

# Probabilistic and progressive deblended far-infrared and sub-millimetre point source catalogues

## I. Methodology and first application in the COSMOS field<sup>★</sup>

Lingyu Wang<sup>1,2</sup>, Antonio La Marca<sup>1,2</sup>, Fangyou Gao<sup>2,3,4</sup>, William J. Pearson<sup>5</sup>,  
Berta Margalef-Bentabol<sup>1</sup>, Matthieu Béthermin<sup>6,7</sup>, Longji Bing<sup>8</sup>, James Donnellan<sup>8</sup>, Peter D. Hurley<sup>8</sup>,  
Seb J. Oliver<sup>8</sup>, Catherine L. Hale<sup>9,10</sup>, Matt J. Jarvis<sup>10,11</sup>, Lucia Marchetti<sup>12,13,14</sup>,  
Mattia Vaccari<sup>13,14,15</sup>, and Imogen H. Whittam<sup>10,11</sup>

<sup>1</sup> SRON Netherlands Institute for Space Research, Landleven 12, 9747 AD Groningen, The Netherlands  
e-mail: 1.wang@sron.nl

<sup>2</sup> Kapteyn Astronomical Institute, University of Groningen, Postbus 800, 9700 AV Groningen, The Netherlands

<sup>3</sup> Department of Astronomy, Nanjing University, Nanjing 210093, PR China

<sup>4</sup> Key Laboratory of Modern Astronomy and Astrophysics (Nanjing University), Ministry of Education, Nanjing 210093, PR China

<sup>5</sup> National Centre for Nuclear Research, Pasteura 7, 02-093 Warszawa, Poland

<sup>6</sup> Aix-Marseille Univ., CNRS, CNES, LAM, Marseille, France

<sup>7</sup> Université de Strasbourg, CNRS, Observatoire astronomique de Strasbourg, UMR 7550, 67000 Strasbourg, France

<sup>8</sup> Astronomy Centre, Department of Physics and Astronomy, University of Sussex, Falmer, Brighton BN1 9QH, UK

<sup>9</sup> Institute for Astronomy, School of Physics and Astronomy, University of Edinburgh, Royal Observatory Edinburgh, Blackford Hill, Edinburgh EH9 3JH, UK

<sup>10</sup> Sub-Department of Astrophysics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK

<sup>11</sup> Department of Physics and Astronomy, University of the Western Cape, Robert Sobukwe Road, 7535 Bellville, Cape Town, South Africa

<sup>12</sup> 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 Astronomy, University of Cape Town, 7701 Rondebosch, Cape Town, South Africa

<sup>14</sup> INAF – Istituto di Radioastronomia, via Gobetti 101, 40129 Bologna, Italy

<sup>15</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 21 December 2023 / Accepted 12 May 2024

## ABSTRACT

**Context.** Single-dish far-infrared (far-IR) and sub-millimetre (sub-mm) point source catalogues and their connections with galaxy catalogues at other wavelengths are of paramount importance to studying galaxy evolution. However, due to the large mismatch in spatial resolution, cross-matching galaxies detected at different wavelengths is not straightforward.

**Aims.** This work aims to develop the next-generation deblended far-IR and sub-mm catalogues in deep extragalactic survey fields, by extracting photometry at the positions of known sources. We present the first application of our methodology in the COSMOS field.

**Methods.** Our progressive deblending used the Monte Carlo Markov chain (MCMC)-based Bayesian probabilistic framework known as XID+. The deblending process started from the *Spitzer*/MIPS 24  $\mu\text{m}$  data, using an initial prior list composed of sources selected from the COSMOS2020 catalogue and radio catalogues from the VLA and the MeerKAT surveys, based on spectral energy distribution (SED) modelling which predicts fluxes of the known sources at the deblending wavelength. To speed up flux prediction, we made use of a neural network-based emulator. After deblending the 24  $\mu\text{m}$  data, we proceeded to the *Herschel* PACS (100 and 160  $\mu\text{m}$ ) and SPIRE wavebands (250, 350 and 500  $\mu\text{m}$ ). Each time we constructed a tailor-made prior list based on the predicted fluxes of the known sources, taking into account the deblended photometry from the previous steps.

**Results.** Using simulated far-IR and sub-mm sky, we detailed the performance of our deblending pipeline. After validation with simulations, we then deblended the real observations from 24 to 500  $\mu\text{m}$  and compared with blindly extracted catalogues and previous versions of deblended catalogues. As an additional test, we deblended the SCUBA-2 850  $\mu\text{m}$  map and compared our deblended fluxes with ALMA measurements, demonstrating a higher level of flux accuracy compared to previous results. We publicly release our XID+ deblended point source catalogues, including best estimates and posterior probability distribution functions. These deblended long-wavelength data, which are cross-matched with multi-band photometry by construction, are crucial for studies such as deriving the fraction of dust-obscured star formation and better separation of quiescent galaxies from dusty star-forming galaxies.

**Key words.** catalogs – surveys – galaxies: general

\* Table 4 is available at the CDS via anonymous ftp to [cdsarc.cds.unistra.fr](https://cdsarc.cds.unistra.fr) (130.79.128.5) or via <https://cdsarc.cds.unistra.fr/viz-bin/cat/J/A+A/688/A20>

## 1. Introduction

Around half of all the luminous power from star formation and supermassive black hole (SMBH) accretion is obscured by dust (Puget et al. 1996; Fixsen et al. 1998; Hauser & Dwek 2001; Driver et al. 2008). The interstellar dust heated by ultraviolet (UV)/optical light re-emits the absorbed photons approximately as a modified blackbody with a peak at far-infrared (IR) and sub-millimetre (sub-mm) wavelengths (Magnelli et al. 2012; Casey 2012; Casey et al. 2014; Wang et al. 2016). Therefore, far-IR and sub-mm observations of dusty star-forming galaxies (DSFG) are of key importance to achieving an unbiased census of the cosmic star-formation history (Madau & Dickinson 2014; Wang et al. 2019; Gruppioni et al. 2020; Zavala et al. 2021). It is also well known that massive galaxies at higher redshifts ( $z > 1$ ) are increasingly dominated by DSFGs due to the correlation between stellar mass and the level of dust obscuration for star-forming galaxies (Pannella et al. 2009; Wuyts et al. 2011; Whitaker et al. 2012, 2017; Algera et al. 2023) and the increasing fraction of the overall star-forming galaxy population with increasing redshift (Martis et al. 2016; Weaver et al. 2023). Thus, the DSFG population also plays a crucial role in understanding massive galaxy formation and evolution (Bourne et al. 2017; Gao et al. 2021; Long et al. 2023; Dudzevičiūtė et al. 2020).

The overarching goal of this work is to develop the next-generation deblended far-IR and sub-mm galaxy photometric catalogues in the premier deep extragalactic survey fields, building on our experience and expertise in analysing single-dish long-wavelength data with coarse spatial resolution ( $\sim 10$  times or more worse than optical and near-IR observations). There are several reasons why deblending is critical. Deep long-wavelength observations often result in confusion-limited maps. Due to the confusion limit which reaches as high as  $\sim 20$  mJy ( $5\sigma$ ) in the *Herschel*/SPIRE bands (Nguyen et al. 2010), blind detections are limited to only a small number of bright sources (Elbaz et al. 2011; Béthermin et al. 2015) accounting for  $< \sim 15\%$  of the cosmic IR background at SPIRE wavelengths (Oliver et al. 2010). Because of the large beam size, blindly detected sources suffer from large positional uncertainty and flux boosting which become increasingly severe with decreasing flux (Smith et al. 2012; Wang et al. 2014). In addition, a single blind detection may not correspond to one source but instead a blend of several sources within a single spatial beam, i.e. the so-called multiplicity issue (Bussmann et al. 2015; Scudder et al. 2016; Hatziminaoglou et al. 2018; Wang et al. 2021). Therefore, it can be extremely challenging to match detections from single-dish far-IR/sub-mm maps with their counterparts in the optical/near-IR. To probe fainter sources below the confusion limit as well as to overcome the difficulty in multi-wavelength source cross-matching, deblending is needed to correctly distribute fluxes among the contributing sources.

In the past decade or so, several techniques have been developed that can deblend sources in low-resolution far-IR/sub-mm maps, with varying degrees of success. These techniques are typically based on the positions of sources detected in high-resolution imaging at other wavelengths, for example at the *Spitzer*/MIPS 24  $\mu\text{m}$  and the radio 1.4 GHz (Magnelli et al. 2010; Béthermin et al. 2010; Roseboom et al. 2010, 2012; Chapin et al. 2011; Lee et al. 2013) wavelengths, thanks to the strong correlation between the 24- $\mu\text{m}/\text{radio}$  emission and the far-IR/sub-mm dust emission (Rieke et al. 2009; Ivison et al. 2010; Delhaize et al. 2017). Most of these techniques use a maximum-likelihood optimisation approach partly due to its computational ease. However, they suffer from two major issues. The first is that variance

and covariance of source fluxes cannot be properly estimated, which is problematic as neighbouring sources are expected to be highly correlated (particularly when the source surface density is high). The second is overfitting when many of the sources are intrinsically faint which can lead to a systematic flux underestimation. The list-driven algorithm DESPHOT (Roseboom et al. 2010, 2012; Wang et al. 2014) specifically developed for the *Herschel* Multi-tiered Extragalactic Survey (HerMES; Oliver et al. 2012) tried to overcome this by using the non-negative weighted least absolute shrinkage and selection operator (LASSO) algorithm (Tibshirani 1996; Zou 2006; Braak et al. 2010), a shrinkage and selection method that introduces an additional penalty term in order to decrease the number of sources needed to account for the emission in the map. However, when multiple sources are very close to each other, this method can have the opposite problem by incorrectly assigning all the flux to one source.

In this paper, we aim to construct state-of-the-art deblended far-IR and sub-mm point source catalogue in the Cosmic Evolution Survey (COSMOS; Scoville et al. 2007) field, using a progressive and probabilistic approach. There are several deblended catalogues in this field already. For example, Jin et al. (2018) presented a super-deblended catalogue from 24  $\mu\text{m}$  to mm wavelengths. Their deblending method is progressive (i.e., moving from short to long wavelengths) and relies on a  $K_s$ -selected prior catalogue, with additional sources selected in the radio. Hurley et al. (2017) presented a deblended *Herschel*/SPIRE catalogue using prior sources selected at 24  $\mu\text{m}$  and a Bayesian Markov chain Monte Carlo (MCMC)-based probabilistic deblending and source extraction tool called XID+. In developing the next generation state-of-the-art deblended catalogues, we need to combine the strengths of the previous approaches, for example, the progressive deblending in Jin et al. (2018) to exploit the information encoded in the multi-wavelength data and the probabilistic deblending in Hurley et al. (2017) to fully explore the posterior, the variance and covariance between sources. At the same time, we want to improve on several aspects of the previous deblending works. For example, we need to take into account the correlation between different far-IR/sub-mm bands rather than fitting each band independently. The prior catalogue itself, which has a significant impact on the quality of the deblended far-IR/sub-mm fluxes, has also changed substantially. For example, in the COSMOS field, there is now a newer and deeper multi-wavelength photometric catalogue with improved photometric redshift (photo- $z$ ) and stellar mass estimates. While we do not make use of the photo- $z$  and stellar mass estimations, the improved photometric data and better spectral energy distribution (SED) coverage lead to a more complete and reliable prior list for deblending the long-wavelength data.

The structure of this paper is as follows. In Sect. 2, we explain how we build the initial prior catalogue for deblending the long-wavelength data based on the latest photometric catalogue in COSMOS and additional radio source catalogues. First, we give an overview of the various datasets and our selection criteria. Then, we present a deep learning-based SED modelling and fitting emulator which is used to predict fluxes at the wavelengths to be deblended. This step is critical as it allows us to select the most relevant sources to be included in the prior. In Sect. 3, we describe the key properties of the far-IR and sub-mm data to be deblended from 24  $\mu\text{m}$  to 500  $\mu\text{m}$  and our probabilistic deblending tool XID+. In Sect. 4, we focus on validation of our deblending methodology using realistic simulations of the far-IR and sub-mm sky. Issues such as flux accuracy, flux precision and accuracy of flux uncertainty are discussed in detail. In Sect. 5,

we deblend the real observations from 24  $\mu\text{m}$  to 500  $\mu\text{m}$ . Our final deblended far-IR and sub-mm point source catalogue is then compared with blindly extracted catalogues as well as the super-deblended catalogue from Jin et al. (2018). As an additional test, we also deblend the SCUBA-2 850  $\mu\text{m}$  map and compare with ALMA measurements. In Sect. 6, we show two example science applications of our deblended long-wavelength catalogue by examining the galaxy star formation main sequence (SFMS) and the far-IR to radio correlation (FIRC). In Sect. 7, we present our conclusions and future directions. Throughout the paper, we adopt the Chabrier (2003) initial mass function (IMF) and the Wilkinson Microwave Anisotropy Probe year 7 (WMAP7) cosmology (Komatsu et al. 2011; Larson et al. 2011) with  $\Omega_{\text{M}} = 0.272$ ,  $\Omega_{\Lambda} = 0.728$ , and  $H_0 = 70.4 \text{ km s}^{-1} \text{ Mpc}^{-1}$ .

