) field and the high sensitivity of the Very Large Array (VLA)-3 GHz COSMOS Large Project (Smolčić et al. 2017) to collect a sample of 323 radio-selected NIR-dark galaxies (RS-NIRdark galaxies in the following). In this paper, we present the photometric catalog of these sources with new accurate photometry extracted from the optical to the radio regimes. We also employ a spectral energy distribution (SED)-fitting procedure to estimate photo- $z$ s and physical properties of our galaxies.

The paper is structured as follows. In Section 2, we introduce the sample selection. In Section 3, we describe the maps employed to extract the photometry in the optical-to-mid-infrared (MIR) regime and the additional photometry retrieved at longer wavelengths. In Section 4, we present the new Photometry Extractor for Blended Objects pipeline (PHOEBO) used to extract the photometry and its validation on simulated data. Section 5 is focused on the SED-fitting procedure employed to assess the photo- $z$ s and the physical properties of the galaxies in our sample. Furthermore, in Section 6, we analyze in detail the possible active galactic nucleus (AGN) contamination in our sample. In Section 7, we discuss what the results presented in this paper tell us about the nature of the RS-NIRdark galaxies and compare them with previous analogous studies in the current literature. Finally, we draw our conclusions in Section 8. Throughout this paper, we use AB magnitudes (Oke & Gunn 1983), employ a Chabrier (2003) initial mass function (IMF), and assume a concordance cosmology  $\Lambda$ CDM with  $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$ , and  $(\Omega_{\text{tot}}, \Omega_{\Lambda}, \Omega_m) = (1, 0.7, 0.3)$ .

## 2. Sample Selection

To assemble a large sample of candidate DSFGs with complete photometry from the optical to the radio, we focus on the galaxies in the COSMOS field. We collect our sources by performing a selection analogous to that employed by Talia et al. (2021) and summarized here:

1. We start from the VLA-COSMOS 3 GHz Large Project catalog (Smolčić et al. 2017). We select 8850 radio sources with an signal-to-noise ratio (S/N)  $> 5.5$  (i.e., brighter than  $12.65 \mu\text{Jy beam}^{-1}$ ) over the full  $2 \text{ deg}^2$  coverage of the survey. This cut allows us to assemble a sample of radio-bright galaxies, limiting the likely contamination of fake sources to 0.4% (Smolčić et al. 2017).
2. To limit the expected presence of galaxies hosting AGN (common in radio-selected catalogs; see, e.g., Bonzini et al. 2013; Novak et al. 2018), we remove from the sample all the sources flagged as “multi-component” in the initial catalog. We underline that this flag is the result of a visual inspection of the sources in the full catalog by Smolčić et al. (2017). Therefore—at this stage—we cannot exclude that some multicomponent radio sources are still present in our sample (see the discussion in Section 6).
3. To take advantage of the multiwavelength coverage of the COSMOS field, we exclude from the sample all the radio sources lying outside the UltraVISTA survey footprint (Laigle et al. 2016). This further limits the sample to 5982 galaxies and the effective area mapped by this study to  $1.38 \text{ deg}^2$ .
4. Finally, we cross-match the resulting sample with the two versions of the COSMOS2020 catalog (the “CLASSIC” and “FARMER”; see Weaver et al. 2022), removing all the sources with a match within  $0''.7$ . The final sample is composed of 323 galaxies.

The last step aims to include in the catalog only sources without a significant counterpart in the optical/NIR bands. However, it is crucial to underline how the source detection in the COSMOS2020 catalog is performed on the  $\chi^2$ -image produced through the SWARP software (Bertin et al. 2002) by combining the maps in the  $i$  and  $z$  band from the Subaru telescope and the  $Y$ ,  $J$ ,  $H$ , and  $Ks$  bands from the VISTA telescope (see Weaver et al. 2022). The flux in the generic pixel  $I$  of this image is computed as a weighted average of the fluxes  $f_i$  in the different photometric bands, with the weights  $w_i$  provided by the uncertainty maps:

$$I = \sqrt{\frac{\sum_{i=1}^N w_i f_i^2}{N}}. \quad (1)$$

Given this definition, we cannot exclude that some of the sources in our catalog could have a significant counterpart in a few individual NIR bands. Moreover, through a visual inspection of the  $\chi^2$ -map employed by Weaver et al. (2022) for detecting the sources in the COSMOS2020, we notice that 60 galaxies of our sample have a detection in that image with an S/N higher than  $1.5\sigma$  (i.e., the threshold used in Weaver et al. 2022). Some of these sources ( $\sim 10\%$ ) are inside regions masked in the COSMOS2020 that were employed in the COSMOS2015, while most of the lasting sources are close to a bright companion; therefore, it is likely that they were not properly deblended in the COSMOS2020. Since these galaxies are not included in the COSMOS2020, they should be part of our sample of RS-NIRdark galaxies. However, due to the different photometry, we expect their properties to differ from the rest of the sample. Therefore, we do not consider these sources in all the statistical analyses performed in this paper aimed to characterize our population of galaxies: for all these studies, we will focus on the lasting 263 galaxies. Finally, we

underline that the S/N cut applied to the radio catalog excludes from the sample all the faintest radio sources. Although this step could potentially cause the lack of significantly high- $z$  galaxies (see, e.g., Casey et al. 2019), it is necessary to exclude a high contamination by fake sources (this rate would increase up to 25% just including the sources with  $5 < S/N < 5.5$ ; see Smolčić et al. 2017).

A significant difference between this study and the previous papers of this series (Talia et al. 2021; Behiri et al. 2023) resides in the last step of the selection: while the previous works selected a sample of 476 RS-NIRdark galaxies without a counterpart in the COSMOS2015 catalog (Laigle et al. 2016), we employ the updated (and deeper) COSMOS2020 catalog (Weaver et al. 2022). This step allows us to exclude from the sample 153 galaxies undetected in the COSMOS2015 but revealed by COSMOS2020. More in detail, 16 sources were only included in the “FARMER” catalog, 21 only in the “CLASSIC,” and 116 in both the catalogs. Finally, improving the works by Talia et al. (2021), Behiri et al. (2023), we do not employ any additional selection aimed to avoid the sources with a bright contaminant in the vicinity, but we analyze the entire sample, performing a more accurate photometry extraction in presence of contamination or source blending. A more comprehensive comparison with Talia et al. (2021), Behiri et al. (2023) can be found in Appendix A.

### 3. Data

#### 3.1. Analyzed Maps

The analysis of DSFGs requires the most comprehensive wavelength coverage to account for all the physical processes taking place in these complex objects (i.e., stellar emission, dust obscuration, thermal emission, and nonthermal processes). For this purpose, we extract accurate photometry in the optical-to-MIR regimes by analyzing the following maps:

1. *Optical*. We analyze the maps produced with the Subaru telescope’s Hyper Suprime Cam (HSC) during the Subaru Strategic Program (DR3; Aihara et al. 2019), targeting the COSMOS field in the  $g$ ,  $r$ ,  $i$ ,  $z$ , and  $y$  bands.
2. *NIR*. The photometry at NIR wavelengths is extracted from the DR4 maps of the UltraVISTA survey (McCracken et al. 2012) performed with the VISTA telescope in the  $Y$ ,  $J$ ,  $H$ , and  $Ks$  bands.
3. *MIR*. We analyze the maps produced with the Infrared Array Camera (IRAC) of the Spitzer space telescope. The maps analyzed in this study are the deepest ones made with this facility as part of the Cosmic Dawn Survey (Moneti et al. 2022), obtained by coadding all the available exposures of the COSMOS field for each of the four channels of the IRAC camera.

Further details on the maps employed in the photometry extraction can be found in Table 1.

#### 3.2. Additional Photometry

To analyze the dust emission in the far-IR (FIR)/(sub)millimeter and the nonthermal processes emitting at radio frequencies, we retrieve additional photometry for our sources by cross-matching our sample with other catalogs in the current literature:

1. *FIR*. We cross-match our catalog with the 2020 version of the superdeblended catalog by Jin et al. (2018). Since the



**Table 1**  
Main Properties of the Maps Analyzed in Section 3

| Instrument/Telescope        | Band      | $\lambda^a$<br>(Å) | $\Delta\lambda^b$<br>(Å) | Depth <sup>c</sup><br>( $2''$ ) | Corr. Fact. <sup>d</sup> | PSF FWHM <sup>e</sup><br>( $''$ ) |
|-----------------------------|-----------|--------------------|--------------------------|---------------------------------|--------------------------|-----------------------------------|
| Hyper-Suprime<br>Cam/Subaru | <i>g</i>  | 4847               | 1383                     | 28.1                            | 1.4                      | 0.79                              |
|                             | <i>r</i>  | 6219               | 1547                     | 27.8                            | 1.4                      | 0.75                              |
|                             | <i>i</i>  | 7699               | 1471                     | 27.6                            | 1.5                      | 0.61                              |
|                             | <i>z</i>  | 8894               | 766                      | 27.2                            | 1.4                      | 0.68                              |
|                             | <i>y</i>  | 9761               | 786                      | 26.5                            | 1.4                      | 0.68                              |
| VIRCAM                      | <i>Y</i>  | 10,216             | 963                      | 25.3/26.6 <sup>f</sup>          | 2.7/2.8 <sup>f</sup>     | 0.82                              |
|                             | <i>J</i>  | 12,525             | 1718                     | 25.2/26.4 <sup>f</sup>          | 2.5/2.7 <sup>f</sup>     | 0.79                              |
| /VISTA                      | <i>H</i>  | 16,466             | 2905                     | 24.9/26.1 <sup>f</sup>          | 2.4/2.6 <sup>f</sup>     | 0.76                              |
|                             | <i>Ks</i> | 21,557             | 3074                     | 25.3/25.7 <sup>f</sup>          | 2.4/2.4 <sup>f</sup>     | 0.75                              |
| IRAC<br>/Spitzer            | ch1       | 35,686             | 7443                     | 26.4                            | ...                      | ...                               |
|                             | ch2       | 45,067             | 10,119                   | 26.3                            | ...                      | ...                               |
|                             | ch3       | 57,788             | 14,082                   | 23.2                            | ...                      | ...                               |
|                             | ch4       | 79,958             | 28,796                   | 23.1                            | ...                      | ...                               |

**Notes.** Part of the data is reproduced from Weaver et al. (2022) with author’s permission.

<sup>a</sup> Median  $\lambda$  of transmission curve.

<sup>b</sup> FWHM of the transmission curve.

<sup>c</sup>  $3\sigma$  depths as reported by Weaver et al. (2022).

<sup>d</sup> Multiplicative correction for the photometric uncertainties (Weaver et al. 2022).

<sup>e</sup> Values taken from Aihara et al. (2019), McCracken et al. (2012). For the IRAC maps, we employ the PSF publicly available on the IRSA website.

<sup>f</sup> Two values for the deep/ultra-deep stripes in the UltraVISTA survey.

procedure followed in building that catalog employs as positional prior the sources in the VLA-COSMOS 3 GHz Large Project, it is possible to retrieve FIR photometry obtained with Spitzer (24  $\mu\text{m}$ ; Le Flocc’h et al. 2009), Herschel (100, 160, 200, 250, and 500  $\mu\text{m}$ ; Lutz et al. 2011), SCUBA2 (850  $\mu\text{m}$ ; Cowie et al. 2017; Geach et al. 2017), AzTEC (1.1 mm; Aretxaga et al. 2011), and MAMBO (1.2 mm; Bertoldi et al. 2007) for all the galaxies in our sample. Further details on the employed maps and on their depth can be found in Jin et al. (2018).

- (Sub)millimeter. We cross-match our sample with the catalog of the Automated Mining of the ALMA Archive in the COSMOS Field (A3COSMOS) survey (v.20200310; Liu et al. 2019). This catalog contains (sub)millimeter continuum fluxes for all the sources in the COSMOS field observed with ALMA. Since the coverage of the A3COSMOS survey is not uniform, only a tiny percentage of our RS-NIRdark sources has a counterpart in this catalog (32 galaxies within a matching radius of  $1''$ ,  $\sim 10\%$ ; Liu et al. 2019), with central bandwidth and depth strongly variable for different sources.
- Radio. Finally, we retrieve additional radio photometry by cross-matching our catalog with the public catalog by Schinnerer et al. (2010) containing 1.4 GHz photometry obtained during the VLA-COSMOS survey (77 galaxies;  $\sim 24\%$  of the full sample). We also include radio photometry at 1.28 GHz from the MIGHTEE Early Science Data Release (170 sources within a matching radius of  $8''$ ;  $\sim 53\%$  of the sample; Jarvis et al. 2016; Heywood et al. 2022).