## 2. Construction of the initial prior catalogue

In this section, first we introduce the multi-wavelength datasets used to construct the initial prior catalogue for deblending the *Spitzer*/MIPS 24  $\mu\text{m}$  data, i.e., the COSMOS2020 catalogue and the joint radio catalogue from combining the VLA and MeerKAT data. The COSMOS2020 catalogue is the main dataset which can be regarded as a proxy for a stellar mass-based selection, while the radio source catalogues provide additional high-redshift (high- $z$ ) sources which may be missing from the COSMOS2020 catalogue (Liu et al. 2018; Jin et al. 2018). Then, we describe the SED fitting procedure used to predict the 24  $\mu\text{m}$  flux, taking into account systematic uncertainties due to varying assumptions in SED fitting. This is another major improvement over previous deblending efforts. In addition, to speed up the flux prediction step, we use a Deep Learning Neural Network (DLNN) as an emulator which is trained on the outputs from the SED fitting step.

### 2.1. The input multi-wavelength datasets

#### 2.1.1. The COSMOS2020 catalogue

The main component in building our prior catalogue for deblending the far-IR and sub-mm data is the latest photometric catalogue from COSMOS. This catalogue named COSMOS2020 (Weaver et al. 2022) is an updated version of the previous catalogue COSMOS2015 (Laigle et al. 2016). It includes multi-band photometry from far-UV (FUV) and near-UV (NUV) band of the Galaxy Evolution Explorer (GALEX) satellite to *Spitzer*/Infrared Array Camera (IRAC) over the wavelength range 0.1–10  $\mu\text{m}$  for about 1.7 million sources. Compared to COSMOS2015, the latest release includes new imaging data, such as the  $U$ -band data from the Canada-France-Hawaii Telescope (CFHT) large area  $U$ -band deep survey (CLAUDS; Sawicki et al. 2019), grizy-band data from the Hyper Suprime-Cam (HSC) Subaru Strategic Program (SSP) PDR2 (Aihara et al. 2019), YJHKs-band data from the fourth data release (DR4) of the UltraVISTA survey (McCracken et al. 2012; Moneti et al. 2023), and *Spitzer*/IRAC data from the Cosmic Dawn Survey (Euclid Collaboration et al. 2022). Source detection was performed on a very deep multi-band chi-squared  $i_zYJHKs$  detection image. There are two versions of COSMOS2020: a Classic one produced by a pipeline similar to COSMOS2015 (with an exception for the IRAC bands), and the Farmer version (limited to the UltraVISTA coverage) produced by the software Tractor (Lang et al. 2016). Photometric redshift (Photo- $z$ ) was derived using either the LePhare (Arnouts et al. 2002; Ilbert et al. 2006) or EAZY code (Brammer et al. 2008). Therefore, there

are in total four possible combinations in choosing a version of the photometric measurements and a version of the photo- $z$ . In this paper, we chose the Farmer catalogue with photo- $z$  derived using LePhare as this combination performs the best in terms of comparisons to spectroscopic redshifts<sup>1</sup> (Weaver et al. 2022). Extinctions due to the Milky Way and photometric offsets are corrected in COSMOS2020.

The photometric catalogue, downloaded from the release site<sup>2</sup>, contains 964 506 sources. Following recommendations from the COSMOS team, we set the flag FLAG\_COMBINED equal to zero to select areas that have coverage from HSC, Ultra-VISTA, and *Spitzer*/IRAC and are not affected by bright stars or large artifacts, which reduced the catalogue to 746 976 sources. We also made use of the star/galaxy separation provided in the Le Phare photo- $z$  and physical parameters. We selected sources which have the star/galaxy separation flag lp\_type value equal to 0 (galaxy) or 2 (X-ray source), which further reduced the catalogue to a total of 711 106 sources. According to the readme file<sup>3</sup>, the area selected by setting FLAG\_COMBINED equal to zero is  $1.278 \text{ deg}^2$ , which is much smaller than the full COSMOS field of  $\sim 2 \text{ deg}^2$ . In the future, with new ground- and space-based observations such as the planned observations from the European Space Agency survey mission *Euclid* (Laureijs et al. 2011) and the *James Webb* Space Telescope (JWST; Gardner et al. 2006), it will be possible extend this work over a larger area in COSMOS.

#### 2.1.2. The joint radio catalogue

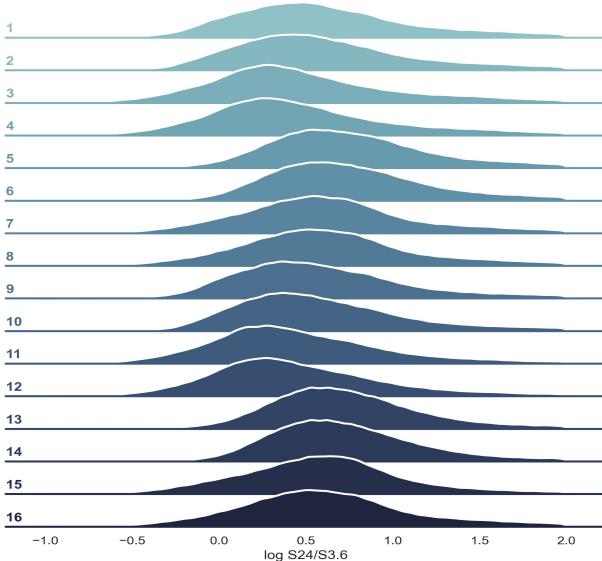
In order to add additional high- $z$  sources which may be missing from COSMOS2020, we made use of the source catalogue from the VLA-COSMOS 3 GHz Large Project (Smolčić et al. 2017), which is based on 384 h of observations with the *Karl G. Jansky* Very Large Array (VLA) and uniformly covers the entire COSMOS field down to a median rms of  $2.3 \text{ }\mu\text{Jy beam}^{-1}$  at an angular resolution of  $0.75''$ . The catalogue contains 10 830 blindly extracted sources down to  $5\sigma$ , with a total fraction of spurious sources  $<2.7\%$ . Positional accuracy is estimated to be around  $0.01''$  for bright sources with  $S/N > 20$ .

We also made use of the MeerKAT International GHz Tiered Extragalactic Exploration (MIGHTEE) survey (Jarvis et al. 2016; Heywood et al. 2022), which is an ongoing flagship Large Survey Project on MeerKAT (Jonas & MeerKAT Team 2016) and precursor to the Square Kilometre Array. When completed, MIGHTEE will cover  $20 \text{ deg}^2$  across four extragalactic deep fields to a depth of  $\sim 1 \text{ }\mu\text{Jy beam}^{-1}$  rms at a central bandwidth frequency of  $\sim 1.3 \text{ GHz}$ . In this paper, we used the COSMOS catalogue extracted from the Data Release 1 total intensity map with an angular resolution of  $\sim 8''$  (Hale, Heywood, Jarvis et al., in prep.). Images at a higher resolution ( $5''$ ) are also available but are shallower by a factor of  $\sim 3$ . The thermal noise in the MIGHTEE COSMOS image reaches  $1.7 \text{ }\mu\text{Jy beam}^{-1}$ . For comparison, the VLA-COSMOS 1.4 GHz Large Project map (Schinnerer et al. 2010) has a depth of  $\sigma = 12 \text{ }\mu\text{Jy beam}^{-1}$  in the deepest central  $50' \times 50'$  area. Consequently, the MIGHTEE continuum images are limited by confusion noise rather than thermal noise. By comparing to sources detected in the shallower but much-higher resolution VLA-COSMOS 3 GHz Large Project,

<sup>1</sup> We do not make use of photo- $z$  in our deblending pipeline. Only source type information provided by LePhare is used.

<sup>2</sup> <https://cosmos2020.calet.org/catalogues/>

<sup>3</sup> [https://cosmos2020.calet.org/catalogues/MASKS/MASKS\\_README.txt](https://cosmos2020.calet.org/catalogues/MASKS/MASKS_README.txt)



**Fig. 1.** Distributions of the ratio of the predicted MIPS 24  $\mu\text{m}$  flux to the observed IRAC 3.6  $\mu\text{m}$  flux from the 16 CIGALE SED fitting runs, showing systematic effects arising from varying assumptions (such as the adopted SFH, AGN templates, dust attenuation and emission models) in the SED modelling and fitting process.

the mean offsets in (RA, Dec) of the MIGHTEE sources have been determined to be  $(-0.25'' \pm 0.01'', -0.08'' \pm 0.01'')$ . Cross-matching MIGHTEE sources with the VLA 1.4 GHz source catalogue (Schinnerer et al. 2010) and comparing measurements of the integrated flux density also demonstrate a high level of photometric consistency (with a median MIGHTEE to VLA flux ratio of 0.95).

To find radio sources which are missing in COSMOS2020, we cross-matched the VLA 3 GHz and the MIGHTEE 1.3 GHz catalogues with COSMOS2020 (after applying the flags as described in Sect. 2.1.1). If a VLA or MIGHTEE source did not have a match in COSMOS2020 within  $1''$ , we treated it as an additional source to be included in the prior catalogue. In total, we found 2791 such sources as most of the radio sources are already included in COSMOS2020. In our previous work (Wang et al. 2021), we used the radio flux as an informative flux prior to deblend the far-FIR/sub-mm data. In this work, we decided to just use the detection information to keep the potential bias in the deblended flux at a minimum. In Sect. 6, we examine whether MIGHTEE sources not included in COSMOS2020 follow the FIRC, which also serves as an independent test of the quality of our deblended photometry.

## 2.2. Flux prediction using SED fitting

The number of sources in COSMOS2020 and the joint radio catalogue (totalling over 700 000 sources) is too large to be directly deblended in the far-IR and sub-mm maps which have much worse spatial resolutions than optical and near-IR maps. We thus need to develop a way of selecting the most relevant sources to be included in the prior catalogue. As discussed in Liu et al. (2018), ideally we want to keep the average number of sources to be deblended under a single spatial beam to be  $\lesssim 1$  in all bands.

### 2.2.1. The CIGALE SED library

Our probabilistic and progressive deblending starts from the *Spitzer*/MIPS 24  $\mu\text{m}$  and then moves to the *Herschel* PACS and

SPIRE bands. This approach has the benefit of staying as close as possible to the measured data without having to extrapolate too far down in wavelength, and at the same time extracting as much information as possible from the data. First we select a subset of sources from COSMOS2020 which are most likely to contribute to the emission in the 24  $\mu\text{m}$  map.

We made use of the SED modelling and fitting tool Code Investigating GALaxy Emission (CIGALE; Noll et al. 2009) with an improved fitting procedure by Serra et al. (2011) to predict the 24  $\mu\text{m}$  flux in the observed frame based on the multi-band photometry in COSMOS2020. To make the process of SED fitting and flux prediction for the entire catalogue faster, we first used CIGALE to predict the observed 24  $\mu\text{m}$  fluxes for a random sample of 200 000 sources. To explore systematic uncertainties introduced by varying assumptions in SED fitting, we carried out a total of 16 CIGALE runs with all possible combinations of two star-formation history models (a delayed  $\tau$  model plus a late starburst or double exponential forms), two dust attenuation models (modified Charlot & Fall 2000 or modified Calzetti et al. 2000 attenuation law), two dust emission models (IR SED templates from the Draine et al. 2014 model or the Dale et al. 2014 model), and two active galactic nuclei (AGN) models (smooth AGN templates from the Fritz et al. 2006 model or clumpy AGN templates from the SKIRTOR model in Stalevski et al. 2012). The SED model components and parameters (representing a range of common configurations in the literature) are listed in Table A.1. Figure 1 shows the distributions of the ratio between the predicted 24  $\mu\text{m}$  flux to the observed IRAC 3.6  $\mu\text{m}$  flux from the 16 CIGALE SED fitting runs. Overall, the systematic differences between the different runs are smaller than the spread in each run, demonstrating that the SEDs are well sampled enough to allow for predictions which are not very sensitive to inherent assumptions in the SED fitting.

### 2.2.2. The deep learning-based SED emulator

The inputs (i.e. the multi-band photometry in COSMOS2020) and outputs (i.e. the predicted flux in a specific far-IR/sub-mm band to be deblended) of the 16 CIGALE runs formed an SED library which was then used to train a DLNN developed in Euclid Collaboration (2023). The DLNN model consisted of four linear fully connected layers with different numbers of neurons and non-linear activation functions, as listed in Table 1. The purpose of using a DLNN as an SED emulator was to speed up the process of predicting the observed 24  $\mu\text{m}$  flux density. To ensure reliability of the predicted fluxes, we selected sources which are detected at a signal-to-noise ratio ( $S/N > 0.1$ ) in at least 40% of the available bands. All flux measurements were converted to log-space. We augmented the COSMOS2020 multi-band photometric data by calculating 16 colours<sup>4</sup>, namely  $u - r$ ,  $g - r$ ,  $r - i$ ,  $i - z$ ,  $z - y$ ,  $J - K_s$ ,  $H - Y$ ,  $J - H$ ,  $Y - K_s$ ,  $r - K_s$ ,  $u - K_s$ ,  $r$  – IRAC channel 1 (3.6  $\mu\text{m}$ ),  $u$  – IRAC channel 4 (8.0  $\mu\text{m}$ ), IRAC channel 1 – channel 3 (5.8  $\mu\text{m}$ ), IRAC channel 1 – channel 4, and IRAC channel 2 (4.5  $\mu\text{m}$ ) – channel 4. In addition, we applied a normalisation process to scale data to a range of 0–1 using the transformation below,

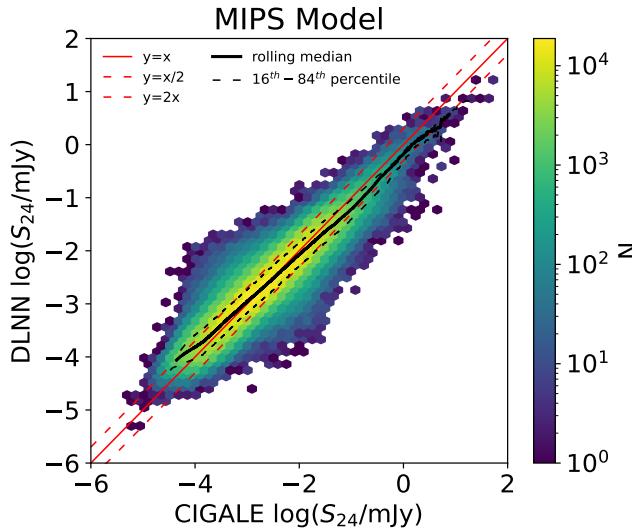
$$x' = \frac{x - \min}{\max - \min}, \quad (1)$$

<sup>4</sup> A popular way of improving the performance of machine learning methods is through combining existing features with new features, which can be obtained from the differences and/or ratios of the existing features.

**Table 1.** DLNN architecture used as an SED emulator.

Layer	Properties
Input	$N_{\text{in}}$ neurons <sup>(a)</sup>
Linear, fully connected	2000 neurons
Dropout	Dropout rate 0.35
Linear, fully connected	1000 neurons
Linear, fully connected	500 neurons
Dropout	Dropout rate 0.35
Output, fully connected	$N_{\text{out}}$ neurons <sup>(b)</sup>

**Notes.** The input and output layers have a number of neurons that depends on the particular task (i.e., predicting fluxes in the MIPS, PACS or SPIRE bands). <sup>(a)</sup>To predict the 24  $\mu\text{m}$  flux, the number of neurons in the input layer  $N_{\text{in}}$  corresponds to the number of the multi-band photometry (satisfying the S/N requirements) in COSMOS2020 and the 16 colours as explained in Sect. 2.2.2. To predict the PACS and SPIRE fluxes,  $N_{\text{in}}$  also includes the deblended far-IR/sub-mm fluxes in the previous step(s). <sup>(b)</sup>To predict the 24  $\mu\text{m}$  flux, the output is a single number. To predict the PACS and SPIRE fluxes, the output consists of two and three numbers, respectively.



**Fig. 2.** Predicted MIPS 24  $\mu\text{m}$  flux from the DLNN on the test set compared to the predicted values from the 16 CIGALE runs. Colour coding is based on number of sources. The running median is close to the one-to-one line. The 16th and 84th percentile ranges are mostly within a factor of two from the one-to-one line.

where  $x$  refers to a given feature of a given source and the min and max refer to the minimum and maximum value for all sources. Missing data were assigned a value of  $-1$ .

We used the TensorFlow framework (Abadi et al. 2016) to build and train the DLNN model. We used as activation function for each layer a Rectified Linear Unit (ReLU; Nair & Hinton 2010) which returns  $\max(x, 0)$  when passed  $x$ . To prevent overfitting, we included two dropout layers (Srivastava et al. 2014) with a dropout rate of 0.35, one after the first layer and the other after the third layer. To train the DLNN we adopted a mean squared error loss function (a typical performance metric for regression problems) which was optimised using the Adam algorithm (Kingma & Ba 2014). The learning rate was set to 0.001. The sample was split into a 40% training set, a 10% validation set, and a 50% test set. The DLNN was trained using the training sample, while the validation sample was employed to derive an independent estimate of the loss function. The training and validation samples were passed to the DLNN in batches of 4000

**Table 2.** DLNN performance.

DLNN Model	$\widetilde{\Delta S}$	NMAD	MFD
MIPS	0.0000	0.0011	-0.0028
PACS100	-0.0001	0.0096	-0.0110
SPIRE250 (opt. – PACS)	-0.0006	0.0066	-0.0230
SPIRE250 (opt. – MIPS)	-0.0001	0.0240	-0.0063

**Notes.** Both the median bias  $\widetilde{\Delta S}$  and the normalised median absolute deviation NMAD are in unit of mJy. The median fractional difference MFD is dimensionless. For the PACS DLNN, we report the metrics measured for the 100  $\mu\text{m}$  band. For the two SPIRE DLNNs (one with optical to MIPS data and one with optical to PACS data), we report the results for the 250  $\mu\text{m}$  band.

objects. Therefore, the parameters of the DLNN were updated after every 4000 objects. The training stopped when the loss function of the validation sample converged, to avoid overfitting the training set. The third dataset, the test sample, was used only to evaluate the network performance.

In Fig. 2, we compare the predicted value from the trained DLNN ( $S_{24}^{\text{DLNN}}$ ) on the test set with the CIGALE predicted 24  $\mu\text{m}$  fluxes ( $S_{24}^{\text{CIGALE}}$ ). Note that as there are 16 CIGALE runs, each galaxy in the test set is plotted 16 times. The running median is close to the one-to-one line across the entire dynamic range with the 16th and 84th percentile ranges are mostly within a factor of two from the 1:1. At the very faint end ( $<10^{-4}$  mJy),  $S_{24}^{\text{DLNN}}$  starts to deviate significantly from  $S_{24}^{\text{CIGALE}}$ . This is not an issue, as we apply a flux cut at  $10^{-2}$  mJy or 10  $\mu\text{Jy}$  to decide whether to keep a source in the prior list (see Sects. 2.3 and 4). We measure the median bias

$$\widetilde{\Delta S} = \text{median}(S_{24}^{\text{DLNN}} - S_{24}^{\text{CIGALE}}), \quad (2)$$

the normalised median absolute deviation (equivalent to the standard deviation for a normal distribution)

$$\text{NMAD} = 1.48 \cdot \text{median}(|(S_{24}^{\text{DLNN}} - S_{24}^{\text{CIGALE}}) - \widetilde{\Delta S}|), \quad (3)$$

and the median fractional difference

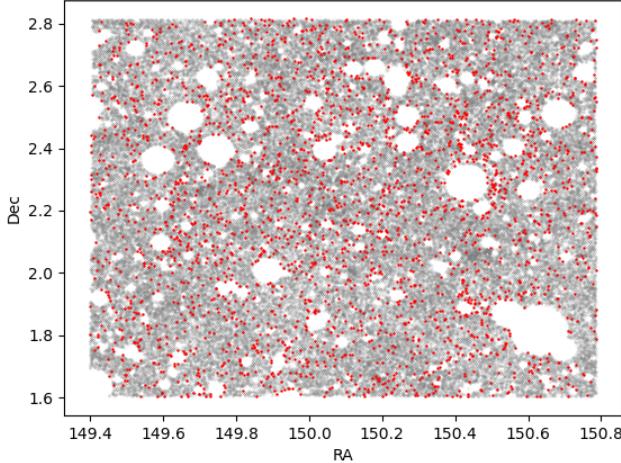
$$\text{MFD} = \text{median}\left(\frac{S_{24}^{\text{DLNN}} - S_{24}^{\text{CIGALE}}}{S_{24}^{\text{CIGALE}}}\right). \quad (4)$$

We report these metrics in Table 2. The perfect algorithm would have zero bias, MFD and NMAD.

### 2.3. The initial prior catalogue for deblending the MIPS 24 $\mu\text{m}$ data

After training the DLNN, we then applied it to the whole dataset of 711 106 sources selected from COSMOS2020 to predict their observed 24  $\mu\text{m}$  fluxes. We constructed an initial prior catalogue for deblending the MIPS 24  $\mu\text{m}$  image as follows:

1. We included 121 227 sources from COSMOS2020 with  $\text{lp\_type} = 0$  (galaxies) and predicted  $S_{24} > 10 \mu\text{Jy}$ . Note that we use the notations  $S_{24}$  and  $S_{24}$  (and similarly for fluxes at other wavelengths) interchangeably. This flux cut is chosen in light of the noise properties of the 24  $\mu\text{m}$  map and the validation results based on simulations (see Sect. 4.1).
2. We included 5141 sources from COSMOS2020 with  $\text{lp\_type} = 0$  but no predicted  $S_{24}$ , in order to catch sources for which we could not predict their 24  $\mu\text{m}$  fluxes due to insufficient SED coverage and/or low S/N.



**Fig. 3.** Sky distribution of the sources in our initial prior catalogue for deblending the MIPS 24  $\mu\text{m}$  data. The grey data points are the sources selected from COSMOS2020. The red data points are the additional radio sources from the VLA-COSMOS 3GHz and MIGHTEE 1.3 GHz surveys which are not included in COSMOS2020.

3. We included 2019 sources from COSMOS2020 with  $lp\_type = 2$  (X-ray sources) for which the SED modelling and fitting procedure may not work as well as for other source types.
4. We included 2791 radio sources from the VLA-COSMOS 3 GHz and MIGHTEE 1.3 GHz surveys which are not present in COSMOS2020. This step is to include additional high- $z$  far-IR/sub-mm emitting sources.

To summarise, we found a total of 131 178 sources (corresponding to 0.29 sources per beam) in the initial prior catalogue for deblending the MIPS 24  $\mu\text{m}$  map. Figure 3 shows the sky distribution of these sources, with the missing radio sources in red. In particular, the extra radio sources seem to cluster around the masked region in COSMOS2020.

### 3. Deblending the far-IR and sub-mm data

We adopt a progressive deblending approach from the 24  $\mu\text{m}$  data to the PACS and SPIRE data. After the 24  $\mu\text{m}$  map is deblended, we add the deblended 24  $\mu\text{m}$  flux to the sources in the prior list and continue with the prediction of the PACS 100 and 160  $\mu\text{m}$  fluxes, using the aforementioned DLNN trained with the SED library from the 16 CIGALE runs. Similarly, after the PACS maps are deblended and the deblended PACS fluxes are added to the sources in the prior list, we continue with the prediction of the SPIRE 250, 350 and 500  $\mu\text{m}$  fluxes. A schematic representation of our progressive deblending is shown in Fig. 4. Below we first give an account of the far-IR and sub-mm data to be deblended. Then we describe the characteristics of XID+ which is a prior-based probabilistic deblending tool and how we use it in this work.

#### 3.1. The long wavelength data

For the MIPS 24  $\mu\text{m}$  imaging data, we used the GO3 image from the COSMOS-Spitzer programme (Sanders et al. 2007). The noise level corresponds to a  $1\sigma$  standard deviation of 14 to 18  $\mu\text{Jy}$  in the point source noise (Le Floc'h et al. 2009). The  $1\sigma$  confusion noise is 11.2  $\mu\text{Jy}$  (Dole et al. 2004). The map is in unit of MJy/sr. To use XID+, we converted to mJy/beam by multiplying a factor of 1.156, assuming an idealised Gaussian beam with a full width at half maximum (FWHM) of 5.7''. However, this assumption ignores the side lobes of the real beam. We thus



**Fig. 4.** Flowchart of our progressive deblending from 24  $\mu\text{m}$  to the sub-mm wavelengths. Our initial prior catalogue for deblending the 24  $\mu\text{m}$  map is constructed using COSMOS2020 and additional radio sources from the VLA-COSMOS 3 GHz and MIGHTEE 1.3 GHz catalogues. At each step of the deblending, we use a DLNN trained on an SED library to predict the fluxes of the sources in the prior catalogue at the relevant wavelength. Based on the predicted flux, we decide whether to keep the source in the prior list.

applied a correction factor of 1.369 to account for the lost flux (Rodighiero et al. 2006).

The PACS maps came from the PACS Evolutionary Probe (PEP; Lutz et al. 2011) survey<sup>5</sup>. The FWHM of the point spread function (PSF) is 7.2'' and 12.0'' at 100 and 160  $\mu\text{m}$ , respectively. The maps, in unit of Jy/pixel, were converted to mJy/beam using a multiplicative factor of 40.8 and 28.3 at 100 and 160  $\mu\text{m}$ , assuming Gaussian beam. We applied a correction factor of 1.68 and 1.64 at 100 and 160  $\mu\text{m}$  to account for aperture correction and flux loss due to the high-pass filtering, as detailed in the data release<sup>6</sup>. The instrument noise level corresponds to a  $1\sigma$  standard deviation of 1.4 and 3.5 mJy at 100 and 160  $\mu\text{m}$ . The  $1\sigma$  confusion noise is 0.15 and 0.68 mJy at 100 and 160  $\mu\text{m}$  (Berta et al. 2011; Magnelli et al. 2013).

The SPIRE maps were obtained from the fourth release of the *Herschel* Database in Marseille hosting data from the *Herschel* Multi-tiered Extragalactic Survey (HerMES; Oliver et al. 2012)<sup>7</sup>. These maps are already in units of mJy/beam. The PSF FWHM is equal to 18.2, 24.9, and 36.3'' at 250, 350, and 500  $\mu\text{m}$ , respectively. The  $1\sigma$  instrument noise is 1.7, 2.7, and 2.9 mJy at 250, 350, and 500  $\mu\text{m}$  (Oliver et al. 2012). In comparison, the  $1\sigma$  confusion noise is 6.8, 6.3, and 5.8 mJy at 250, 350,

<sup>5</sup> <https://www.mpe.mpg.de/ir/Research/PEP/DR1>

<sup>6</sup> [https://www.mpe.mpg.de/resources/PEP/DR1\\_tarballs/readme\\_PEP\\_global.pdf](https://www.mpe.mpg.de/resources/PEP/DR1_tarballs/readme_PEP_global.pdf)

<sup>7</sup> <https://hedam.lam.fr/HerMES/>

**Table 3.** Summary of the long-wavelength imaging data.

Wavelength	Telescope/instrument	PSF <i>FWHM</i>	Pixel size	$S_{\text{cut}}$	$N_{\text{fit}}$	$\rho_{\text{fit}}$	$\sigma_{\text{instrument}}$	$\sigma_{\text{confusion}}$
24 $\mu\text{m}$	<i>Spitzer</i> /MIPS	5.7''	1.2''	10 $\mu\text{Jy}$	131 178	0.29	18.8 $\mu\text{Jy}$	11.2 $\mu\text{Jy}$
100 $\mu\text{m}$	<i>Herschel</i> /PACS	7.2''	1.2''	1.0 mJy	35 819	0.13	1.4 mJy	0.15 mJy
160 $\mu\text{m}$	<i>Herschel</i> /PACS	12.0''	2.4''	—	35 819	0.35	3.5 mJy	0.68 mJy
250 $\mu\text{m}$	<i>Herschel</i> /SPIRE	18.2''	6.0''	7.0 mJy	14 869	0.34	1.7 mJy	6.8 mJy
350 $\mu\text{m}$	<i>Herschel</i> /SPIRE	24.9''	8.3''	—	14 869	0.63	2.7 mJy	6.3 mJy
500 $\mu\text{m}$	<i>Herschel</i> /SPIRE	36.3''	12.0''	—	14 869	1.34	2.9 mJy	5.8 mJy

**Notes.** The columns (from left to right) are the nominal wavelength, telescope/instrument, FWHM of the PSF, image pixel size, flux cut applied to the predicted flux, number of sources to be fitted, source density per beam area, the  $1\sigma$  instrument noise and the  $1\sigma$  confusion noise.

and 500  $\mu\text{m}$  (Nguyen et al. 2010). Therefore, the SPIRE maps in COSMOS are dominated by confusion.