## 4. PHOEB0: A New Pipeline for Photometry Extraction

### 4.1. Description of the Pipeline

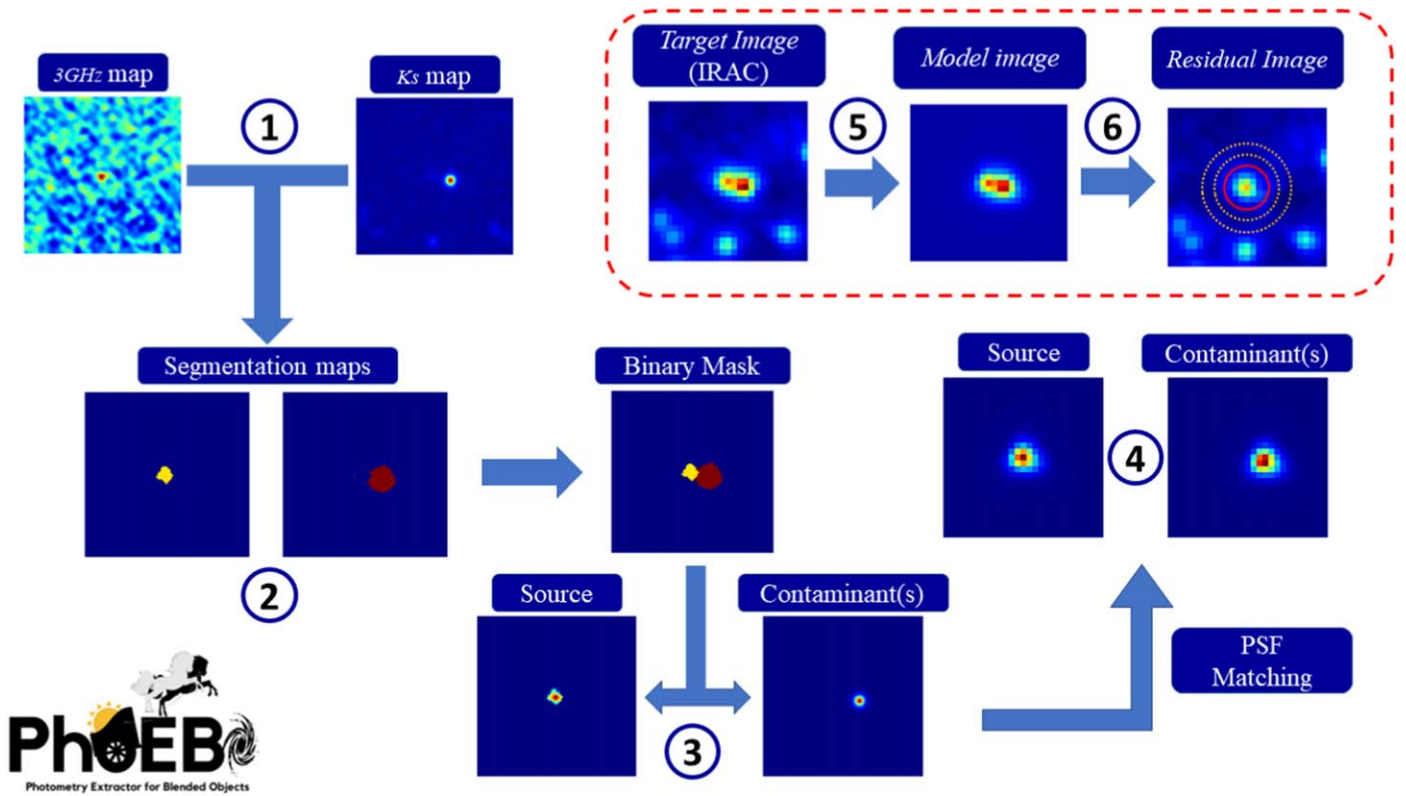
Extracting accurate photometry for the sources in our catalog is a challenging task. The main limitation is represented by the possible presence of bright contaminants close to our

(extremely faint or even undetected) galaxies in the optical/NIR bands. This issue becomes highly significant in the IRAC channels, where the large (up to  $\text{FWHM} = 2''$  in ch3 and ch4) and irregular point-spread function (PSF) makes it almost impossible to blindly deblend multiple sources without priors on their positions and shapes.

Basic deblending algorithms such as that implemented in SEXTRACTOR (Bertin & Arnouts 1996) need a minimum contrast between the different blended components inside the same cluster of pixel to recognize the presence of multiple sources. This level of contrast is not reached in the IRAC bands, making these algorithms unfit to analyze our galaxies. More complex softwares such as TRACTOR (Lang et al. 2016) and FARMER (Weaver et al. 2022), relying on profile-fitting, can overcome this problem by extracting prior information on the position and shape of the different objects through a high-resolution image in which all the blended components are present and distinguishable at the same time (e.g., the  $\chi^2$ -map employed by Weaver et al. 2022). As before, these techniques cannot be applied to our RS-NIRdark galaxies, since the only bands in which our galaxies are robustly detected—together with the contaminants—are the IRAC channels, where it is generally impossible to distinguish the different sources. Therefore, in order to build the photometric catalog, we developed a new deblending pipeline called PHOEB0 (Gentile et al. 2023)<sup>14</sup> relying on a slightly modified implementation of the method employed in several previous studies (e.g., Labbé et al. 2006; Endsley et al. 2021; Whittler et al. 2023). The whole procedure is summarized in Figure 1 and follows these steps:

1. We start from two high-resolution *detection images* and a low-resolution *target image* to be deblended. One *detection image* must contain information on the RS-NIRdark galaxy we want to analyze; the other must

<sup>14</sup> The code is freely available in a GITHUB repository: <https://github.com/fab-gentile/PhoEBO>; a current copy has been deposited to Zenodo (Gentile et al. 2023).



**Figure 1.** Scheme summarizing the workflow of the debblending algorithm PHOEBE employed to extract the optical/NIR/MIR photometry for the RS-NIRdark galaxies in the sample. Further details on the numbered steps and on the full procedure are given in Section 4.1.

contain information on the contaminant sources. In our case, we choose the 3 GHz and the  $K_s$  maps, respectively. We underline that we minimize the risk of biases by employing two maps with a comparable PSF FWHM ( $0''.75$  and  $0''.78$ , respectively).

2. We employ the basic debblending algorithm included in the SEP library (Bertin & Arnouts 1996; Barbary 2016a) to identify the different sources in the two *detection images* and to produce the segmentation maps (i.e., a set of images assigning each pixel to one of the galaxies in the field). To include all the galaxies in the maps, we set a low detection threshold ( $2\sigma$ ) and a minimum of 5 contiguous pixels in each detected source.
3. We combine the two segmentation maps and use their product as a binary mask to isolate the sources present in the *detection images* and produce a different image for each of them.
4. We convolve each image produced in the previous step with a matching kernel to homogenize the PSF with that of the *target image*. The matching kernel is produced with the “PHOTUTILS” library (Bradley et al. 2020) using the ratio of Fourier transforms (Gordon et al. 2008; Aniano et al. 2011). The PSF of the *detection images* is modeled as a 2D Gaussian with an FWHM equal to the PSF FWHM of the  $K_s$  band reported in McCracken et al. (2012), Weaver et al. (2022). The PSF of the optical/NIR bands is obtained in the same way (i.e., with Gaussians with fixed widths, see Table 1), while the PSFs of the four IRAC channels are downloaded from the IRSA archive.<sup>15</sup>

5. We normalize all the resulting images and coadd them into a *model image*. This one is resampled to match the pixel size of the *target image* (if needed) and then fitted to the *target image* by multiplying each normalized source by a free-parameter  $\alpha_i$ . The fit is performed with the SCIPY library (Virtanen et al. 2020), aiming to minimize the  $\chi^2$  between the *model* and the *target image*.
6. We multiply all the components of the *model image* for the relative  $\alpha_i$  obtained through the fitting procedure. Then, we subtract all the resulting images of the contaminants from the original *target image*. In doing so, we get a *residual image* containing only the source present in our sample.
7. Finally, we perform aperture photometry with PHOTUTILS on this residual image by employing a fixed diameter of  $2''$ . The local background is computed in an annulus between 1.5 and 2 times the radius of the aperture and subtracted from the extracted counts. The counts are then converted to AB magnitudes and to microJansky through the photometric zero-points employed in Weaver et al. (2022), corrected for the systematic offset reported in Weaver et al. (2022). This last correction is needed to account for the systematic mismatch between the spec-zs and the photo-zs computed in the COSMOS field (Laigle et al. 2016; Weaver et al. 2022).

The whole procedure described above allows us to estimate the fluxes in all the NIR and IRAC bands reported in Section 3.1. As prescribed by the IRAC Instrument Handbook, we correct the fluxes measured through aperture photometry by the factors reported in the IRAC Instrument Handbook.<sup>16</sup> For

<sup>15</sup> <https://irsa.ipac.caltech.edu/data/SPITZER/docs/irac/>

<sup>16</sup> <https://irsa.ipac.caltech.edu/data/SPITZER/docs/irac/iracinstrumenthandbook/>

the optical bands—with a smaller PSF FWHM than the radio map at 3 GHz and the  $K_s$  map—we subtract the contaminants with PHOEB0 by using the HSC- $i$  band as a *detection image*. Finally, we estimate the photometric uncertainties as follows:

$$\Delta F = \sqrt{\sum_{i \in A} \sigma_i^2} \quad (2)$$

where the  $\sigma$  is obtained from the weight maps, and the sum is extended to all the pixels in the aperture employed in the estimation of the flux. As prescribed by Weaver et al. (2022), we correct the photometric uncertainties in the optical/NIR bands with the multiplicative factors reported in Table 1. This step is required to account for the expected underestimation of the uncertainties through Equation (2) due to the presence of correlated noise in the analyzed maps (e.g., Leauthaud et al. 2007). Since the weight maps are not affected by the subtraction of the contaminants, we underline that the procedure followed here to estimate the photometric uncertainties is totally consistent with that employed by Weaver et al. (2022) for the optical/NIR bands. Regarding the IRAC bands, we can compare our uncertainties with those by Weaver et al. (2022) by running PHOEB0 on the 153 RS-NIRdark galaxies selected by Talia et al. (2021) and excluded from this study being revealed in the COSMOS2020 (see Section 2). We obtain that the median ratio between our uncertainties and those included in the *classic* COSMOS2020 is in the order of  $\sim 3$ . This result can be explained by the likely underestimation of the IRAC uncertainties found by Weaver et al. (2022) in the COSMOS2020 catalog.

#### 4.2. Validation of PHOEB0

Thanks to the small difference in the PSFs, the photometry extraction performed by PHOEB0 in the optical/NIR bands does not differ significantly from that performed by “classic” algorithms such as SEXTRACTOR (Bertin & Arnouts 1996). Nevertheless, the extraction in the IRAC channels is quite different. Therefore, in order to obtain reliable results on the physical properties extracted from the photometric catalog, we need to validate the performances of the pipeline in the MIR regime. We validate the results of the PHOEB0 pipeline performing extensive simulations of blended galaxies in the four IRAC channels. The simulation procedure recalls the philosophy discussed in Section 4.1:

1. We simulate two noise maps through a random Gaussian generator, requiring an rms compatible with the 3 GHz images observed in the VLA-COSMOS survey at 3 GHz and with the  $K_s$ -band images observed in the UltraVISTA survey (conservatively, we employ the sensitivity reached in the deep stripes).
2. We simulate a radio source and an NIR-bright contaminant. Both the galaxies are simulated as 2D Gaussians on the noise maps generated in the previous step. Since the vast majority of the RS-NIRdark galaxies are unresolved at 3 GHz, we choose an FWHM =  $0''.7$  for the radio sources (i.e., the width of the synthesized beam in the VLA-COSMOS survey). On the contrary, to account for the presence of partially resolved contaminants, the FWHMs of the Gaussians in the  $K_s$  band are uniformly sampled in the range  $[0.7, 1.4]$  arcsec (i.e., between 1 and 2 PSF FWHM).

The normalizations of the two Gaussians are chosen to obtain an S/N uniformly sampled in the range  $[5.5, 8]$  and  $[3, 8]$ , for the radio and  $K_s$  images, respectively. The radio sources are placed in the center of each image, while the contaminants are allowed to space in the range  $[0.7, 1.4]$  arcsec from the center. This range allows us to study the accuracy of the algorithm as a function of the blending between the two sources. The lower limit on the distance is chosen as  $0''.7$ , since we recall that—due to the selection described in Section 2—all the NIR sources with a separation lower than this threshold are considered NIR counterparts of the radio signal.

3. We convolve each of these Gaussians with a set of matching kernels computed as prescribed in Section 4.1 to obtain the sources as observed in each of the four IRAC channels.
4. We rescale the flux of the convolved radio source to obtain an integrated flux in the range  $[21, M]$  mag, where  $M$  is the limiting magnitude in each IRAC channel. Consequently, we rescale the flux of the convolved contaminant to obtain a flux ratio uniformly sampled in the range  $[0.1, 1]$  (with the radio source brighter than the other galaxy).
5. Finally, we coadd the two images and add Gaussian noise with the same rms expected in the four IRAC channels.

To explore the whole parameter space of the randomly sampled quantities, we simulate  $\sim 10^3$  images. Then, we run the PHOEB0 pipeline on the IRAC-like images, employing as *detection images* the high-resolution data in the radio and  $K_s$  bands. In order to assess the performances of the pipeline, we compare the fluxes reported by PHOEB0 with those obtained by performing a standard aperture photometry on the isolated IRAC-like images (i.e., those obtained in point 3 of the simulation procedure, before adding the contaminant).

Computing the  $\Delta$  mag on the whole data set, we obtain a median( $\Delta$ mag)  $\sim 0.03$  in all the IRAC channels, without any significant difference between the different bands. Similarly, we obtain a std( $\Delta$ mag)  $\sim 0.15$  in all the channels. Moreover, the employment of simulations allows us to estimate the  $\Delta$  mag as a function of the different parameters employed during the simulation procedure (Figure 2). We find interesting—albeit expected—trends with the IRAC flux and with the angular separation between the high-resolution images, with lower accuracy achieved on more blended sources and IRAC-fainter RS-NIRdark galaxies.

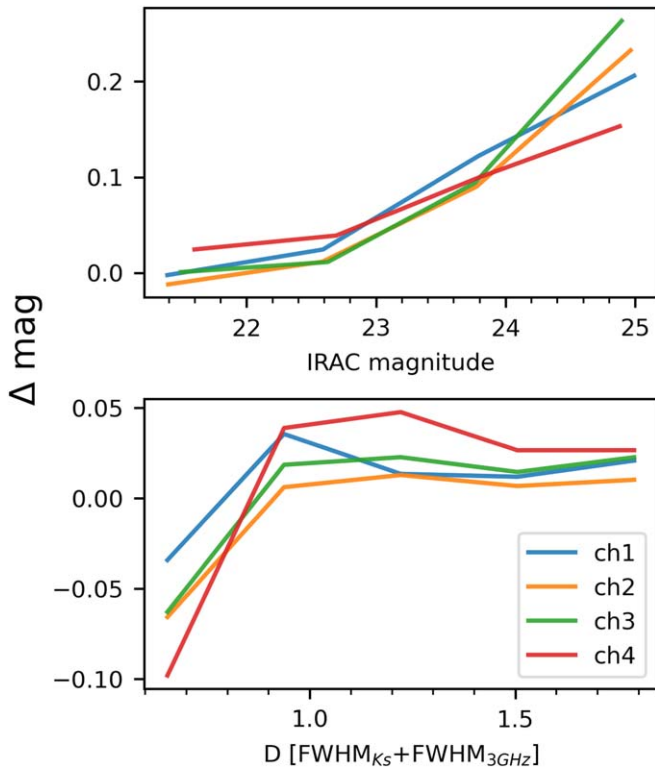
## 5. Physical Properties from SED Fitting

To determine the nature of the RS-NIRdark galaxies, we need to assess their photo- $z$ s and physical properties. We do this through an SED-fitting procedure. To test the robustness of our results against the characteristics of different codes (see some examples in Figure 3), we base our analysis on two algorithms: MAGPHYS+PHOTO- $z$  (da Cunha et al. 2008; Battisti et al. 2019) and CIGALE (Boquien et al. 2019).

### 5.1. SED Fitting with MAGPHYS

MAGPHYS (da Cunha et al. 2008) is a physically motivated SED-fitting code based on the energy balance between stellar attenuation and thermal dust emission. The software estimates the physical properties of a galaxy by comparing its optical-to-radio broadband photometry with more than a million





**Figure 2.** Accuracy of the PHOEBE pipeline as a function of the IRAC flux of the RS-NIRdark galaxies and of the angular distance between these sources and the blended contaminant (a proxy for the blending between the sources, once parameterized through the sum of the FWHMs of the two sources). These results are obtained by applying the PHOEBE pipeline to the set of simulated images described in Section 4.2.

templates, including stellar emission, dust attenuation, thermal dust emission, and nonthermal radio emission. The stellar emission is considered by combining the single stellar populations (SSPs) by Bruzual & Charlot (2003) with an exponentially declining star formation history (SFH) with random bursts of star formation superimposed on the continuum. The dust attenuation is included as prescribed by Charlot & Fall (2000), with the addition of a 2175Å feature accounting for the young stars born in denser clouds. A three-component model accounts for thermal dust emission. These components include the *hot dust* (i.e., warmed up by young stars in the birth clouds), the *cold dust* (i.e., present in the diffuse interstellar medium), and the characteristic emission by the polycyclic aromatic hydrocarbons (PAHs). The first two components are modeled with a modified graybody, while the PAH emission is modeled with an empirical template (see da Cunha et al. 2008 for details). The code computes the dust temperature ( $T_d$ ) as the luminosity-average of these three components. Finally, MAGPHYS includes the radio emission from star formation as prescribed by da Cunha et al. (2015). All the ranges of the free parameters employed in the templates can be found in da Cunha et al. (2008, 2015). In this work, we use the PHOTO-z version of MAGPHYS (Battisti et al. 2019), able to estimate the photometric redshift together with the physical properties of the analyzed galaxies.

### 5.2. SED Fitting with CIGALE

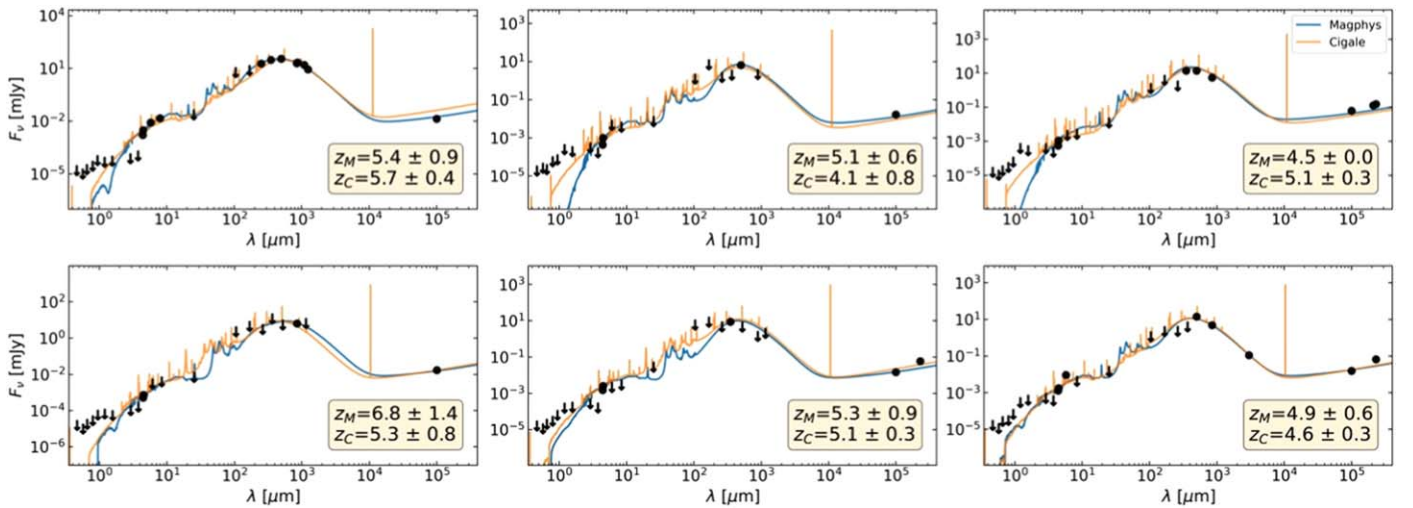
To avoid possible biases arising from the use of a single SED-fitting code, we involve the software CIGALE (Boquien et al. 2019)

in the analysis. This code is based on a similar energy-balance principle as MAGPHYS but allows a larger customization of the libraries employed in building the templates. In this paper, we start by using a setup as close as possible to MAGPHYS to achieve consistent results. A detailed investigation of the different modules included in CIGALE will be discussed in a forthcoming paper. The SSPs, SFH, dust attenuation, and radio emission are the same as discussed in Section 5.1. A significant difference between the two codes resides in the treatment of the thermal dust emission. CIGALE does not have a single model including both the graybody thermal emission and the PAH emission as MAGPHYS. The choice is limited to the models by Draine & Li (2007) and their updated version by Draine et al. (2014) parameterized through the intensity of the radiation field and including the PAHs emission. A second possibility is the analytical model by Casey (2012) parameterized through the dust temperature but not including the PAHs. Since our photometric catalog includes a point at  $24\ \mu\text{m}$  from the superdeblended catalog, sampling the typical PAHs emission at  $z \sim 3$ ; and since this feature is generally crucial for determining a robust redshift, we decide to employ the Draine et al. (2014) model. We compute the mass-weighted dust temperature starting from the intensity of the radiation field reported in output by CIGALE following the framework described in Draine & Li (2007), Draine et al. (2014).

### 5.3. SED-fitting Results and Comparison between the Codes

The two codes are applied to the photometric catalog presented in Section 4. Before running the codes, we correct the photometry in the catalog for the galactic extinction. We employ the dust maps by Lenz et al. (2017) and the extinction law by Fitzpatrick & Massa (2007). This correction is performed through the python libraries DUSTMAPS (Green 2018) and EXTINCTION (Barbary 2016b). To account for the possible biases in the photometry extraction (see Section 4.2), to account for the known underestimation of the photometric uncertainties by the PHOTUTILS library employed in PHOEBE (see, e.g., Leauthaud et al. 2007; Laigle et al. 2016; Weaver et al. 2022), and to allow the SED-fitting codes to explore a wider region in the photometry space, we add in quadrature 0.15 mag to the photometric uncertainties included in the catalog for the optical/NIR/MIR bands, following Laigle et al. (2016), Weaver et al. (2022). The outputs of the SED-fitting codes MAGPHYS and CIGALE are summarized in Figure 4, while the median values of the photo- $z$ s and the main physical properties for both codes are reported in Table 2. In generating the histograms and in computing the medians, we exclude from the sample 57 sources that could host an AGN (see the discussion in Section 6), in order to avoid possible biases coming from the SED fitting performed with templates not including nuclear activity. Similarly, we do not include in our analysis 10 sources with no other robust detections than in the radio (the so-called “type 0” of Behiri et al. 2023; see that study for the possible models of these sources) for which the properties estimated through SED fitting would be highly unreliable. The good convergence of the SED-fitting procedure is ensured by the median reduced  $\chi^2 < 2.5$  obtained by both codes on the analyzed sample. Since, for our galaxies, the constraints on stellar population in the optical/NIR regimes are limited—in the majority of the cases—to upper limits, we compute the SFR starting from the infrared luminosity through the relation by Kennicutt & Evans (2012) rescaled to a Chabrier (2003) IMF. This quantity is more robustly estimated, since





**Figure 3.** Some examples of SEDs of high- $z$  candidates fitted with MAGPHYS (blue solid lines; da Cunha et al. 2008; Battisti et al. 2019) and CIGALE (orange solid lines; Boquien et al. 2019). The photometry is represented by the black dots (for the detections at  $S/N > 3$ ) and the black arrows (for the upper limits). The yellow boxes report the photo- $z$ s computed with the two codes and the relative uncertainties.