In Table 3, we show a summary of the long-wavelength imaging data to be deblended in the COSMOS field, including the nominal wavelength, the FWHM of the PSF, the image pixel size, the flux cut applied to the predicted flux in a given band, the number of sources to be fitted, the source density per beam area, the  $1\sigma$  instrument noise and confusion noise.

### 3.2. Probabilistic deblending with XID+

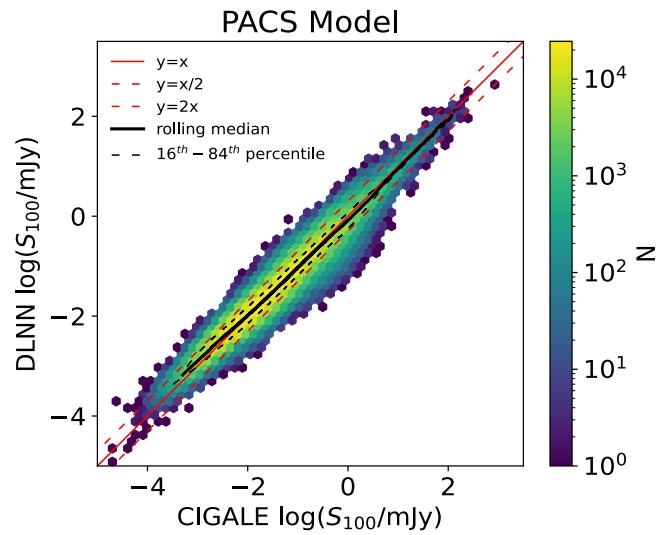
Hurley et al. (2017) developed a prior-based source extraction tool, XID+<sup>8</sup>, to extract flux information at the positions of known sources. XID+ uses a probabilistic Bayesian framework that provides a natural way to include prior information, and uses the Bayesian inference tool Stan (Stan Development Team 2015, 2018) to obtain the full posterior probability distribution on flux estimates. The basic model of our data (i.e., far-IR and sub-mm maps) can be built by the sum of three components,

$$\mathbf{d} = \sum_{i=1}^S \mathbf{P} \mathbf{f}_i + \mathcal{N}(0, \Sigma_{\text{instrumental}}) + \mathcal{N}\left(\mathbf{B}, \Sigma_{\text{confusion}}^{\text{residual}}\right). \quad (5)$$

The first component represents the contribution of the known sources which can be derived by multiplying the pointing matrix (calculated using a Gaussian PSF) and the flux vector (composed of the flux density from each source). The second term represents the instrumental noise which is provided by the uncertainty map of the observations. The third component is a global estimate for the background term with Gaussian fluctuations accounting for the unknown sources<sup>9</sup>. This model can then be compared with the observed data.

A key feature of XID+ (in contrast with maximum-likelihood based photometry methods) is that in addition to estimates of the source flux, it also determines the full posterior probability distributions. The original XID+ used a flat flux prior. Later works added informative flux prior (Pearson et al. 2017, 2018; Wang et al. 2021) which has the advantage of being able to go down to fainter flux levels. In this paper, we use the informative flux prior only in deciding whether a given source should be kept in the prior catalogue. When actually running XID+, we use a flat flux prior in order not to introduce bias in the deblended flux. Finally, we fit the two PACS bands and the three SPIRE bands simultaneously to take into account the correlations between the bands that are close together.

For deblending the PACS maps, we took the initial prior catalogue (used for deblending the 24  $\mu\text{m}$  map) and added



**Fig. 5.** DLNN predicted PACS 100  $\mu\text{m}$  fluxes on the test set compared to the values from the 16 CIGALE runs. The running median is close to the 1:1 ratio. The 16th and 84th percentile ranges are mostly within a factor of two from the 1:1.

the deblended 24  $\mu\text{m}$  fluxes. We then fed these data to the CIGALE-trained DLNN to predict simultaneously the 100 and 160  $\mu\text{m}$  flux  $S_{100}$  and  $S_{160}$ . In Fig. 5, we show the comparison between the predicted PACS 100  $\mu\text{m}$  fluxes by the DLNN ( $S_{100}^{\text{DLNN}}$ ) and the predicted values from the 16 CIGALE SED fitting runs ( $S_{100}^{\text{CIGALE}}$ ) for the test set. Performance metrics such as the median bias, MFD and NMAD of  $(S_{100}^{\text{DLNN}} - S_{100}^{\text{CIGALE}})$  are reported in Table 2. We constructed the prior catalogue for deblending the PACS maps as follows:

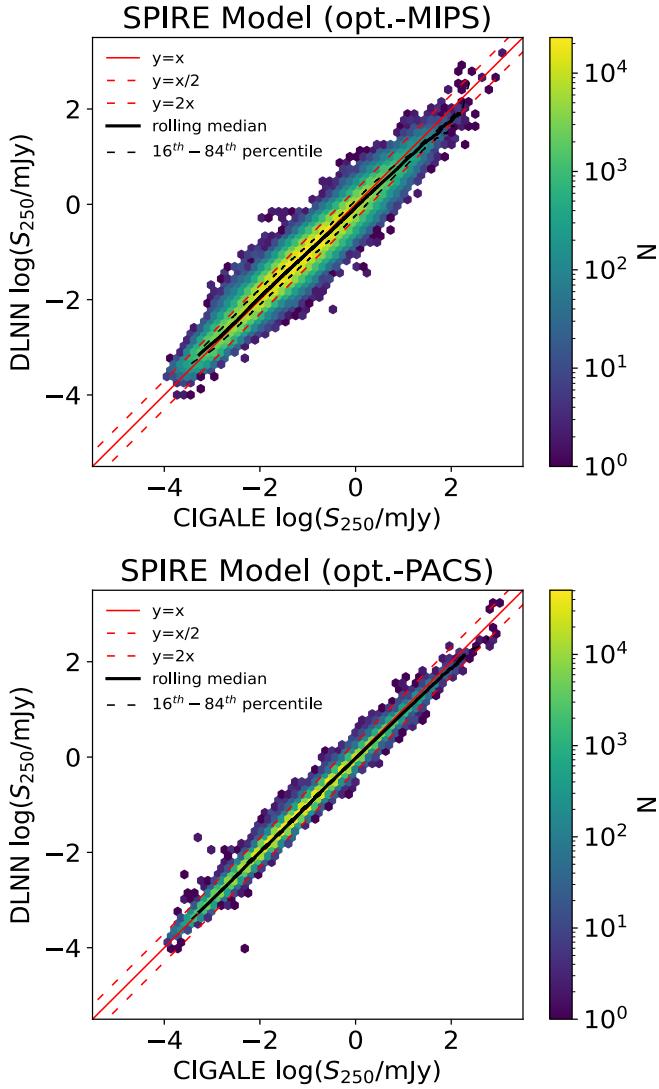
1. We included sources with predicted  $S_{100} > 1.0$  mJy. This flux cut is chosen in light of the noise properties and the validation results based on simulations (see Sect. 4.2).
2. We included sources with predicted  $S_{100} < 1.0$  mJy or without predicted  $S_{100}$ , but present in the sample of 2019 X-ray sources.
3. We included the 2791 radio sources not present in COSMOS2020.

In total, there were 35 819 sources in the prior catalogue for deblending the PACS maps, corresponding to 0.13 sources per beam at 100  $\mu\text{m}$  and 0.35 sources per beam at 160  $\mu\text{m}$ .

To build the prior catalogue for SPIRE, we took the initial prior catalogue and added the deblended 24  $\mu\text{m}$  and PACS fluxes. Then we fed these data to the DLNN to predict the 250, 350, and 500  $\mu\text{m}$  flux  $S_{250}$ ,  $S_{350}$ , and  $S_{500}$ . When the deblended PACS fluxes were available, we used the DLNN trained using

<sup>8</sup> [https://github.com/H-E-L-P/XID\\_plus](https://github.com/H-E-L-P/XID_plus)

<sup>9</sup> The unknown sources could include sources which are fainter than our applied flux cut and sources which are not detected in the COSMOS2020 catalogue nor the joint radio catalogues.



**Fig. 6.** DLNN predicted SPIRE 250  $\mu\text{m}$  flux on the test set compared to the values from the 16 CIGALE runs. The top panel corresponds to predictions from the DLNN using photometry from COSMOS2020 and the deblended 24  $\mu\text{m}$  flux. The bottom panel corresponds to predictions from the DLNN using all available photometry out to the PACS wavelengths.

also the 100 and 160  $\mu\text{m}$  CIGALE fluxes. Otherwise, we predicted SPIRE fluxes using the DLNN trained with data up to 24  $\mu\text{m}$ . In Fig. 6, we show the comparison between the predicted SPIRE 250  $\mu\text{m}$  fluxes by the two DLNNs ( $S_{250}^{\text{DLNN}}$ ) and the predicted values from the 16 CIGALE SED fitting runs ( $S_{250}^{\text{CIGALE}}$ ) for the test set. Performance metrics such as the median bias, MFD and NMAD of ( $S_{250}^{\text{DLNN}} - S_{250}^{\text{CIGALE}}$ ) for the two DLNNs (one with optical to MIPS data and one with optical to PACS data) are shown in Table 2. We constructed the prior catalogue for deblending the SPIRE maps as follows:

1. We included sources with predicted  $S_{250} > 7.0 \text{ mJy}$ . This flux cut is chosen in light of the noise properties and the validation results based on simulations (see Sect. 4.2).
2. We included sources with predicted  $S_{250} < 7.0 \text{ mJy}$  or without predicted  $S_{250}$  but present in the sample of 2019 X-ray sources.
3. We included the 2791 radio sources not present in COSMOS2020.

In total, there were 14 869 sources in the prior catalogue for deblending the SPIRE maps, corresponding to a source density of 0.34 sources per beam at 250  $\mu\text{m}$ , 0.63 sources per beam at 350  $\mu\text{m}$ , and 1.34 sources per beam at 500  $\mu\text{m}$ .

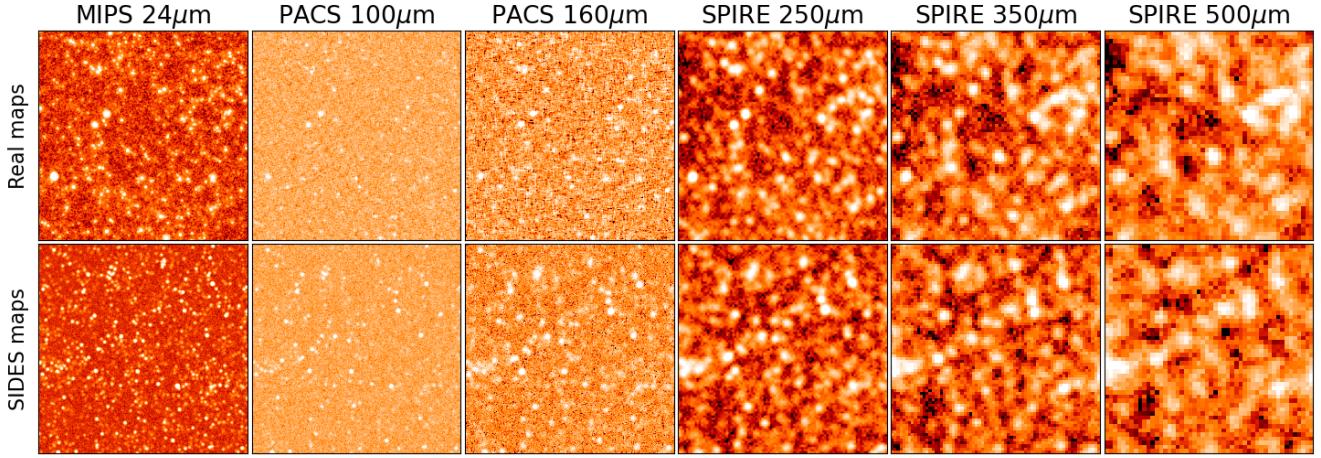
#### 4. Validation

In this section, we validate our deblending methodology using simulations. Specifically, we make use of the continuum observables in Simulated Infrared Dusty Extragalactic Sky (SIDES; Béthermin et al. 2017, 2022) which is a 2 deg<sup>2</sup> simulation of the extragalactic sky, including clustering, based on dark matter simulations and empirical prescriptions such as abundance matching between the galaxy stellar mass function (SMF) and the dark matter halo mass function, the fraction of star-forming galaxies as a function of stellar mass and redshift, the SFMS and the fraction of starbursts as a function of redshift. This empirical model is designed to reproduce a range of observed statistical properties of far-IR and sub-mm galaxies, such as the distribution of the surface brightness in the pixels (P(D)), the number counts and the power spectra of the cosmic IR background anisotropies (Gkogkou et al. 2023). To generate simulated far-IR and sub-mm maps, we adopted idealised Gaussian beams which are commonly used as representations of the PSF at these wavelengths. The FWHM of the beams and the pixel sizes are set to the same values as in the observations (Table 3). We also added Gaussian random noise based on the uncertainty map from the real observations. A realistic level of the confusion noise is already included as the SIDES simulation reproduces the observed number counts. The top row of Fig. 7 shows cutouts of the same 10'  $\times$  10' patch of the sky in COSMOS, observed at different wavelengths by MIPS, PACS, and SPIRE. In the bottom row, we show cutouts of the same size from the simulated MIPS, PACS, and SPIRE maps, demonstrating that the simulated maps can statistically reproduce the observed far-IR and sub-mm sky reasonably well.

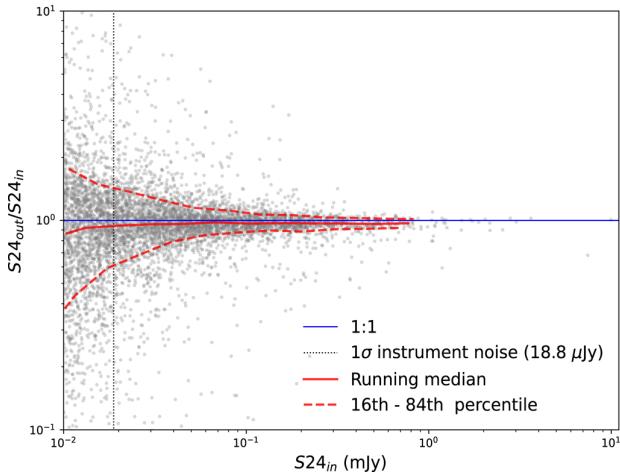
##### 4.1. Deblending results from the simulated MIPS map

To make the prior catalogue, we selected a total of 154 748 sources with a true 24  $\mu\text{m}$  flux  $S_{24} > 10 \mu\text{Jy}$ , corresponding to 0.22 sources per beam. This flux cut is equal to the 1 $\sigma$  confusion noise at 24  $\mu\text{m}$ . For validation, only the first 100 tiles<sup>10</sup> (out of 2470 in total) were run. Figure 8 shows the relation between the true 24  $\mu\text{m}$  flux ( $S_{24\text{in}}$ ) and the XID+ deblended flux ( $S_{24\text{out}}$ ). The running median of the flux ratio ( $S_{24\text{out}}/S_{24\text{in}}$ ) is very close to the 1:1 line (with deviation <5%) above the 1 $\sigma$  instrument noise of 18.8  $\mu\text{Jy}$ , demonstrating a high level of flux accuracy. The 16th–84th percentile ranges widens towards fainter sources, as expected. We can also notice that even at the bright end, XID+ can significantly underestimate or overestimate the fluxes of a few sources. This is due to very nearby sources, as illustrated in Fig. 9. The black circles represent the sources without any neighbours within 1.2'' (i.e., the pixel size). The red dots are the sources without any neighbours within 3.0''. We can see that the number of sources for which the XID+ flux is very different from the true flux is greatly reduced. So, as long as the sources are separated by at least two pixels, XID+ can correctly deblend them in most cases. Finally, the blue dots represent the sources without any neighbours within 5.7'' (i.e., the PSF FWHM). The

<sup>10</sup> In XID+, the map to be fitted is segmented into tiles using the Hierarchical Equal Area isoLatitude Pixelization of a sphere (HEALPix), as it is computationally unfeasible to process the whole map simultaneously. The tiles have equal areas determined by the HEALPix level.



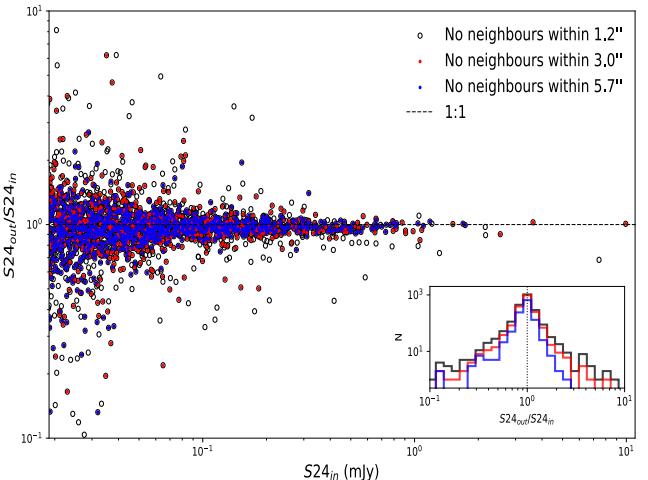
**Fig. 7.** Comparison of the sky ( $10' \times 10'$  cutouts) at different wavelengths. Top row: the same patch of the sky from  $24\text{ }\mu\text{m}$  to  $500\text{ }\mu\text{m}$  from the real observations (top). Bottom row: similar to the top row but for the SIDES simulation including Gaussian random noise. The simulation can statistically reproduce the observed far-IR and sub-mm sky reasonably well. One can also notice that the PACS 100 and 160  $\mu\text{m}$  maps are much more dominated by instrument noise compared to the maps at other wavelengths.



**Fig. 8.** Ratio of the output XID+ flux to the input true flux vs. the true flux at  $24\text{ }\mu\text{m}$ , down to the flux cut of  $10\text{ }\mu\text{Jy}$ . The blue line is the one-to-one correspondence. The vertical dotted line corresponds to the  $1\sigma$  instrument noise. The solid red line is the running median and the dashed red lines are the 16th and 84th percentiles.

deblended fluxes for these sources show further improvement, i.e., better agreement with the true fluxes.

Hurley et al. (2017) showed that for faint sources comparable to or below the noise level, their flux posterior distributions will be non-Gaussian, due to the limit imposed by noise. To illustrate this, we calculated from the posterior  $\sigma^+$  (from the difference between the 84th percentile and the median) and  $\sigma^-$  (from the difference between the median and 16th percentile). Figure 10 shows the ratio of  $\sigma^+$  to  $\sigma^-$  as a function of the median (i.e. what we assign as the best estimate of the XID+ deblended flux). The flux uncertainties become roughly Gaussian at fluxes above the  $1\sigma$  instrument noise. We also checked the accuracy of the flux uncertainty. First, we derived the  $1\sigma$  flux uncertainty estimate from the maximum of  $\sigma^+$  and  $\sigma^-$ . Following Liu et al. (2018) and Jin et al. (2018), we plot in Fig. 11 the probability distribution of the difference between the XID+ flux  $S_{24\text{out}}$  and the true flux  $S_{24\text{in}}$  divided by the  $1\sigma$  flux uncertainty, for sources with  $S_{24\text{out}} > 18.8\text{ }\mu\text{Jy}$  which is the  $1\sigma$  instrument noise. The distribution (the blue histogram) is wider than the standard Gaussian with a mean of zero and standard deviation of

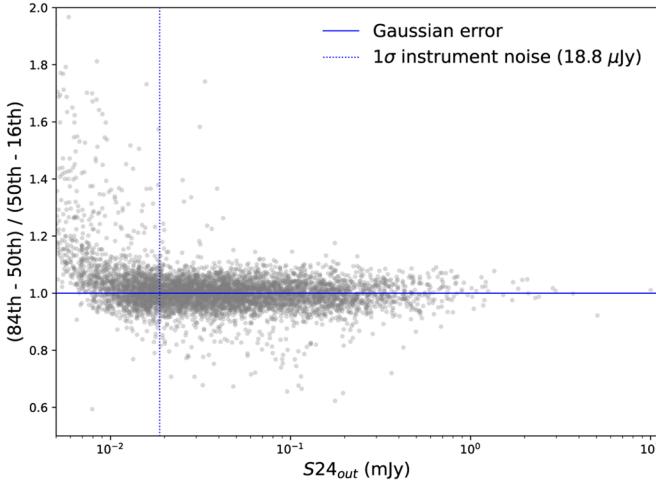


**Fig. 9.** Similar to Fig. 8, but only for sources which have true flux  $S_{24\text{in}} > 18.8\text{ }\mu\text{Jy}$  (i.e., above the  $1\sigma$  instrument noise level) and do not have any neighbours in the prior list within a certain radius (black circles – no neighbours within  $1.2''$ ; red dots – no neighbours within  $3.0''$ ; blue dots – no neighbours within  $5.7''$ ). The inset shows the histograms of the ratio of the output XID+ flux to the input true flux for the three groups.

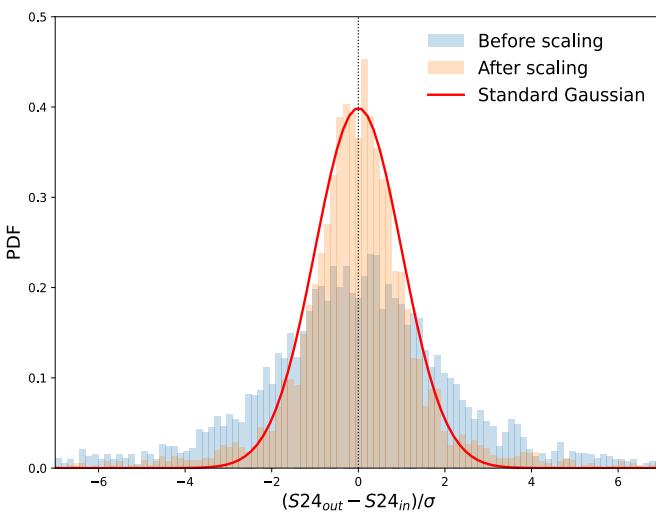
one, indicating that the flux uncertainties are underestimated<sup>11</sup>. If we multiply the  $1\sigma$  error by a factor of two, the distribution (the brown histogram) then closely follows the standard Gaussian. In fact, the same multiplicative scaling factor is also needed in the *Herschel* SPIRE bands. For a detailed explanation of why we recommend applying a scale factor of two to the  $1\sigma$  error in the MIPS and SPIRE bands, please refer to Appendix B.

To investigate how well constrained the XID+ deblended flux is (i.e. flux precision), we show in Fig. 12 the interquartile range (IQR) divided by the true flux and how it varies as a function of the true flux. The IQR is the difference between the 84th and 16th percentile of the posterior. To account for the systematic underestimation in the flux uncertainty, we multiply the IQR by a factor of two. The ratio of the IQR to the

<sup>11</sup> Based on a frequentist interpretation, we expect the estimated fluxes to fall within  $1\sigma$  roughly 68% of the time and within  $2\sigma$  roughly 95 of the time.



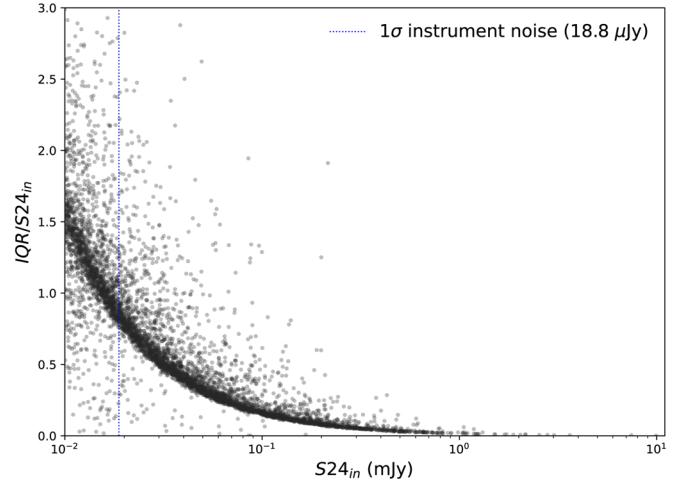
**Fig. 10.**  $\frac{84\text{th}-50\text{th}}{50\text{th}-16\text{th}}$  percentile ratio (i.e.  $\sigma^+/\sigma^-$ ) of the flux posterior distribution as a function of the output XID+ deblended flux. If the flux uncertainties are Gaussian, this ratio would be equal to one (the horizontal line), which is roughly the case at fluxes above the  $1\sigma$  instrument noise of  $18.8 \mu\text{Jy}$ .



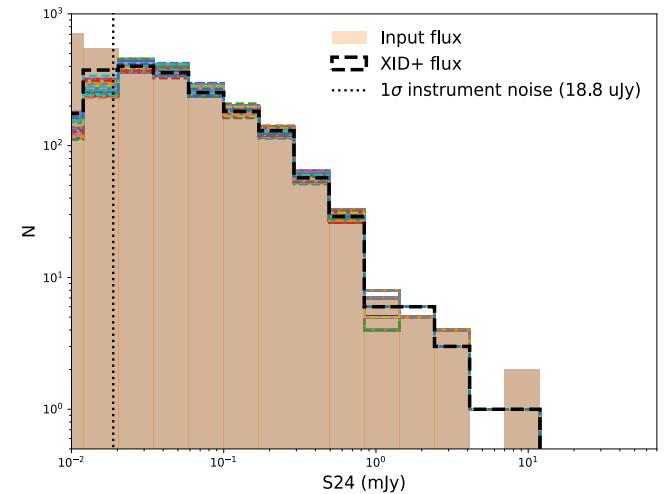
**Fig. 11.** Distribution of the flux difference (between the output XID+ flux and the true flux) normalised by the XID+ flux uncertainty estimate, before and after multiplying the  $1\sigma$  uncertainty by a factor of two. The red line corresponds to a standard Gaussian.

true flux generally decreases (indicating higher precision) for intrinsically brighter sources. When the sources are very faint (similar to the  $1\sigma$  instrument noise), their IQR values are comparable to (or even larger than) the true fluxes. Finally, we examined how well we can recover the true number counts. In Fig. 13, we show the true counts as the brown histogram which rise steeply with decreasing flux. The black dashed line represents the output counts derived from the best estimate of the deblended flux. The dashed lines with different colours correspond to the extracted counts using the 3000 samplings from the XID+ posterior probability distribution functions (PDFs). To account for the systematic flux uncertainty underestimation, we rescaled the 3000 samplings from the posterior for each source as follows,

$$S'_i = S_{50\text{th}} + 2 \times (S_i - S_{50\text{th}}), \text{ with } i = 1, \dots, 3000, \quad (6)$$



**Fig. 12.** Ratio of the IQR to the true flux vs. the true flux, after scaling by a factor of two. The vertical line is the  $1\sigma$  instrument noise.