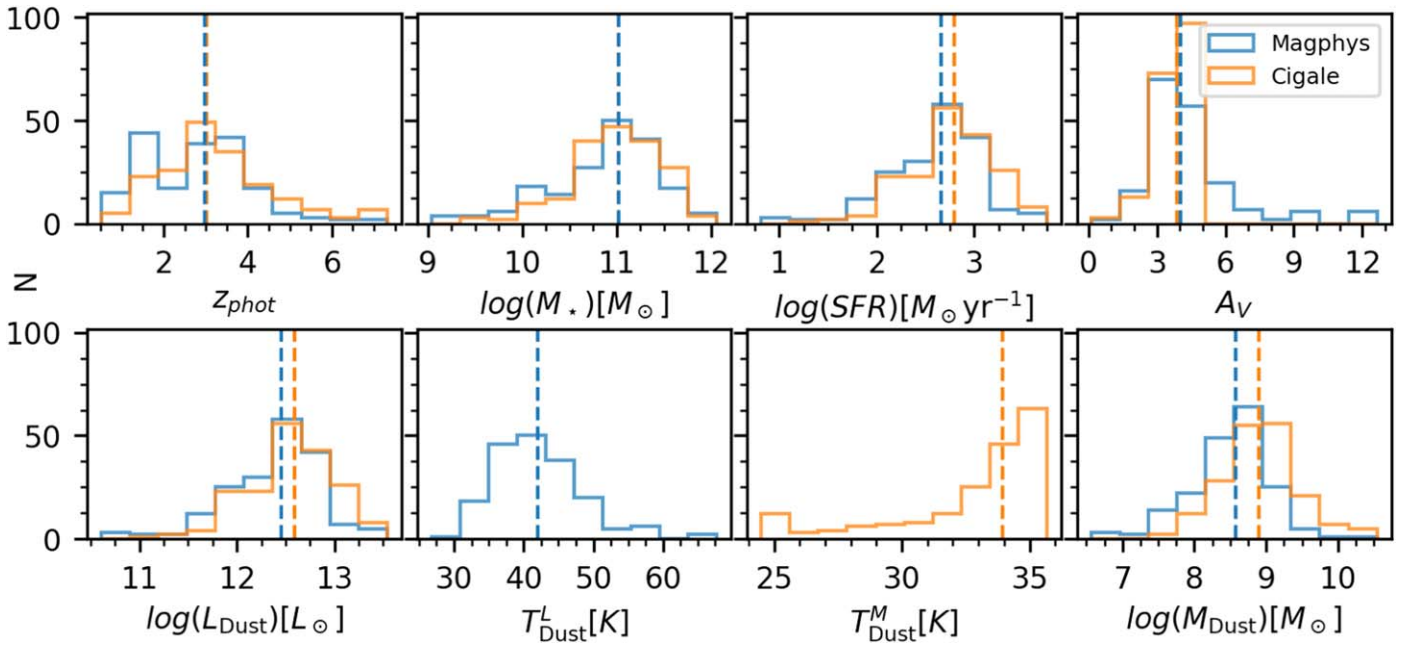
$\sim 60\%$  of the galaxies in the sample have a detection at FIR/(sub)millimeter wavelengths. However, we must underline how this quantity is more unconstrained for the other  $\sim 40\%$  of the sources in the sample without these detections. Nevertheless, since the two codes employed in the SED-fitting procedure rely on the energy balance and include the radio fluxes, some constraints on the infrared luminosity can also come by the modeled dust attenuation and by the radio luminosity (da Cunha et al. 2015; Battisti et al. 2019; Boquien et al. 2019). Finally, we notice that no significant discrepancy is visible when comparing the SFR estimated through the  $L_{\text{IR}}$  and the  $L_{1.4 \text{ GHz}}$  (see, e.g., Kennicutt & Evans 2012; Novak et al. 2017) for the galaxies with FIR/(sub)millimeter detections and those without. This result can be partly explained by the fact that MAGPHYS assumes a constant radio–infrared correlation with a  $q_{\text{TIR}}$  centered on 2.34 and with a  $1\sigma$  dispersion of 0.25 (da Cunha et al. 2015; Battisti et al. 2019). This value is broadly compatible with the median  $q_{\text{TIR}} = 2.1 \pm 0.3$  computed for the galaxies in our sample (see Section 6.3).

As it can be seen, the results of the two software for what concerns the photo- $z$ s, and most of the physical properties, are broadly compatible. The slight difference between the outputs for these quantities can be explained by some minor differences between the codes (e.g., the treatment of the upper limits, representing the majority of the constraints in the optical/NIR regime for our galaxies; see, e.g., Battisti et al. 2019; Boquien et al. 2019). Major differences hold for the dust mass and temperature. As discussed in Sections 5.1 and 5.2, MAGPHYS and CIGALE report the luminosity- ( $T_{\text{Dust}}^L$ ) and mass-weighted ( $T_{\text{Dust}}^M$ ) dust temperature, respectively. Since these quantities weight differently the hot and cold components of the dust (see, e.g., Liang et al. 2019; Sommovigo et al. 2020), it is not surprising the difference in the two distributions reported in Figure 4. Moreover, this difference in the two codes can also explain the discrepancy in the two estimates of the dust mass, being this quantity computed starting from the dust temperature. In the following, for consistency with the previous papers of the series (Talia et al. 2021; Behiri et al. 2023), we will consider the results obtained by MAGPHYS. A final note concerns the accuracy of our photo- $z$ s. Unfortunately, due to the elusive nature of the sources in our sample, it is not easy to

retrieve spectroscopic redshifts for our RS-NIRdark galaxies from the current literature. One spectroscopic redshift can be found in Jin et al. (2022), where they targeted one of our galaxies with a spectral scan performed with the ALMA and NOEMA interferometers. For our galaxy RSN-436 (ID 3117 in Jin et al. 2022), they obtained  $z_{\text{spec}} = 3.545$ , which is in good agreement with our estimates ( $z_M = 3.3 \pm 0.3$ , and  $z_C = 3.4 \pm 0.2$ ). To increase the spectroscopic coverage of our sample, we already have planned approved spectroscopic follow-ups with (sub)millimeter facilities such as ALMA and NOEMA (Gentile et al. 2023).

## 6. AGN Contribution

While in Section 1 we discussed the possible biases affecting the FIR and (sub)millimeter selection of DSFGs, this section focuses on the main bias possibly affecting our radio selection: the presence of strong AGNs. Since both star formation and radio activity can generate radio emission, we cannot exclude that some of our galaxies are hosting an AGN. For instance, Bonzini et al. (2013), Novak et al. (2017), and Enia et al. (2022), analyzing samples of radio-bright galaxies selected at 1.4 GHz, reported significant fractions of AGN in their catalogs, spanning from 10% (for the faintest radio fluxes) to 100% (for the radio-brightest sources). Fortunately, our selection procedure focuses on galaxies without a significant optical/NIR counterpart. This step allows us to remove from the sample all the brightest AGN. This selection, however, does not allow us to remove from the sample the obscured AGN, where a significant amount of dust absorbs the optical/NIR emission (see, e.g., the review by Hickox & Alexander 2018 and references therein). The presence of these sources in our sample can be unveiled by searching for AGN tracers at longer wavelengths (namely the IR and radio regimes), or in the X-ray, taking advantage of the broad wavelength coverage of our sample. Estimating the fraction of obscured AGN and properly accounting for their presence are crucial for two main reasons. First, the presence of an extra IR component due to a dusty torus surrounding the AGN and/or a radio excess due to nuclear activity can bias the SFR when obtained from the IR/radio luminosity through the relation by Kennicutt & Evans (2012). Similarly, the employment of galaxy-only templates



**Figure 4.** Distribution of the photo- $z$ s and of the main physical properties (in order; stellar mass, star formation rate, extinction, dust luminosity, dust temperature—and dust mass) of the RS-NIRdark as estimated by SED fitting with MAGPHYS+PHOTO- $z$  (da Cunha et al. 2015; Battisti et al. 2019) and CIGALE (Boquien et al. 2019) in blue and orange, respectively. The dashed lines of the same colors mark the median values of the distributions. Further details are given in Section 5.

**Table 2**

Comparison between the Median Properties Estimated by MAGPHYS and CIGALE

| Property                | MAGPHYS          |          | CIGALE           |          | Unit                     |
|-------------------------|------------------|----------|------------------|----------|--------------------------|
|                         | Median           | $\sigma$ | Median           | $\sigma$ |                          |
| $z_{\text{phot}}$       | $2.96 \pm 0.04$  | 1.2      | $3.02 \pm 0.04$  | 1.3      | ...                      |
| $\log(M_*)$             | $11.01 \pm 0.02$ | 0.61     | $11.01 \pm 0.02$ | 0.44     | $M_\odot$                |
| $\log(\text{SFR})$      | $2.67 \pm 0.02$  | 0.48     | $2.79 \pm 0.04$  | 0.45     | $M_\odot \text{yr}^{-1}$ |
| $A_V$                   | $4.01 \pm 0.04$  | 1.3      | $3.89 \pm 0.05$  | 0.31     | mag                      |
| $\log(L_{\text{Dust}})$ | $12.45 \pm 0.02$ | 0.48     | $12.59 \pm 0.02$ | 0.44     | $L_\odot$                |
| $T_{\text{Dust}}^L$     | $42.05 \pm 0.2$  | 5.8      | ...              | ...      | K                        |
| $T_{\text{Dust}}^M$     | ...              | ...      | $33.9 \pm 0.2$   | 2.7      | K                        |
| $\log(M_{\text{Dust}})$ | $8.57 \pm 0.02$  | 0.5      | $8.9 \pm 0.02$   | 0.46     | $M_\odot$                |

**Note.** The uncertainties on the median properties are estimated as the median absolute deviation:  $\text{MAD} = 1.482 \times \text{median}(|x_i - \text{median}(x_i)|)$  (Hoaglin et al. 1983) divided by  $\sqrt{N}$ , where  $N$  is the number of galaxies in the sample. For each quantity, we also report the dispersion computed as half the symmetrized interval between the 16th and the 84th percentiles.

could bias all the other physical properties (and the photo- $z$ s) estimated through SED fitting. Second, the contamination affecting the selection procedure has to be considered in the determination of the statistical properties of the sample (e.g., the luminosity function and the contribution to the SFRD).

### 6.1. Visual Inspection

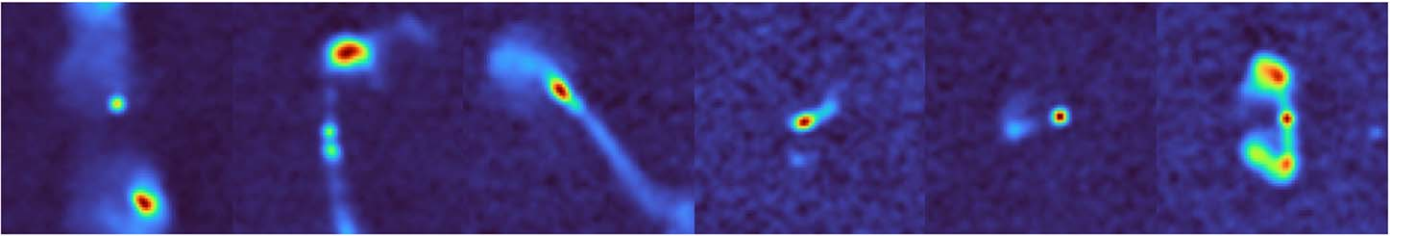
The first selection of galaxies hosting AGN in our sample is performed through a visual inspection of the 3 GHz radio maps. This procedure aims to select all the galaxies with radio morphologies generally associated with AGN (e.g., radio blobs and relativistic jets). This visual inspection allows us to mark as likely AGN 17 galaxies. Some examples of these sources are reported in Figure 5. We underline that—although this

selection is 100% pure (since nonactive galaxies do not have these peculiar radio morphologies)—it is rather far from complete since most of the galaxies in the sample are unresolved at the  $0''.7$  resolution of the 3 GHz maps by Šmolčić et al. (2017). Moreover, 9 RS-NIRdark galaxies are also part of the catalog by Vardoulaki et al. (2021) containing AGN in the COSMOS field selected through visual inspection thanks to their morphology in the same radio maps employed for our selection. Cross-matching our sample with this catalog, we find out that all the 9 sources are classified by AGN by both the selections. These results strengthen the reliability of our classification.

### 6.2. X-Ray Stacking

Another test to unveil the presence of AGN in our sample can be performed in the X-ray regime. We conduct two complementary tests to characterize the sample of RS-NIRdark galaxies as a whole and the individual sources.

1. *Sample characterization.* The first test is based on the median X-ray flux computed on all the galaxies in our sample and on the possibility of explaining it by only invoking the star formation. We follow this procedure:
  - (a) We extract the X-ray flux for each RS-NIRdark galaxy through aperture photometry on the event file in the range  $[0.5, 7]$  keV produced by the Chandra telescope in the C-COSMOS (Elvis et al. 2009) and COSMOS legacy (Civano et al. 2016) surveys. We perform the extraction through the DMEXTRACT function of the CIAO library (Fruscione et al. 2006), employing circular apertures centered on the radio position of each source. We choose the radius of each aperture by assuming that each source is not resolved (a reasonable hypothesis for possible AGN) and employing the radius used in Elvis et al. (2009), Civano et al. (2016)



**Figure 5.** Some examples of the sources marked as likely AGN in the final catalog because of their morphology during the visual inspection of the radio maps at 3 GHz. The postage has a  $15''$  side. Further details are given in Section 6.

to extract the flux of the X-ray source closest to the considered galaxy. More in detail, we employ the median radius (considering all the observations in which the source fell) corresponding to an area encompassing 90% of the PSF at the X-ray closest source. The local background is computed and subtracted from each galaxy through an annulus with an outer radius 1.5 times the circular aperture.