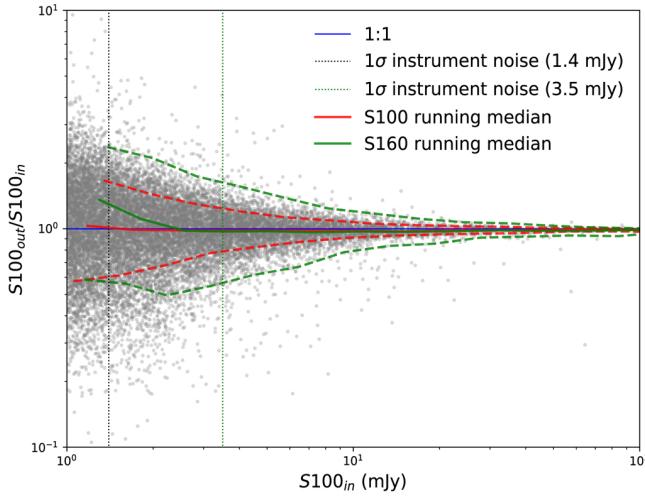


**Fig. 13.** Comparison of the true  $24 \mu\text{m}$  number counts (filled histogram) with the counts extracted from the output XID+ flux (black line). The dashed histograms of different colours represent the 3000 samplings from the posterior PDFs. The vertical dotted line is the  $1\sigma$  instrument noise.

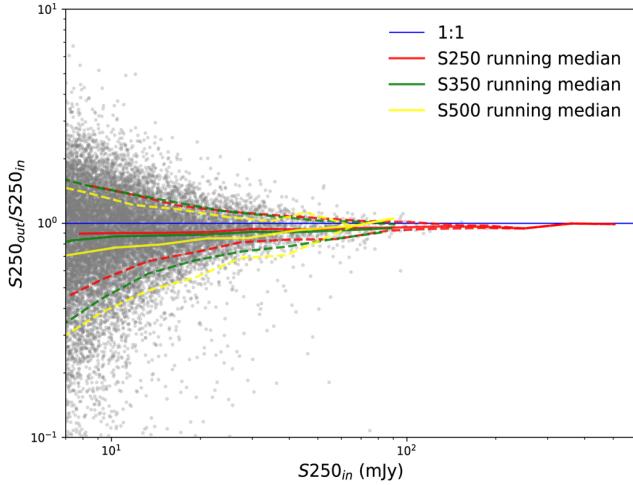
where  $S_{50\text{th}}$  is the median (i.e., the best estimate) of the source flux. The output counts are close to the true counts above the  $1\sigma$  instrument noise. Below the  $1\sigma$  instrument noise (where both flux accuracy and flux precision are poorer), the output counts fall increasingly under the true counts.

#### 4.2. Deblending results from the simulated Herschel PACS and SPIRE maps

The PACS maps are dominated by instrument noise. Therefore, we selected sources with a true  $100 \mu\text{m}$  flux  $> 1 \text{ mJy}$  (close to the  $1\sigma$  instrument noise) which gave us a total of 43 615 sources (0.10 sources per beam). Figure 14 compares the output to input flux ratio at  $100 \mu\text{m}$  as a function of the true  $100 \mu\text{m}$  flux. The red solid line is the running median and the red dashed lines correspond to the 16th–84th percentile ranges. We also overplot the running median and the 16th–84th percentiles at  $160 \mu\text{m}$  as the green solid line and green dashed lines, respectively. The black and green vertical dotted lines represent the instrument noise



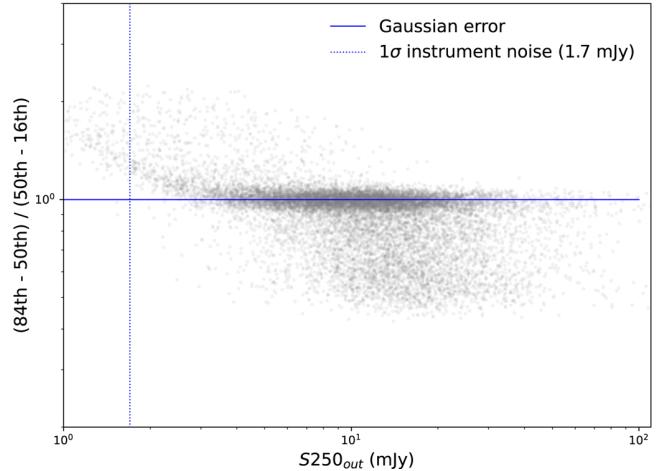
**Fig. 14.** Ratio of the output XID+ flux to the input flux vs. the input flux at 100  $\mu\text{m}$ , down to the flux cut of 1 mJy. The blue line is the one-to-one correspondence. The black and green vertical dotted line corresponds to the 1 $\sigma$  instrument noise level of 1.4 and 3.5 mJy at 100 and 160  $\mu\text{m}$  respectively. The red/green solid line is the running median at 100/160  $\mu\text{m}$ . The red/green dashed lines are the 16th and 84th percentiles at 100/160  $\mu\text{m}$ .



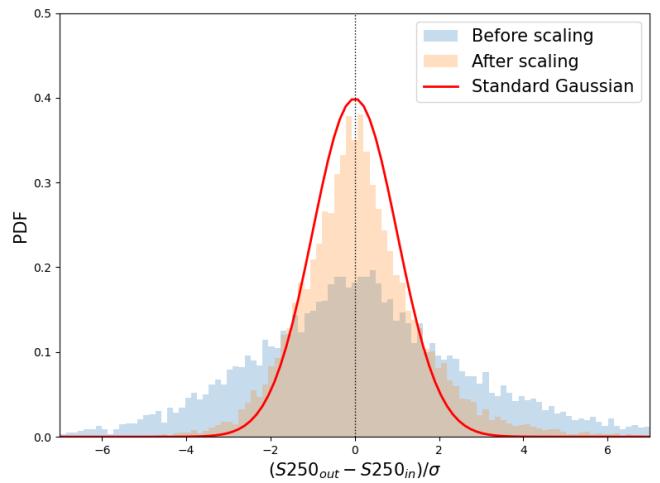
**Fig. 15.** Ratio of the output XID+ flux to the input true flux vs. the true flux at 250  $\mu\text{m}$ , down to the flux cut of 7 mJy. The blue line is the 1:1 ratio. The solid and dashed red/green/yellow lines are the 250/350/500  $\mu\text{m}$  running median and the 16th–84th percentiles, respectively.

level at 100 and 160  $\mu\text{m}$ , which is 1.4 and 3.5 mJy, respectively. Again, we find a very good agreement between the output flux and the true flux above 1 $\sigma$  instrument noise in both PACS bands.

In contrast to PACS, the SPIRE maps are dominated by confusion noise. Thus, we selected sources with a true 250  $\mu\text{m}$  flux  $> 7$  mJy (equal to the 1 $\sigma$  total noise =  $\sqrt{\sigma_{\text{instrument}}^2 + \sigma_{\text{confusion}}^2}$ ), which gave us 14 583 sources (0.54 sources per beam). Figure 15 compares the XID+ deblended flux to the input flux ratio as a function of the true 250  $\mu\text{m}$  flux. The red solid line represents the running median, while the red dashed lines are the 16th–84th percentiles. We over-plot the same quantities for the other two SPIRE bands. The 350 and 500  $\mu\text{m}$  running medians and percentiles have qualitatively the same behavior. At the bright end, the output flux is median unbiased relative to the true flux across the three bands as the running median is close to the 1:1.



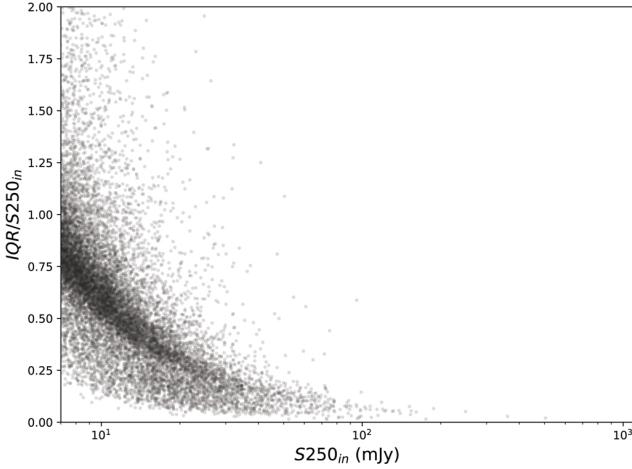
**Fig. 16.**  $\frac{84\text{th}-50\text{th}}{50\text{th}-16\text{th}}$  percentile ratio (i.e.  $\sigma^+/\sigma^-$ ) of the flux posterior as a function of the output XID+ deblended flux. If the flux uncertainties are Gaussian, this ratio would be equal to one (the horizontal blue line).



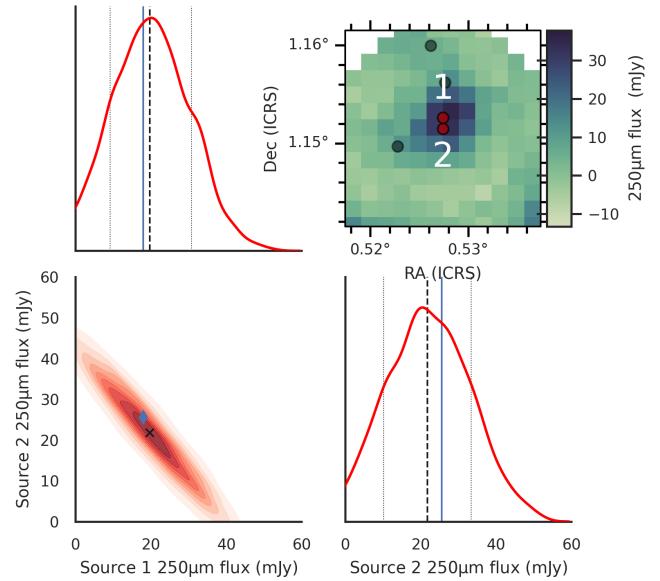
**Fig. 17.** Distribution of flux difference (between the output XID+ flux and the input true flux) normalised by the XID+ flux uncertainty estimate, before and after scaling by a factor of two. The red line corresponds to a standard Gaussian distribution.

Towards fainter fluxes, the median flux bias increases from a few percent to  $\sim 10\%$ ,  $\sim 15\%$  and  $\sim 25\%$  at the faintest flux level at 250, 350 and 500  $\mu\text{m}$ , respectively. Thus, the offset representing a systematic flux underestimation increases towards longer wavelengths, which is expected as the surface density of the sources in the prior catalogue increases by almost a factor of four from 250 to 500  $\mu\text{m}$ .

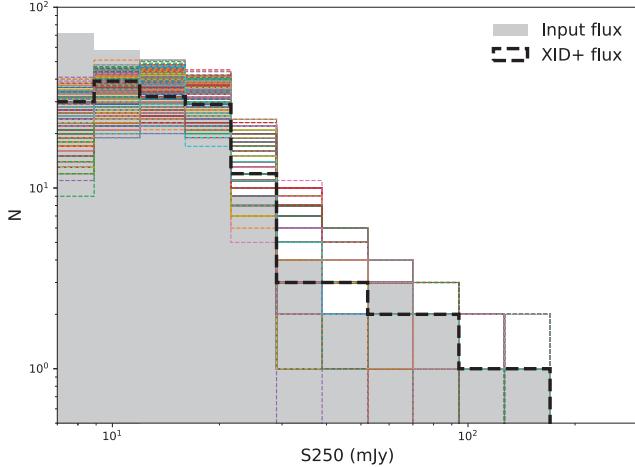
After demonstrating the level of flux accuracy in the SPIRE bands, we explored further issues such as accuracy of flux uncertainty, flux precision and number counts, using the 250  $\mu\text{m}$  band as an example. Figure 16 shows the  $\frac{84\text{th}-50\text{th}}{50\text{th}-16\text{th}}$  percentile ratio of the XID+ flux posterior distribution as a function of the median at 250  $\mu\text{m}$ . The flux uncertainties of most sources are roughly Gaussian. However, there are some relatively bright sources with a much lower  $\sigma^+$  flux uncertainty. On the other hand, there are some sources with a much lower  $\sigma^-$  flux uncertainty at the faint end. In Fig. 17, we plot the probability distribution of difference between the output flux  $S_{250\text{out}}$  and the true flux at 250  $\mu\text{m}$   $S_{250\text{in}}$  normalised by the 1 $\sigma$  flux uncertainty estimate (from the maximum of  $\sigma^+$  and  $\sigma^-$ ) for sources with  $S_{250\text{out}} > 7$  mJy,



**Fig. 18.** Ratio of the IQR to the input flux as a function of the true 250  $\mu\text{m}$  flux, after applying the scaling factor.



**Fig. 20.** Joint and marginalised posterior plot of two correlated sources ( $\sim 4''$  apart) in the simulated 250  $\mu\text{m}$  maps. Note that the joint and marginalised posteriors have been scaled by a factor of two using Eq. (6). The dashed black lines and the black cross indicate the fluxes extracted by XID+. The solid blue lines and the blue diamond indicate the true fluxes. The vertical dotted lines correspond to the 16th–84th percentiles in the posterior.



**Fig. 19.** Comparison of the input 250  $\mu\text{m}$  number counts (filled histogram) with the counts extracted from the output XID+ flux (black line). The dashed histograms of different colours represent the 3000 samplings from the posterior.

before and after scaling by a factor of two. Without scaling, the distribution is much wider than the standard Gaussian (the red line). After scaling by a factor of two, the distribution of the normalised flux difference is much closer to the standard Gaussian. Figure 18 shows that ratio of the IQR to the input true 250  $\mu\text{m}$  flux as a function of the true flux, demonstrating that flux precision is in general higher for brighter sources.

In Fig. 19, we show the input counts at 250  $\mu\text{m}$  as the grey histogram. The black dashed line represents the output counts derived from the best estimate of the output flux. The dashed histograms with different colours correspond to the extracted counts using the 3000 samplings from the posterior, scaled using Eq. (6). The output counts are close to the true counts above the 10 mJy. Because of the worse scatter and flux accuracy of the fainter sources, the output counts falls increasingly below the true counts at  $< 10$  mJy. Finally, we illustrate how posterior PDFs of the output flux determined by XID+ can be used to investigate sources which are very close to each other, as their flux densities will be highly correlated. In Fig. 20, we show the joint and marginalised posteriors of two close-by sources (separated by  $\sim 4''$  which is smaller than the pixel size of  $6''$ ) in the simulated 250  $\mu\text{m}$  map. The joint posterior is highly elongated which could

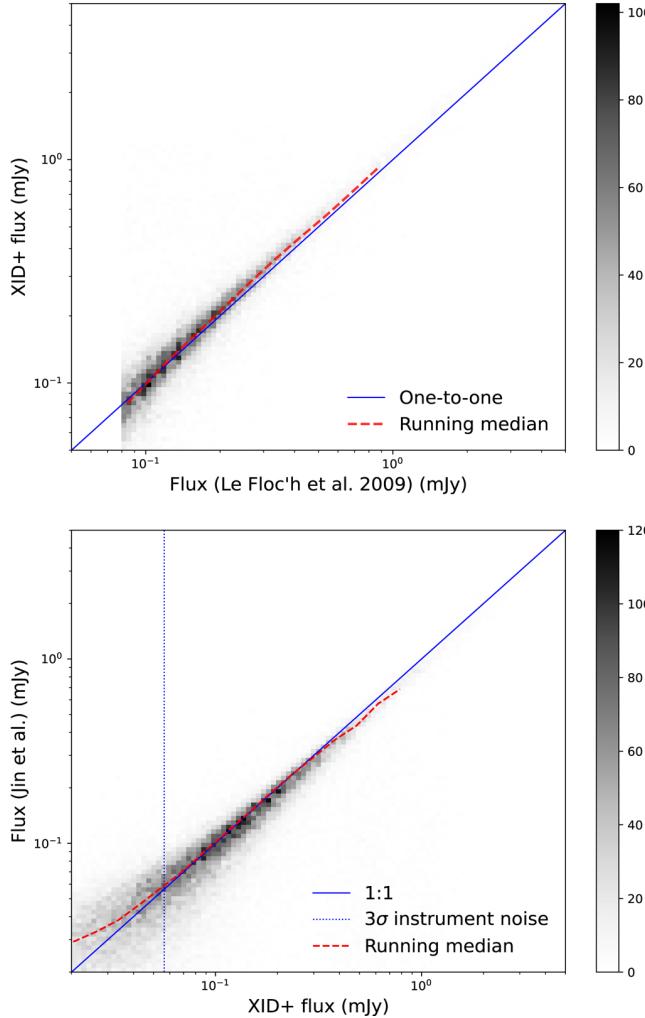
cause some source photometry methods to assign all of the flux to one source. With XID+, one can explore the correlation of the neighbouring sources.

## 5. The final deblended point source catalogue

After validating our deblending methodology using the SIDES simulations, we proceed to deblend the real maps from 24  $\mu\text{m}$  to 500  $\mu\text{m}$  and compare with blindly extracted catalogues which contain relatively bright sources and the super-deblended catalogue from Jin et al. (2018) which extends to very faint sources (down to  $\sim 1\sigma$  instrument noise). As an additional test, we deblend the SCUBA-2 850  $\mu\text{m}$  map and compare with ALMA measurements which should be free of source blending issues. This gives us an independent validation which does not rely on simulations (and any inbuilt assumptions).

### 5.1. Deblending the real Spitzer MIPS 24 $\mu\text{m}$ map

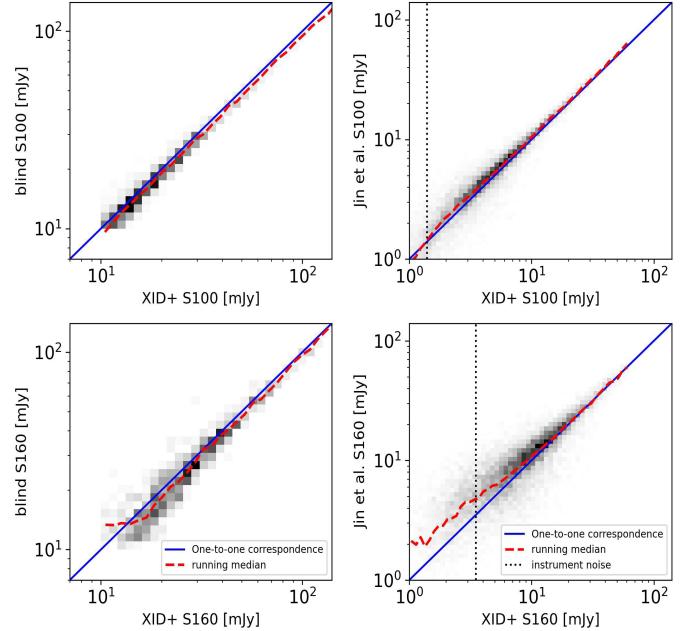
We first compare our XID+ deblended 24  $\mu\text{m}$  fluxes with the blindly extracted catalogue from (Le Floc'h et al. 2009) which contains  $\sim 30\,000$  sources down to the limit of 80  $\mu\text{Jy}$  (roughly  $4\sigma$  instrument noise). By construction, blind catalogues are limited to relatively bright sources in order to keep the rate of spurious detections at a minimum. We cross-matched our deblended catalogue with the blind catalogue by finding the closest match within  $2''$  which results in 21 703 matches. The top panel in Fig. 21 shows that there is a very good agreement between our XID+ deblended 24  $\mu\text{m}$  fluxes and the blindly extracted fluxes down to the limit of the blind catalogue. The median difference is 5.4  $\mu\text{Jy}$  which is much smaller than the  $1\sigma$  uncertainty of the blind catalogue. In terms of fractional difference (i.e., flux difference divided by the blind flux) the median difference is 3.3%.



**Fig. 21.** *Spitzer* MIPS 24  $\mu\text{m}$  flux comparison. Top panel: A 2D histogram plot comparing the XID+ deblended fluxes with the blind catalogue from Le Floc'h et al. (2009). There is a good agreement down to the limit of the blind catalogue (80  $\mu\text{Jy}$ ). Bottom panel: Comparison with the super-deblended catalogue from Jin et al. (2018). Our deblended fluxes agree quite well with the super-deblended fluxes, but with increasing scatter towards fainter sources.

We also compare our XID+ deblended catalogue with the super-deblended catalogue from Jin et al. (2018). There are several differences between the two deblended catalogues. First, the construction of the prior catalogue is different. Jin et al. (2018) used COSMOS2015 while we used the updated COSMOS2020 benefiting from deeper imaging and improved SED coverage. The initial selection of prior sources in Jin et al. (2018) is based on photo- $z$  and stellar mass estimates<sup>12</sup>. While a stellar mass-based selection is reasonable because of the SFMS which shows a strong correlation between star-formation rate (SFR) and stellar mass for star-forming galaxies, it also includes passive galaxies which are not significant far-IR/sub-mm emitters. The fraction of passive galaxies increases with increasing stellar mass and

<sup>12</sup> Jin et al. (2018) used photo- $z$  and stellar mass estimates in the initial selection of prior sources to deblend the 24  $\mu\text{m}$  map and the radio maps. After that, they used SED fitting and the photo- $z$  information to predict the flux of a given prior source at the wavelength to be deblended (e.g., PACS or SPIRE maps). If the predicted flux is above a critical flux value, then the source is kept in the prior list.

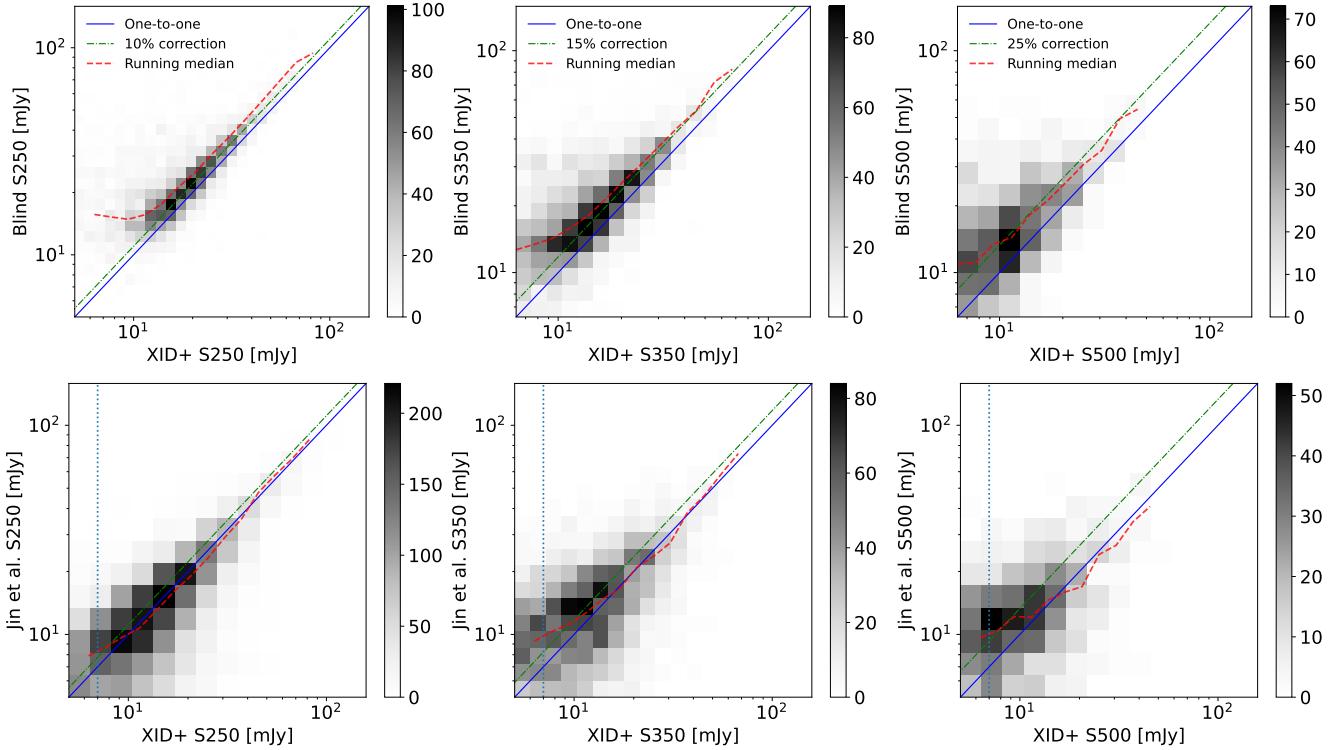


**Fig. 22.** *Herschel* PACS 100 and 160  $\mu\text{m}$  flux comparison. Left column: 2D histogram plots comparing the XID+ deblended fluxes with the blind catalogue. Right column: comparison with the super-deblended catalogue from Jin et al. (2018). The vertical dotted line corresponds to the 1 $\sigma$  instrument noise level, which is 1.4 mJy at 100  $\mu\text{m}$  and 3.5 mJy at 160  $\mu\text{m}$ .

decreasing redshift. Furthermore, relying directly on the photo- $z$  and stellar mass estimates would be problematic for some sources for which these properties could not be reliably derived. For example, from COSMOS2015 to COSMOS2020, more than 20% of the sources ( $\sim 30\,000$  sources) have their photo- $z$  estimates changed by  $>10\%$ . In our deblending methodology, we only used the observed fluxes and colours to predict the flux in the waveband to be deblended, taking into account systematic uncertainties in the SED modelling. Thus, we can avoid relying on derived properties and automatically fold in differences in the SEDs between star-forming galaxies and passive galaxies. There are also differences in the deblending methodology. Jin et al. (2018) used a maximum-likelihood based GALFIT PSF fitting and rely on Monte Carlo simulations to correct the output flux. The Monte Carlo simulations would break down the correlation between sources as the injected source has no correlation with the real sources to be deblended. Bearing these differences in mind, we can see a very good agreement in the bottom panel of Fig. 21 between the two deblended catalogues with increasing scatter towards fainter fluxes. Below the 3 $\sigma$  instrument noise level ( $3 \times 18.8 \mu\text{Jy}$ ), our XID+ deblended fluxes are slightly lower than the super-deblended fluxes, which could be due to the systematic underestimation effect (which is  $<5\%$  above 1 $\sigma$  instrument noise) as seen in the validations using simulations (Sect. 4.1).