- (b) Once the net counts are obtained, these are converted to X-ray luminosities through the PIMMS<sup>17</sup> software by assuming a power-law model with a slope of  $\Gamma = 1.8$  (i.e., that expected for possible AGN), and a galactic  $n_{\text{H}} = 1.7 \times 10^{20} \text{ cm}^{-2}$  (Civano et al. 2016).
- (c) Finally, we convert the luminosities in fluxes by assuming the photometric redshifts computed by MAGPHYS (Section 5) and these in SFR through the empiric relation by Ranalli et al. (2003).

Considering the median SFR obtained by these X-ray fluxes, we obtain a value of  $\log(\text{SFR}_{\text{X}}) = (2.24 \pm 0.02) M_{\odot} \text{ yr}^{-1}$ , slightly lower than that obtained through the FIR flux  $\log(\text{SFR}_{\text{IR}}) = (2.38 \pm 0.02) M_{\odot} \text{ yr}^{-1}$ . This result confirms the lack of strong unobscured AGN activity in our sample.

2. *Source characterization.* If the analysis performed in the previous paragraph ensures that the bulk of the sample of RS-NIRdark is composed of SFGs, it does not give insights on the presence of single AGN within the total sample. To address this point, we perform an additional analysis. We cross-match the two catalogs of X-ray sources in the COSMOS field by Elvis et al. (2009), Civano et al. (2016) with our sample by employing as matching radii the positional uncertainty included in these catalogs (for the X-ray sources) and  $0''.7$  (for the NIRdark galaxies). Given the relatively shallow depth of the X-ray coverage in the COSMOS field, all these sources have an X-ray luminosity higher than  $10^{42} \text{ erg s}^{-1}$ ; therefore, we can assume that their X-ray flux is largely due to the presence of an AGN (see, e.g., Hickox & Alexander 2018). Once the list of RS-NIRdark galaxies with a possible X-ray counterpart is obtained, we visually inspect the NIR and radio maps at 3 GHz. We obtain two main cases:

3. The positional uncertainty on the X-ray source includes only the RS-NIRdark galaxy. In this case, we can safely assume that the X-ray signal is produced by the galaxy in our sample, suggesting the presence of an AGN in that galaxy. The 3 sources in this class are marked in the photometric catalog with an appropriate flag.
4. The positional uncertainty of the X-ray source includes both an RS-NIRdark galaxy and an NIR-bright galaxy. In this case, we cannot unambiguously associate the X-ray

signal to one of these galaxies. The 15 NIR-dark galaxies in this class are marked in the catalog with a different flag accounting for the possibility of hosting an AGN.

### 6.3. $q_{\text{TIR}}$ Analysis

A standard method to identify AGN relies on the so-called infrared–radio correlation. It is well established that the radio luminosity measured at 1.4 GHz and the infrared luminosity measured in the range  $[8; 1000] \mu\text{m}$  are tightly ( $\sigma \sim 0.16$  dex; e.g., Molnár et al. 2021) correlated in star-forming galaxies (see, e.g., de Jong et al. 1985; Helou et al. 1985). This correlation is generally measured through the  $q_{\text{TIR}}$  parameter defined as follows (see, e.g., Helou et al. 1985; Yun et al. 2001):

$$q_{\text{TIR}} = \log\left(\frac{L_{\text{IR}}[\text{W}]}{3.75 \times 10^{12} \text{ Hz}}\right) - \log\left(\frac{L_{1.4 \text{ GHz}}}{\text{W Hz}^{-1}}\right) \quad (3)$$

and mainly arises because of the connection between the star formation, the radio synchrotron emission in star-forming regions, and the thermal FIR emission of dust in the same regions. Several authors studied the possible evolution of the  $q_{\text{TIR}}$  parameter with cosmic time or possible correlations with other physical quantities. In this work, we explore two of the main studies on this point. Delhaize et al. (2017) found a possible evolution with the redshift through the relation

$$q_{\text{TIR}}(z) = (2.88 \pm 0.03) \times (1 + z)^{(-0.19 \pm 0.01)}. \quad (4)$$

On the contrary, more recently, Delvecchio et al. (2021) suggested a  $q_{\text{TIR}}$  almost constant with the redshift, but strongly dependent on the stellar mass of the hosting galaxy through the relation

$$q_{\text{TIR}}(M_{\star}, z) = (2.646 \pm 0.024) \times A^{(-0.023 \pm 0.008)} + -B \times (0.148 \pm 0.013), \quad (5)$$

where  $A = (1 + z)$ , and  $B = (\log(M/M_{\odot}) - 10)$ .

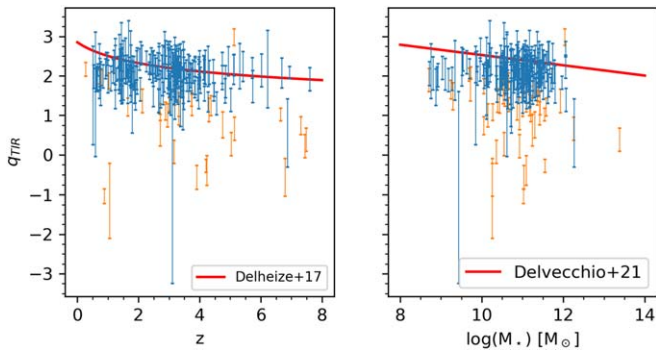
We compute the radio luminosity at 1.4 GHz for the galaxies with a counterpart in the 1.4 GHz catalog by Schinnerer et al. (2010) through the relation

$$L_{1.4 \text{ GHz}} = \frac{4\pi D_L^2(z)}{(1 + z)^{1+\alpha}} S_{1.4 \text{ GHz}} \quad (6)$$

by employing the photometric redshifts estimated by MAGPHYS and the spectral slope  $\alpha$  computed through the flux densities reported in the 3 and 1.4 GHz catalogs. For the galaxies without a counterpart in the 1.4 GHz sample, we evaluate the 1.4 GHz flux starting from the 3 GHz flux density through the

<sup>17</sup> <https://heasarc.gsfc.nasa.gov/docs/software/tools/pimms.html>





**Figure 6.** Behavior of the  $q_{\text{TIR}}$  as a function of the redshift (left panel) and stellar mass (right panel). The blue solid lines show the relation by Delhaize et al. (2017), Delvecchio et al. (2021), respectively. The objects distant more than  $3\sigma$  from each relation are highlighted in red.

relation

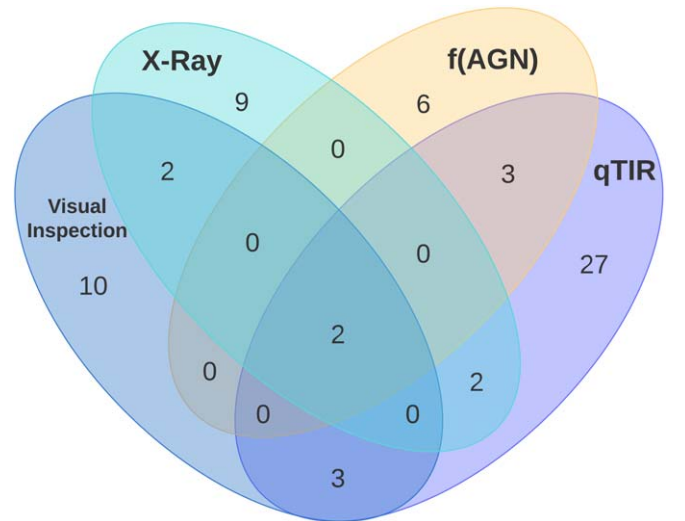
$$L_{1.4 \text{ GHz}} = \frac{4\pi D_L^2(z)}{(1+z)^{1+\alpha}} \left( \frac{1.4 \text{ GHz}}{3 \text{ GHz}} \right)^\alpha S_{3 \text{ GHz}}, \quad (7)$$

and assuming a spectral slope of  $\alpha = -0.7$  (generally considered for star-forming galaxies; see, e.g., Novak et al. 2017).

Once we applied Equations (4) and (5) to our sample (Figure 6), marking as likely AGN the sources distant more than  $3\sigma$  from the relations (see, e.g., Delvecchio et al. 2017; Enia et al. 2022), we obtain that 40 sources are classified as AGN following Delhaize et al. (2017), and 57 follow Delvecchio et al. (2021). We include the results from both the tests in the final catalog. However, given the high uncertainties affecting our photo- $z$ s and our stellar masses, in the following, we consider as likely AGNs only the 37 sources classified as AGN according to both the relations. A final interesting remark concerns the overall distribution of the  $q_{\text{TIR}}$  as a function of the photometric redshift. As visible in Figure 6, the bulk of the population of the RS-NIRdark galaxies have a median  $q_{\text{TIR}}$  compatible with that expected at the median redshift of the sample from the relation by Delhaize et al. (2017) ( $2.1 \pm 0.3$  and  $2.21 \pm 0.05$ , respectively). However, the distribution of these values at low redshift appears to significantly differ from that expected from the relation. This discrepancy can be explained with the different selection operated in this study with respect to Delhaize et al. (2017). In particular, our selection of NIR-dark sources is expected to produce a sample of more extremely obscured sources at low- $z$  to account for the extinction of redder rest-frame bands. Therefore, we do not expect in this regime the properties of our (incomplete) sample to completely resemble those of the total population of radio-selected galaxies analyzed in Delhaize et al. (2017).

#### 6.4. SED Decomposition

A further sign of AGN hosted in our galaxies can be found at MIR wavelengths. It arises from the presence of a dusty torus surrounding the supermassive black hole and heated up from the high-energy radiation coming from there. Trying to model the SED with a template not accounting for this additional component generally produces a best-fitting SED underestimating all the MIR fluxes (see, e.g., Hickox & Alexander 2018). This problem can be solved by adding a torus component to the templates fitted to the galaxy photometry. We perform this test with CIGALE (MAGPHYS does not allow the addition of an AGN component). The overall



**Figure 7.** Diagram reporting the number of likely AGN unveiled by the different methods described in Section 6 and the relative overlap. Further details in Section 6.

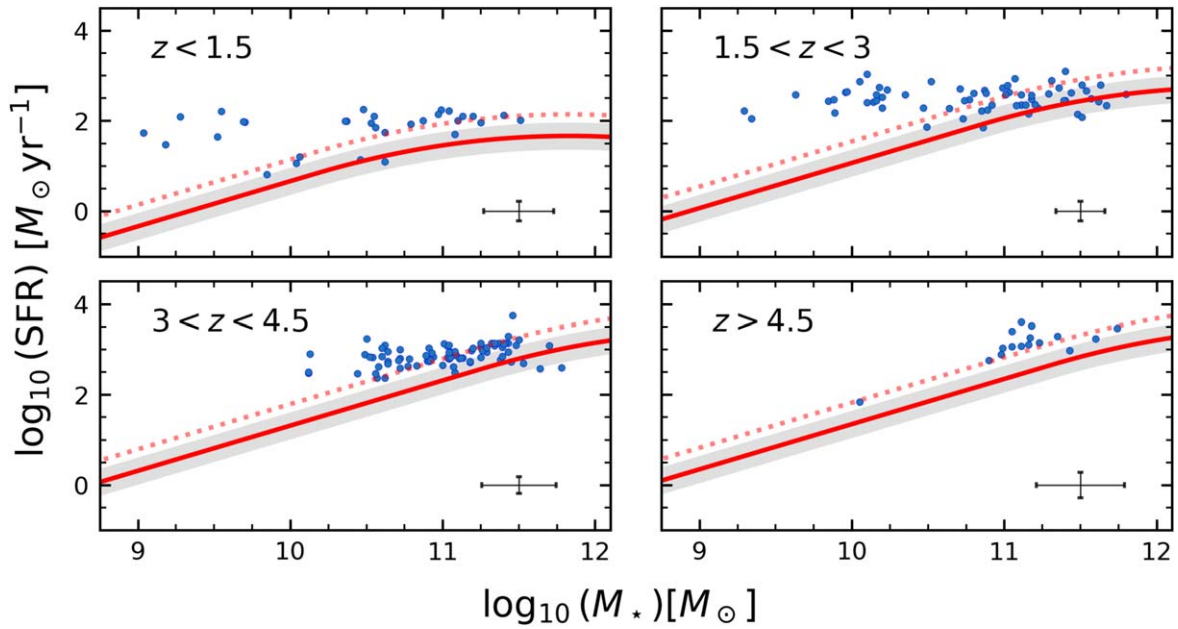
setup is the same as discussed in the Section 5.2, but adding the dusty torus component as modeled by Fritz et al. (2006). The model's parameters are the same as employed in Donevski et al. (2020) in modeling a sample of likely DSFGs. The main parameter describing the effect of the AGN on the modeled SED is the AGN fraction  $f_{\text{AGN}}$ , defined as the ratio between the torus luminosity and the dust luminosity in the range  $[5, 40] \mu\text{m}$ . Defining as likely AGN all the galaxies with a  $f_{\text{AGN}} > 10\%$ , we mark 11 sources among the entire sample of RS-NIRdark galaxies. We underline that the median value of  $f_{\text{AGN}}$  computed on the whole sample is compatible with zero, suggesting an overall small contribution of AGN in the RS-NIRdark galaxies. However, we underline that—due to the limited coverage of the MIR regime in our sample—also this estimation should be considered as a lower limit on the actual AGN contribution in our galaxies.

#### 6.5. Final Remarks on AGN Contamination

The sample of likely AGN reported by the different methods has some overlap, as shown in Figure 7. Considering all the galaxies marked as possible AGN by at least one method, we obtain a sample of 64 sources ( $\sim 23\%$  of the full sample). Excluding the sources with an uncertain X-ray counterpart (i.e., those at point 2 in Section 6.2) and with no other indication of AGN activity from the other methods, we obtain 57 sources ( $\sim 20\%$  of the sample).

As a final remark, we can compare the estimated fraction of RS-NIRdark galaxies hosting an AGN with a theoretical prediction based on analogous works present in the literature. Bonzini et al. (2013), Novak et al. (2018), working on radio-selected catalogs of galaxies, reported in their work the fraction of AGN in the analysis of their sources as a function of the radio flux at 1.4 GHz. Considering these relations and integrating them on our<sup>18</sup> flux distribution at 1.4 GHz, we estimated an expected fraction of AGN in our sample lower than 28.3% (further divided in 27% of radio-quiet AGN and 1.3% of radio-loud AGN). We underline that this last

<sup>18</sup> Here, we employ the radio fluxes at 1.4 GHz provided by Schinnerer et al. (2010). For the sources without a counterpart, we conservatively assume a flux equal to the sensitivity of the survey.



**Figure 8.** Comparison between the SFRs and stellar masses of the RS-NIRdark galaxies (as computed by MAGPHYS) and those expected from the main sequence of the star-forming galaxies (red solid line; Schreiber et al. 2015; rescaled to a Chabrier 2003 IMF). The gray shaded area represents the  $1\sigma = 0.3$  dex scatter, while the red dotted line reports our threshold for selecting starburst galaxies (i.e., 3 times the SFR expected from MS galaxies). It is possible to notice how the vast majority of the galaxies in the sample lie above the main-sequence line. The median uncertainty on the SFR and on the stellar mass is reported in the lower corner of each panel. Further details are given in Section 7.1.

estimation represents an upper limit on the expected number of AGN in our sample, since we expect that focusing on radio sources without an optical/NIR counterpart contributes to exclude a significant fraction of these objects.

## 7. Discussion

### 7.1. Analysis of the Physical Properties

The results, shown in Figure 4 and Table 2, allow us to infer some general properties of the RS-NIRdark galaxies:

1. The high value of the median stellar attenuation  $A_V \sim 3.5$  confirms the initial hypothesis that our selection can provide a sample of highly obscured galaxies.
2. The high value of the infrared luminosity ( $L_{\text{IR}} > 10^{12} L_{\odot}$ ) and—hence—the high median SFR ( $\sim 300 M_{\odot} \text{yr}^{-1}$ ) suggests that our galaxies are actively forming stars (as expected from a radio selection). To determine the nature of the RS-NIRdark galaxies, we need to compare the SFRs obtained through the SED fitting with those expected from main-sequence galaxies in the same redshift range (e.g., Schreiber et al. 2015). As shown in Figure 8, most galaxies lie above the main-sequence line at all the redshifts. Moreover, when computing the  $\text{SFR}/\text{SFR}_{\text{MS}}$  (i.e., the ratio between the SFR and that expected from a main-sequence galaxy of the same mass and in the same redshift bin), we obtain that more than 50% of the RS-NIRdark galaxies have an  $\text{SFR}/\text{SFR}_{\text{MS}} > 3$ , being candidate starburst galaxies.
3. An additional result, strictly connected to the previous one, concerns the high median stellar mass estimated through SED fitting ( $M_{\star} \sim 10^{11} M_{\odot}$ ). This quantity—albeit quite uncertain due to the weak constraints in the optical/NIR wavelengths—suggests that the RS-NIRdark galaxies are a population of massive star-forming galaxies, with a high- $z$  tail suitable for playing a

significant role in the evolution of the massive and passive galaxies at  $z \sim 3$  (see, e.g., Straatman et al. 2014; Schreiber et al. 2018b; Valentino et al. 2020).

All these findings, once combined, confirm the initial hypothesis that the RS-NIRdark galaxies represent a significant population of high- $z$  DSFGs.

### 7.2. High- $z$ Tail

The redshift distribution of the RS-NIRdark galaxies is worth a deeper discussion. The median redshift around  $z \sim 3$  tells us that we are looking at a population whose bulk is located at the so-called *cosmic noon*. However, the presence of a significant tail of high- $z$  sources (namely, 99 galaxies at  $z > 3$  and 17 galaxies at  $z > 4.5$ , once excluded the possible AGNs selected in Section 6) can provide some insights on the possible evolutionary path of these sources.

The main result concerns the number density of these sources. We compute this quantity through the  $V_{\text{max}}$  method (Schmidt 1968), considering the galaxies located at  $z > 3.5$ , and that could—therefore—play a role in the evolution of the massive galaxies at  $z \sim 3.5$ . We account for the uncertainties in the photo- $z$ s and in the radio fluxes through a Monte Carlo integration. Specifically, we perform a large number of realizations ( $\sim 500$ ) of the total number density, sampling each time and for each source a couple of values for the redshift and for the radio flux from their distribution (namely, the  $p(z)$  computed by MAGPHYS and the values and uncertainties reported in Smolčić et al. 2017). At the end of this procedure, we consider as our number density the median value of the distribution and as its uncertainty the symmetrized interval between the 16th and the 84th percentiles. We obtain a number density of  $n = (3.3 \pm 0.9) \times 10^{-6} \text{ Mpc}^{-3}$  for the galaxies at  $z > 3.5$ . This quantity is by a factor 6 lower than those computed by Straatman et al. (2014), Schreiber et al. (2018b),

and Valentino et al. (2020) for the passive and massive galaxies at  $z \sim 3.5$  ( $\sim 2 \times 10^{-5} \text{ Mpc}^{-3}$ ). It is important to notice—however—that our estimation must be considered as a lower limit on the actual number density of the RS-NIRdark galaxies, since it is not corrected for the expected *duty cycle* of the galaxies and for the incompleteness of the selection. This issue will be discussed in detail in the forthcoming paper of the series (F. Gentile et al. 2024, in preparation). Moreover, when looking at the stellar masses of our sources, we obtain that the RS-NIRdark galaxies located at  $z > 3.5$  have a median stellar mass of  $\log(M_*) = 11.0 M_\odot$  with a  $1\sigma$  dispersion of 0.45 dex. Comparing this quantity with the expected properties of the progenitors of the massive galaxies at  $z \sim 3.5$  (see, e.g., the forecasts by Valentino et al. 2020), we can notice how the low-mass end of those objects cannot be formed by the galaxies in our sample. This result also can be used to explain the difference between the number densities of the two populations.

### 7.3. Comparison with the Literature

An interesting final point to discuss concerns the RS-NIRdark galaxies and their possible overlap with other important populations of DSFGs:

1. *H-dropout*. The first population is composed of the H-dropout galaxies (Wang et al. 2019). These galaxies are selected as H-dark sources in the CANDELS survey (the limiting magnitude at  $5\sigma$  in the  $H$  band is 27 mag; Grogin et al. 2011; Koekemoer et al. 2011) with a counterpart in the second channel of the IRAC camera from the SEDS survey ( $[4.5] < 24$  mag; 80% completeness limit; Ashby et al. 2013). When considering the full photometric catalog of 323 RS-NIRdark galaxies, we obtain that only 266 sources satisfy the selection criteria exposed in Wang et al. (2019), the others being too faint at  $4.5 \mu\text{m}$  to be selected with the cut on the  $[4.5]$  flux performed by Wang et al. (2019). This result indicates that a significant fraction of the RS-NIRdark galaxies would not be selected as H-dropout. On the contrary, when examining the 18 H-dropout galaxies selected by Wang et al. (2019) in the COSMOS field, we find that only 2 sources have a significant ( $S/N > 5.5$ ) radio counterpart at 3 GHz in the catalog by Smolčić et al. (2017). These two results, when combined, ensure that the two selections of RS-NIRdark galaxies and H-dropout are different, with just some sources belonging to both.
2. *SMGs*. The second population is composed of the so-called submillimeter galaxies (SMGs). Since this definition can be applied to all the sources detected in a (sub) millimeter survey, it strongly depends on the considered instrument’s sensitivity. To obtain results comparable with those reported by Wang et al. (2019), we consider the ALESS survey (Swinbank et al. 2014) targeting bright galaxies in the SCUBA survey (i.e., galaxies with a flux density at  $870 \mu\text{m}$   $S_{870} > 4.2$  mJy). We can estimate the (sub)millimeter flux density of the RS-NIRdark galaxies at  $870 \mu\text{m}$  through the best-fitting SED provided by MAGPHYS. In doing so, we obtain that only 76 galaxies have  $S_{870} > 4.2$  mJy, being consistent with the selection by Swinbank et al. (2014). On the other side, Thomson et al. (2014) pointed out that only  $\sim 70\%$  of SMGs are radio-bright galaxies, while Gruppioni et al. (2020)

pointed out that only a tiny percentage of SMGs are NIR-dark galaxies. As before, these results suggest that the RS-NIRdark galaxies and SMGs are two different populations of galaxies with just some sources in common.

Finally, we can compare the physical properties estimated by Wang et al. (2019), da Cunha et al. (2015) for the population of H-dropout and SMGs with those presented in this study. Figure 9 shows the comparison between the SFR and stellar mass of the RS-NIRdark galaxies in the redshift range  $3.5 < z < 4.5$  (i.e., around the median redshift  $z \sim 4$  reported by Wang et al. 2019 for the H-dropout) with those reported by da Cunha et al. (2015), Wang et al. (2019) and with the main sequence at this redshift (Schreiber et al. 2018a; rescaled to a Chabrier 2003 IMF). For completeness, we also include in the plot the sample of “NIR-faint” ( $K_s > 23.9$ ) SMGs selected in COSMOS by Smail et al. (2021).

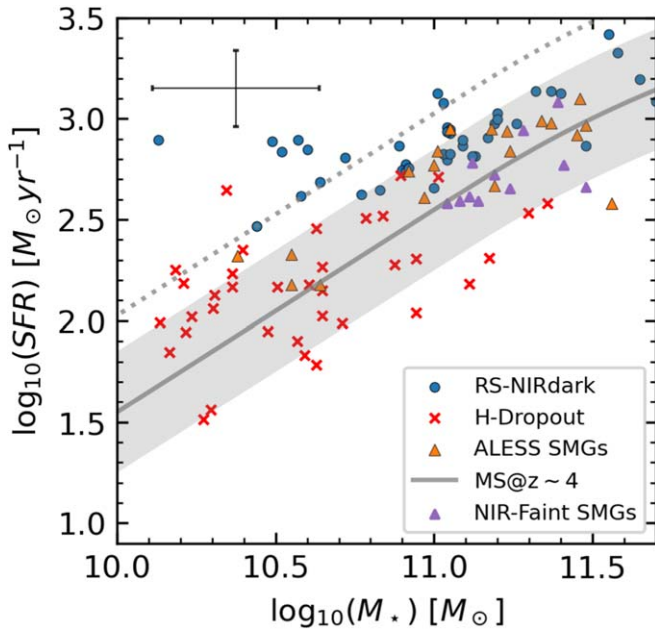
We can notice that the RS-NIRdark galaxies are—on average—more star-forming than the H-dropout ( $\Delta \log(\text{SFR}/M_\odot \text{ yr}^{-1}) \sim 0.6$ ), with a median SFR comparable with the SMGs ( $\Delta \log(\text{SFR}/M_\odot \text{ yr}^{-1}) \sim 0.02$ ), but with a significant tail of less star-forming sources. This result can be explained by the radio-selection, cutting out most of the sources below the main sequence (belonging mainly to the H-dropout). Additionally, we can see how the RS-NIRdark galaxies cover most of the mass range covered by the other selections, being on average more massive than the H-dropout and less massive than the SMGs. This result is in agreement with the discussed overlap with both of the other populations.

## 8. Summary

In this paper, we presented the first panchromatic study of 263 RS-NIRdark galaxies discovered in the COSMOS field following the selection by Talia et al. (2021). The development of a new deblending tool (Photometry Extractor for Blended Objects or PHOEB0) allowed us to extract accurate photometry in the optical-to-MIR regime, even for the sources with a close bright contaminant. This procedure, in particular, allowed us to analyze a wider sample of galaxies missed in previous studies on the RS-NIRdark galaxies in the COSMOS field (Talia et al. 2021; Behiri et al. 2023). The complete photometric catalog<sup>19</sup> has been employed to estimate the photo- $z$ s and physical properties of all the galaxies in the sample through an SED-fitting procedure performed with two complementary codes (MAGPHYS and CIGALE). The results obtained with these algorithms confirmed the initial hypothesis that the RS-NIRdark galaxies are a population of starburst DSFGs, lying above the main sequence in all the redshift bins and with a significant amount of dust absorbing their optical/NIR emission. Moreover, by studying in detail the redshift distribution of the galaxies in the sample and their number density, we obtain precious insights on the possible evolutionary path of these sources, collecting significant clues that the RS-NIRdark galaxies could play a key role in the evolution of the massive and passive galaxies discovered at  $z \sim 3$ . In addition, the analysis of the multiwavelength counterpart in all the wavelength regimes (from the X-rays to the radio) allowed us to estimate the possible AGN contribution in our sample.

<sup>19</sup> The complete catalog, together with the other materials supporting the findings of this study, is available on the website of the collaboration: <https://sites.google.com/inaf.it/rsnirdark/>.





**Figure 9.** Comparison between the SFRs and stellar masses of several populations of DSFGs. The different symbols show the RS-NIRdark galaxies (blue dots), the H-dropout (Wang et al. 2019; red crosses), the SMGs found in ALESS (da Cunha et al. 2015; orange triangles), and the NIR-faint SMGs discovered by Smail et al. (2021) in the UKIRT Infrared Deep Sky Survey (UKIDSS; purple triangles). The solid gray line represents the main sequence at  $z \sim 4$  (Schreiber et al. 2015; rescaled to a Chabrier 2003 IMF), and the gray shaded area its  $1\sigma = 0.3$  dex scatter. The dotted gray line reports our threshold to define starburst galaxies (i.e., 3 times the SFR of a main-sequence galaxy). The median uncertainty on the SFR and on the stellar mass is reported in the lower corner of each panel. For consistency with the cited studies, we only report galaxies with  $3.5 < z < 4.5$ . Further details are given in Section 7.1.

Finally, through a comparison with other populations of DSFGs, we confirmed that the radio-selection produces a population of galaxies with different physical properties with respect to the SMGs (e.g., da Cunha et al. 2015) and the H-dropout (Wang et al. 2019).

The catalog presented in this work will be employed in the next papers of the series to derive the luminosity function of the RS-NIRdark galaxies, to estimate their contribution to the cosmic SFRD, and to study their role in the evolution of the massive galaxies. In the near future, the results presented in this work will be reinforced by the plethora of new data coming from state-of-the-art facilities. About half of the RS-NIRdark galaxies will be observed by the James Webb Space Telescope as a part of the COSMOS-Web survey (Casey et al. 2023). Moreover, a first spectral analysis of a pilot sample of these galaxies performed with ALMA will allow a robust determination of the spectroscopic redshifts (Gentile et al. 2023).

### Acknowledgments

We warmly thank the anonymous referee for their comments on the first version of this paper, which really allowed us to increase the overall quality of our study. F.G. and M.T. thank Eleni Vardoulaki for sharing her sample of AGN in COSMOS. F.G. and M.T. thank Ian Smail for sharing his sample of NIR-faint SMGs in COSMOS. F.G. thanks Alberto Traina for the valuable support on the SED fitting. F.G., M.T., A.L., M.M., and A.C. acknowledge the support from grant PRIN MIUR 2017-20173ML3WW\_001, “Opening the ALMA window on the cosmic evolution of gas, stars, and supermassive black

holes.” A.L. is supported by the EU H2020-MSCA-ITN-2019 project 860744 “BiD4BEST: Big Data applications for black hole Evolution STudies.” M.V. acknowledges financial support from the Inter-University Institute for Data Intensive Astronomy (IDIA), a partnership of the University of Cape Town, the University of Pretoria, the University of the Western Cape and the South African Radio Astronomy Observatory, and from the South African Department of Science and Innovation’s National Research Foundation under the ISARP RADIO-SKY2020 Joint Research Scheme (DSI-NRF grant No. 113121) and the CSUR HIPPO Project (DSI-NRF grant No. 121291).

## Appendix A

### Comparison with Talia+21 and Behiri+23

As previously discussed in Section 1, this paper continues the previous studies on the RS-NIRdark galaxies in the COSMOS field started by Talia et al. (2021), Behiri et al. (2023). Here, we discuss the main differences between this work and the previous ones. The main upgrade of this work concerns the selection of the sample of RS-NIRdark galaxies. First, by cross-matching the radio catalog with the NIR-selected COSMOS2020 (in place of the COSMOS2015 employed by Talia et al. 2021; Behiri et al. 2023), we have been able to exclude from the sample 153 NIR-faint galaxies missed in the previous version of the COSMOS catalog. These galaxies were undetected in the COSMOS2015 because of the brighter limiting magnitude reached in the NIR bands ( $\sim 1$  mag in the  $K_s$  band) and in the shallower  $\chi^2$ -image employed as the detection image (see Laigle et al. 2016; Weaver et al. 2022). Second, both Talia et al. (2021), Behiri et al. (2023) employed an additional selection step after the three discussed in this paper (Section 2). These step were aimed to only analyze the sources without contamination from a nearby source. More in detail, Talia et al. (2021) removed from the sample all the galaxies with a nearby source whose  $3\sigma$  isophote in the  $\chi^2$ -image overlapped the  $3\sigma$  isophote in the radio map at 3 GHz. Thanks to this selection, Talia et al. (2021) analyzed a subsample of the full catalog composed by 197 galaxies. Similarly, Behiri et al. (2023) slightly enlarged the sample by including 75 “slightly contaminated” sources (i.e., galaxy with an NIR-bright contaminant overlapping the  $5\sigma$  radio isophote). Differently from these works, in this paper, we developed an ad hoc pipeline (PHOEBOS; see Section 4) to extract the photometry of all the galaxies, regardless of the possible blending with nearby contaminants. Other important differences are as follows:

1. Differently from Talia et al. (2021), Behiri et al. (2023), in this paper, we employ the IRAC maps from the DAWN survey (Section 5.1). These maps are significantly deeper than those (from the SPLASH survey<sup>20</sup>) used in the previous studies, allowing us to obtain more detection in the IRAC bands for our galaxies and put more stringent constraints on the upper limits for the IRAC-undetected sources. Similarly, the flux densities in the (sub)millimeter regime are obtained from the most up-to-date version of the A3COSMOS catalog, reflecting in a more significant number of counterparts for our sample of galaxies.

<sup>20</sup> <https://splash.caltech.edu/index.html>

**Table 3**

Comparison between the Median Properties Obtained through MAGPHYS in This Paper and in the Previous Works Based on Subsamples of the RS-NIRdark Galaxies (Talia et al. 2021; Behiri et al. 2023)

| Property                | This Work        | Talia+21         | Behiri+23        | Unit                        |
|-------------------------|------------------|------------------|------------------|-----------------------------|
| $\bar{z}_{\text{phot}}$ | $2.96 \pm 0.04$  | $3.1 \pm 0.1$    | $3.3 \pm 0.2$    | ...                         |
| $\log(M_*)$             | $11.01 \pm 0.02$ | $11.3 \pm 0.02$  | $11.1 \pm 0.02$  | $M_{\odot}$                 |
| $\log(\text{SFR})$      | $2.67 \pm 0.02$  | $2.48 \pm 0.02$  | $2.79 \pm 0.02$  | $M_{\odot} \text{ yr}^{-1}$ |
| $A_v$                   | $4.01 \pm 0.04$  | $4.2 \pm 0.1$    | $4.0 \pm 0.03$   | mag                         |
| $\log(L_{\text{Dust}})$ | $12.45 \pm 0.02$ | $12.50 \pm 0.02$ | $12.93 \pm 0.02$ | $L_{\odot}$                 |
| $T_{\text{Dust}}^L$     | $42.0 \pm 0.2$   | $39.5 \pm 0.7$   | ...              | K                           |
| $\log(M_{\text{Dust}})$ | $8.57 \pm 0.02$  | ...              | ...              | $M_{\odot}$                 |

2. Finally, in Talia et al. (2021), the photometry is extracted ex novo only for the IRAC bands, while the flux densities and the upper limits in the optical and NIR bands are obtained from preexisting catalogs or the average depth of the analyzed maps. In this work, the pipeline presented in Section 4 allows us to perform forced photometry on the exact position of the sources (thanks to the radio prior). This procedure (analogous to that followed by Behiri et al. 2023 for the NIR-dark galaxies) enables us to pose more stringent upper limits in the optical and NIR band, reducing the degrees of freedom in the SED fitting.

These improvements allow us to build a more complete photometric catalog employed in the SED-fitting procedure. Analyzing the results obtained with MAGPHYS (i.e., the same software used in the previous works; see Table 3), we can notice some interesting differences:

1. The median redshift of the whole sample of RS-NIRdark galaxies is slightly lower (see Table 3) than that reported in Talia et al. (2021) and Behiri et al. (2023).
2. The median stellar attenuation is slightly lower than that reported in the previous studies (Table 3). This result can be explained by the more stringent upper limits obtained by the new-deblending algorithm in the optical and NIR bands and in the MIR regime thanks to the deeper IRAC maps employed in this work.

The main difference between the results of the various studies concerns the number density of the RS-NIRdark galaxies. Following the procedure discussed in Section 7.2, we can compute this quantity for the galaxies located at  $z > 4.5$  (i.e., the highest redshift bin adopted in Talia et al. 2021). We obtain  $n = (1.1 \pm 0.6) \times 10^{-6} \text{ Mpc}^{-3}$ . This quantity is in moderate tension with that obtained by Talia et al. (2021):  $n = (5.2 \pm 1.3) \times 10^{-5} \text{ Mpc}^{-3}$ . This result can be explained by the lower number of galaxies analyzed in this study with respect to the previous one and with the lower median redshift obtained here.

## Appendix B Data Release


















The full photometric catalog of the RS-NIRdark galaxies in the COSMOS field, together with the other data supporting this study, can be freely accessed on the website of the collaboration.<sup>21</sup> More in detail, the data release consists of the following:

1. One catalog containing all the photometry from optical to radio wavelengths; the catalog has the following structure.
  - (a) *Column 1*. A progressive ID. For consistency with the previous studies by Talia et al. (2021), Behiri et al. (2023), the IDs refer to the total sample of 476 RS-NIRdark galaxies selected in the COSMOS field by Talia et al. (2021).
  - (b) *Columns 2–27*. Fluxes and photometric uncertainties obtained through the PHOEB0 for all the bands discussed in Table 1.
  - (c) *Columns 28–85*. Fluxes and photometric uncertainties at FIR/(sub)millimeter/radio wavelengths obtained through cross-matching with the preexisting catalogs listed in Section 3.2.

All the fluxes and uncertainties are reported in Jy. The catalog can be directly employed to perform the SED fitting with MAGPHYS; therefore, all the entries with an S/N < 1 are treated as upper limits (i.e., with a missing flux and with a positive uncertainty equal to the  $1\sigma$  upper limit).

2. Two catalogs containing all the properties estimated through SED fitting with MAGPHYS and CIGALE; a complete description of all the columns can be found in the FITS header of each file.
3. A catalog with the AGN classifications performed in Section 6; the description of the columns is contained in the FITS header of the file.
4. A series of probability density function files showing the best-fitting SEDs of all the RS-NIRdark galaxies as computed by MAGPHYS and CIGALE and the cutouts in most of the bands analyzed with PHOEB0.

## ORCID iDs

Fabrizio Gentile  <https://orcid.org/0000-0002-8008-9871>  
 Margherita Talia  <https://orcid.org/0000-0003-4352-2063>  
 Meriem Behiri  <https://orcid.org/0000-0002-6444-8547>  
 Giovanni Zamorani  <https://orcid.org/0000-0002-2318-301X>  
 Luigi Barchiesi  <https://orcid.org/0000-0003-3419-538X>  
 Cristian Vignali  <https://orcid.org/0000-0002-8853-9611>  
 Francesca Pozzi  <https://orcid.org/0000-0002-7412-647X>  
 Matthieu Bethermin  <https://orcid.org/0000-0002-3915-2015>  
 Andrea Enia  <https://orcid.org/0000-0002-0200-2857>  
 Andreas L. Faist  <https://orcid.org/0000-0002-9382-9832>  
 Marika Giulietti  <https://orcid.org/0000-0002-1847-4496>  
 Carlotta Gruppioni  <https://orcid.org/0000-0002-5836-4056>  
 Andrea Lapi  <https://orcid.org/0000-0002-4882-1735>  
 Marcella Massardi  <https://orcid.org/0000-0002-0375-8330>  
 Vernesa Smolčić  <https://orcid.org/0000-0002-3893-8614>  
 Mattia Vaccari  <https://orcid.org/0000-0002-6748-0577>  
 Andrea Cimatti  <https://orcid.org/0000-0002-4409-5633>

## References

- Aihara, H., AlSayyad, Y., Ando, M., et al. 2019, *PASJ*, 71, 114  
 Aniano, G., Draine, B. T., Gordon, K. D., & Sandstrom, K. 2011, *PASP*, 123, 1218  
 Aretxaga, I., Wilson, G. W., Aguilar, E., et al. 2011, *MNRAS*, 415, 3831  
 Ashby, M. L. N., Willner, S. P., Fazio, G. G., et al. 2013, *ApJ*, 769, 80  
 Barbary, K. 2016a, *JOSS*, 1, 58  
 Barbary, K. 2016b, *extinction v0.3.0*, Zenodo, doi:10.5281/zenodo.804967  
 Battisti, A. J., da Cunha, E., Grasha, K., et al. 2019, *ApJ*, 882, 61  
 Behiri, M., Talia, M., Cimatti, A., et al. 2023, *ApJ*, 957, 63

<sup>21</sup> <https://sites.google.com/inaf.it/rsnirdark/>

- Bertin, E., & Arnouts, S. 1996, *A&AS*, **117**, 393
- Bertin, E., Mellier, Y., Radovich, M., et al. 2002, in ASP Conf. Ser. 281, *Astronomical Data Analysis Software and Systems XI*, ed. D. A. Bohlender, D. Durand, & T. H. Handley (San Francisco, CA: ASP), **228**, A.D.T.H.
- Bertoldi, F., Carilli, C., Aravena, M., et al. 2007, *ApJS*, **172**, 132
- Béthermin, M., Daddi, E., Magdis, G., et al. 2015, *A&A*, **573**, A113
- Blain, A. W., & Phillips, T. G. 2002, *MNRAS*, **333**, 222
- Bonzini, M., Padovani, P., Mainieri, V., et al. 2013, *MNRAS*, **436**, 3759
- Boquien, M., Burgarella, D., Roehly, Y., et al. 2019, *A&A*, **622**, A103
- Bradley, L., Sipőcz, B., Robitaille, T., et al. 2020, *astropy/photutils: v1.0.0*, Zenodo, doi:10.5281/zenodo.4044744
- Bruzual, G., & Charlot, S. 2003, *MNRAS*, **344**, 1000
- Burgarella, D., Buat, V., Gruppioni, C., et al. 2013, *A&A*, **554**, A70
- Casey, C. M. 2012, *MNRAS*, **425**, 3094
- Casey, C. M., Kartaltepe, J. S., Drakos, N. E., et al. 2023, *ApJ*, **954**, 31
- Casey, C. M., Narayanan, D., & Cooray, A. 2014, *PhR*, **541**, 45
- Casey, C. M., Zavala, J. A., Aravena, M., et al. 2019, *ApJ*, **887**, 55
- Casey, C. M., Zavala, J. A., Manning, S. M., et al. 2021, *ApJ*, **923**, 215
- Chabrier, G. 2003, *PASP*, **115**, 763
- Charlot, S., & Fall, S. M. 2000, *ApJ*, **539**, 718
- Civano, F., Marchesi, S., Comastri, A., et al. 2016, *ApJ*, **819**, 62
- Cowie, L. L., Barger, A. J., Hsu, L. Y., et al. 2017, *ApJ*, **837**, 139
- da Cunha, E., Charlot, S., & Elbaz, D. 2008, *MNRAS*, **388**, 1595
- da Cunha, E., Walter, F., Smail, I. R., et al. 2015, *ApJ*, **806**, 110
- de Jong, T., Klein, U., Wielebinski, R., & Wunderlich, E. 1985, *A&A*, **147**, L6
- Delhaize, J., Smolčić, V., Delvecchio, I., et al. 2017, *A&A*, **602**, A4
- Delvecchio, I., Daddi, E., Sargent, M. T., et al. 2021, *A&A*, **647**, A123
- Delvecchio, I., Smolčić, V., Zamorani, G., et al. 2017, *A&A*, **602**, A3
- Donevski, D., Lapi, A., Malek, K., et al. 2020, *A&A*, **644**, A144
- Draine, B. T., Aniano, G., Krause, O., et al. 2014, *ApJ*, **780**, 172
- Draine, B. T., & Li, A. 2007, *ApJ*, **657**, 810
- Dudzevičiūtė, U., Smail, I., Swinbank, A. M., et al. 2021, *MNRAS*, **500**, 942
- Elvis, M., Civano, F., Vignali, C., et al. 2009, *ApJS*, **184**, 158
- Endsley, R., Stark, D. P., Charlot, S., et al. 2021, *MNRAS*, **502**, 6044
- Enia, A., Talia, M., Pozzi, F., et al. 2022, *ApJ*, **927**, 204
- Faisst, A. L., Capak, P. L., Yan, L., et al. 2017, *ApJ*, **847**, 21
- Fitzpatrick, E. L., & Massa, D. 2007, *ApJ*, **663**, 320
- Franco, M., Elbaz, D., Béthermin, M., et al. 2018, *A&A*, **620**, A152
- Fritz, J., Franceschini, A., & Hatziminaoglou, E. 2006, *MNRAS*, **366**, 767
- Fruscione, A., McDowell, J. C., Allen, G. E., et al. 2006, *Proc. SPIE*, **6270**, 62701V
- Geach, J. E., Dunlop, J. S., Halpern, M., et al. 2017, *MNRAS*, **465**, 1789
- Gentile, F., Talia, M., & Saponi, M. 2023, *PhoEBO: The Photometry Extractor for Blended Objects v1.0*, Zenodo, doi:10.5281/zenodo.8305556
- Giulietti, M., Massardi, M., Lapi, A., et al. 2022, *MNRAS*, **511**, 1408
- Gordon, K. D., Engelbracht, C. W., Rieke, G. H., et al. 2008, *ApJ*, **682**, 336
- Green, G. 2018, *JOSS*, **3**, 695
- Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, *ApJS*, **197**, 35
- Gruppioni, C., Béthermin, M., Loiacono, F., et al. 2020, *A&A*, **643**, A8
- Gruppioni, C., Pozzi, F., Rodighiero, G., et al. 2013, *MNRAS*, **432**, 23
- Helou, G., Soifer, B. T., & Rowan-Robinson, M. 1985, *ApJL*, **298**, L7
- Heywood, I., Jarvis, M. J., Hale, C. L., et al. 2022, *MNRAS*, **509**, 2150
- Hickox, R. C., & Alexander, D. M. 2018, *ARA&A*, **56**, 625
- Hoaglin, D. C., Mosteller, F., & Tukey, J. W. 1983, *Understanding Robust and Exploratory Data Analysis* (New York: Wiley)
- Hughes, D. H., Serjeant, S., Dunlop, J., et al. 1998, *Natur*, **394**, 241
- Jarvis, M., Taylor, R., Agudo, I., et al. 2016, in *Proceedings of MeerKAT Science: On the Pathway to the SKA* (Trieste: SISSA), **6**
- Jin, S., Daddi, E., Liu, D., et al. 2018, *ApJ*, **864**, 56
- Jin, S., Daddi, E., Magdis, G. E., et al. 2022, *A&A*, **665**, A3
- Kennicutt, R. C., & Evans, N. J. 2012, *ARA&A*, **50**, 531
- Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, *ApJS*, **197**, 36
- Labbé, I., Bouwens, R., Illingworth, G. D., & Franx, M. 2006, *ApJL*, **649**, L67
- Lagache, G., Puget, J.-L., & Dole, H. 2005, *ARA&A*, **43**, 727
- Laigle, C., McCracken, H. J., Ilbert, O., et al. 2016, *ApJS*, **224**, 24
- Lang, D., Hogg, D. W., & Mykytyn, D., 2016 *The Tractor: Probabilistic Astronomical Source Detection and Measurement*, Astrophysics Source Code Library, ascl:1604.008
- Le Floc'h, E., Aussel, H., Ilbert, O., et al. 2009, *ApJ*, **703**, 222
- Leauthaud, A., Massey, R., Kneib, J.-P., et al. 2007, *ApJS*, **172**, 219
- Lenz, D., Hensley, B. S., & Doré, O. 2017, *ApJ*, **846**, 38
- Liang, L., Feldmann, R., Kereš, D., et al. 2019, *MNRAS*, **489**, 1397
- Liu, D., Lang, P., Magnelli, B., et al. 2019, *ApJS*, **244**, 40
- Lutz, D., Poglitsch, A., Altieri, B., et al. 2011, *A&A*, **532**, A90
- Madau, P., & Dickinson, M. 2014, *ARA&A*, **52**, 415
- Manning, S. M., Casey, C. M., Zavala, J. A., et al. 2022, *ApJ*, **925**, 23
- McCracken, H. J., Milvang-Jensen, B., Dunlop, J., et al. 2012, *A&A*, **544**, A156
- Molnár, D. C., Sargent, M. T., Leslie, S., et al. 2021, *MNRAS*, **504**, 118
- Moneti, A., Euclid Collaboration, McCracken, H. J. 2022, *A&A*, **658**, A126
- Novak, M., Smolčić, V., Delhaize, J., et al. 2017, *A&A*, **602**, A5
- Novak, M., Smolčić, V., Schinnerer, E., et al. 2018, *A&A*, **614**, A47
- Oke, J. B., & Gunn, J. E. 1983, *ApJ*, **266**, 713
- Ranalli, P., Comastri, A., & Setti, G. 2003, *A&A*, **399**, 39
- Schinnerer, E., Sargent, M. T., Bondi, M., et al. 2010, *ApJS*, **188**, 384
- Schmidt, M. 1968, *ApJ*, **151**, 393
- Schreiber, C., Elbaz, D., Pannella, M., et al. 2018a, *A&A*, **609**, A30
- Moneti, A., Glazebrook, K., Nanayakkara, T., et al. 2018b, *A&A*, **618**, A85
- Schreiber, C., Pannella, M., Elbaz, D., et al. 2015, *A&A*, **575**, A74
- Scoville, N., Aussel, H., Brusa, M., et al. 2007, *ApJS*, **172**, 1
- Smail, I., Dudzevičiūtė, U., Stach, S. M., et al. 2021, *MNRAS*, **502**, 3426
- Smail, I., Ivison, R. J., & Blain, A. W. 1997, *ApJL*, **490**, L5
- Smolčić, V., Novak, M., Bondi, M., et al. 2017, *A&A*, **602**, A1
- Sommovigo, L., Ferrara, A., Pallottini, A., et al. 2020, *MNRAS*, **497**, 956
- Straatman, C. M. S., Labbé, I., Spitler, L. R., et al. 2014, *ApJL*, **783**, L14
- Swinbank, A. M., Simpson, J. M., Smail, I., et al. 2014, *MNRAS*, **438**, 1267
- Talia, M., Cimatti, A., Giulietti, M., et al. 2021, *ApJ*, **909**, 23
- Thomson, A. P., Ivison, R. J., Simpson, J. M., et al. 2014, *MNRAS*, **442**, 577
- Toft, S., Smolčić, V., Magnelli, B., et al. 2014, *ApJ*, **782**, 68
- Valentino, F., Tanaka, M., Davidzon, I., et al. 2020, *ApJ*, **889**, 93
- van der Lugt, D., Hodge, J. A., Algera, H. S. B., et al. 2023, *ApJ*, **951**, 131
- Vardoulaki, E., Jiménez Andrade, E. F., Delvecchio, I., et al. 2021, *A&A*, **648**, A102
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *NatMe*, **17**, 261
- Wang, T., Schreiber, C., Elbaz, D., et al. 2019, *Natur*, **572**, 211
- Weaver, J. R., Kauffmann, O. B., Ilbert, O., et al. 2022, *ApJS*, **258**, 11
- Weiβ, A., De Breuck, C., Marrone, D. P., et al. 2013, *ApJ*, **767**, 88
- Whitler, L., Stark, D. P., Endsley, R., et al. 2023, *MNRAS*, **519**, 5859
- Yun, M. S., Reddy, N. A., & Condon, J. J. 2001, *ApJ*, **554**, 803
- Zavala, J. A., Casey, C. M., Manning, S. M., et al. 2021, *ApJ*, **909**, 165