## 5.2. Deblending the real *Herschel* PACS & SPIRE maps

First, we compare our deblended PACS catalogues with the blindly extracted catalogues from the PEP survey which contains 7443 sources with 100  $\mu\text{m}$  flux  $>4.4$  mJy. Selecting the closest match within 2'', we found a total of 3654 and 2489 matches at 100 and 160  $\mu\text{m}$ , respectively. The left column of Fig. 22 shows a good agreement on flux down to the limit of the blind catalogue



**Fig. 23.** *Herschel* SPIRE 250, 350 and 500  $\mu\text{m}$  flux comparison. Top panels: 2D histogram plots comparing the XID+ deblended SPIRE fluxes with the blind catalogue. Bottom panels: Comparison of the XID+ deblended SPIRE fluxes with the Jin et al. (2018) super-deblended catalogue. The blue line corresponds to the one-to-one correspondence. The red line is the shifted one-to-one line, after taking into account the maximum level of the systematic underestimation of our XID+ deblended SPIRE fluxes (see Sect. 4.2).

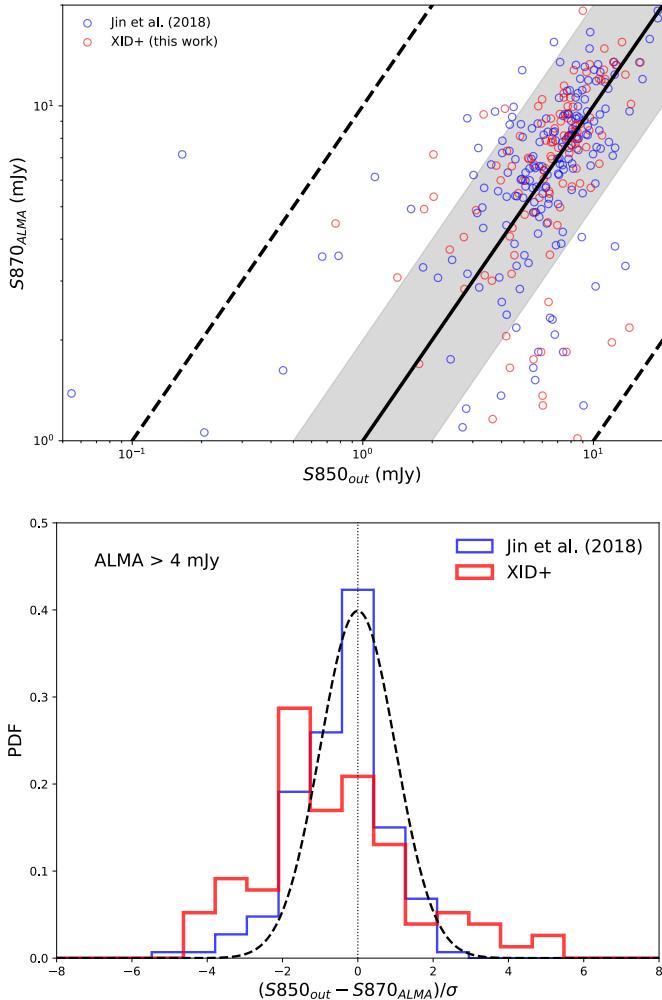
(5 and 10 mJy at 100 and 160  $\mu\text{m}$  respectively). In the right column of Fig. 22 we compare our results with the super-deblended catalogue. At 100  $\mu\text{m}$ , the flux agreement is excellent throughout the explored range. At 160  $\mu\text{m}$ , the super-deblended flux is slightly higher than our deblended flux at  $< \sim 10$  mJy. Based on the validation in Sect. 4.2, we expect our deblended flux to be median unbiased relative to the true flux down to the  $1\sigma$  instrument noise which is 3.5 mJy at 160  $\mu\text{m}$ . Therefore, we conclude that the Jin et al. (2018) super-deblended 160  $\mu\text{m}$  fluxes may be slightly overestimated in this flux range.

We then compare our XID+ deblended results of the SPIRE maps with the blind catalogue from the *Herschel* Extragalactic Legacy Project (HELP; Shirley et al. 2021). The sources in the blind catalogue are detected as peaks in the matched filter images which maximises the S/N of individual point sources taking into account instrument noise and confusion noise (Chapin et al. 2011). The blind catalogue in COSMOS contains 9382 sources directly detected from the 250  $\mu\text{m}$  map, with a flux cut  $S_{250\mu\text{m}} \geq 13.5$  mJy corresponding to the 85% completeness level. Selecting the closest match within 5'' based on the positional uncertainty expected of blind SPIRE sources (Wang et al. 2014), we found a total of 3591 matches. The top row in Fig. 23 shows the comparison of our deblended fluxes with the blind photometry, revealing a generally good agreement particularly at 250  $\mu\text{m}$ . There is a systematic underestimation in our deblended flux which increases with increasing wavelength. The scatter also increases toward fainter fluxes and longer wavelengths. This is expected as the density of the sources in the prior catalogue increases from 250 to 500  $\mu\text{m}$  (from 0.34 to 1.34 sources per beam). Based on the validation in Sect. 4.2, we expect that there is a maximum median offset of  $\sim 10\%$ ,  $15\%$  and  $25\%$  at 250, 350 and 500  $\mu\text{m}$ , respectively, at the faintest flux levels. We add

the red dashed lines to guide the eye in gauging the maximum level of median bias in our deblended flux relative to the blind photometry. The bottom row of Fig. 23 compares our deblended SPIRE fluxes with the Jin et al. (2018) super-deblended catalogue. At 250  $\mu\text{m}$ , we see a good agreement between our deblended fluxes and the super-deblended fluxes throughout the explored flux range. At 350 and 500  $\mu\text{m}$ , the agreement between the two deblended catalogues is slightly worse (with larger scatter). The running median lines cross the maximum level of median bias (based on the validation in Sect. 4.2) around 7 mJy. Overall, the good agreement of the two deblended catalogues demonstrate the good quality and robustness of the deblended photometry.

### 5.3. Deblending the real SCUBA 850 $\mu\text{m}$ maps and comparison with ALMA photometry

We also apply our XID+ deblending to the SCUBA-2 850  $\mu\text{m}$  map and then compare with ALMA measurements. The ALMA measurements regarded as the truth allow us to independently verify the quality of our deblended photometry. We used the 850  $\mu\text{m}$  map from the SCUBA-2 (Holland et al. 2013) COSMOS survey (S2COSMOS; Simpson et al. 2019), conducted with the East Asian Observatory's James Clerk Maxwell Telescope (JCMT). Our prior catalogue is the same as used for deblending the SPIRE maps. For discussions on completeness and precision in using  $K$ -band/mid-IR/radio detected sources as counterparts of sub-mm sources, we refer the reader to Hodge et al. (2013); An et al. (2018, 2019). The median instrumental noise level of the 850  $\mu\text{m}$  map is 1.2 mJy/beam over the main area of 1.6 deg $^2$  (with the deepest area reaching  $\sigma_{\text{inst}} \sim 0.5$  mJy/beam $^{-1}$ , which is where our prior catalogue is located. We used a Gaussian



**Fig. 24.** Testing deblended 850  $\mu\text{m}$  flux densities using AS2COSMOS. Top panel: Comparison between deblended 850  $\mu\text{m}$  fluxes (red: XID+; blue: Jin et al. 2018) and ALMA 870  $\mu\text{m}$  measurements. The solid line corresponds to the 1:1 ratio. The shaded region is within a factor of two from the 1:1 and the dashed lines are within a factor of 10. Bottom panel: probability distributions of the difference between the deblended fluxes and ALMA measurements normalised by the  $1\sigma$  flux uncertainty, for sources with ALMA measured fluxes  $> 4 \text{ mJy}$ .

PSF with a FWHM of  $14.9''$ . Following Geach et al. (2017) and Simpson et al. (2019), we also applied a multiplicative factor of 1.13 to account for the flux loss due to filtering in map making.

We compare our XID+ deblended fluxes with ALMA Band 7 continuum measurements centred at 870  $\mu\text{m}$  of  $\sim 180$  brightest sources from S2COSMOS, undertaken as a pilot of the ALMA-S2COSMOS (AS2COSMOS) survey (Simpson et al. 2020). No correction was applied to convert the ALMA fluxes to 850  $\mu\text{m}$ . Benefiting from the high resolution (with a median synthesised beam of  $0.80'' \times 0.79''$ ) and sensitivity of interferometric observations (with a median sensitivity of  $0.19 \text{ mJy/beam}^{-1}$ ), ALMA measurements are free from issues such as blending of multiple sources and are highly complete at  $> 6.2 \text{ mJy}$ <sup>13</sup>. Within  $1''$ , we found 113 matches between our deblended catalogue and the AS2COSMOS catalogue. The top panel of Fig. 24 compares

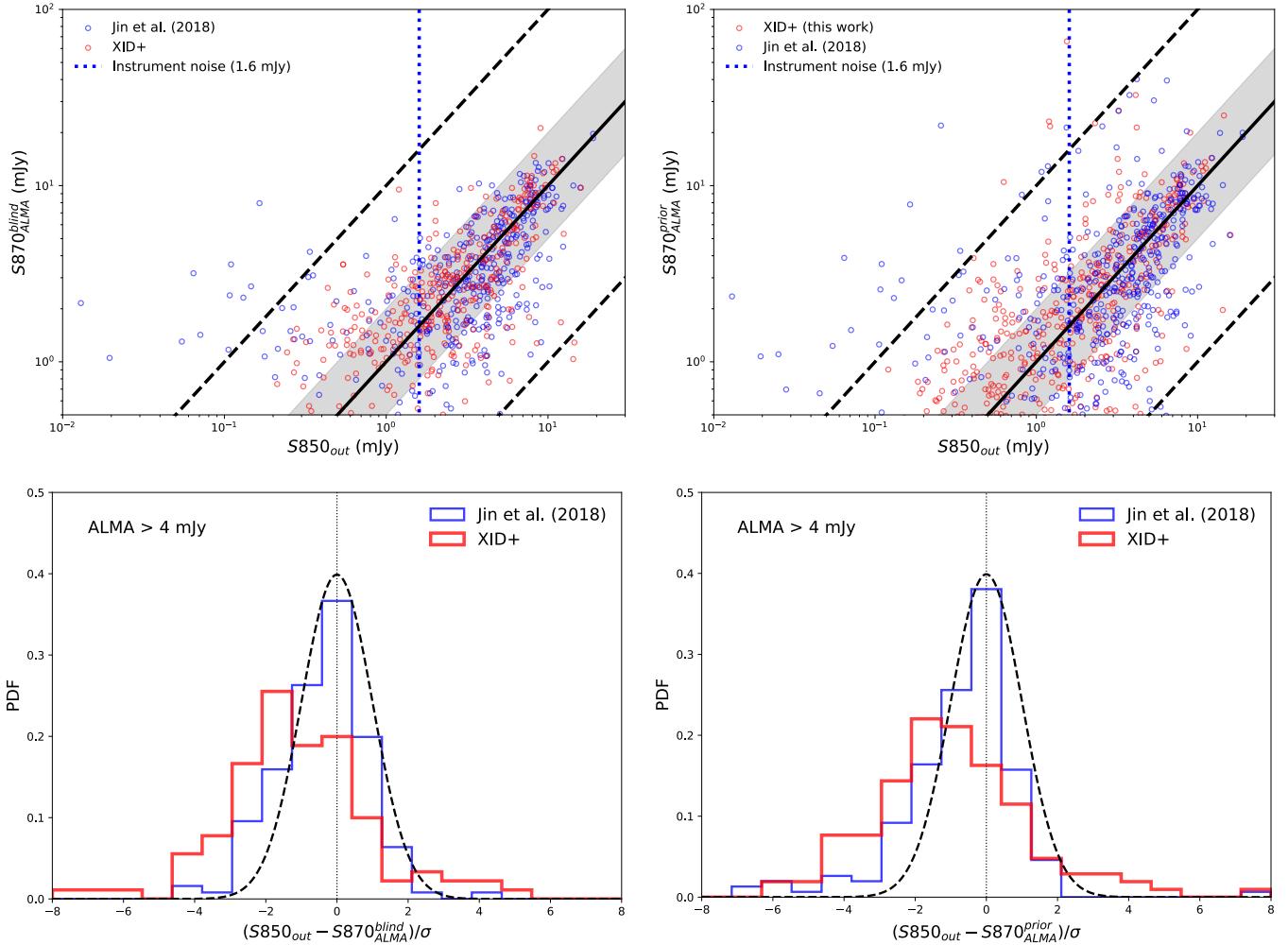
<sup>13</sup> The main limiting factor in the completeness level is the completeness of the parent SCUBA-2 sample, which is estimated to be 87%.

ALMA fluxes with our deblended fluxes. The bottom panel of Fig. 24 shows the probability distribution of the flux difference normalised by the uncertainty estimates of our deblended fluxes, for sources with ALMA measured fluxes  $> 4 \text{ mJy}$ . We also compare with the super-deblended photometry from Jin et al. (2018) which used the images from the SCUBA-2 Cosmology Legacy Survey (S2CLS) COSMOS program (Geach et al. 2017) with an rms noise of  $1.6 \text{ mJy/beam}^{-1}$ . Within  $1''$ , we got 201 matches between the super-deblended catalogue and the AS2COSMOS catalogue. The larger number of matches is mostly due to the larger area covered by the super-deblended catalogue. Overall, the level of agreement between our deblended fluxes and the ALMA measurements is similar to that between the super-deblended fluxes and the ALMA measurements. In the case of our deblended photometry, 92% of the matches (with ALMA fluxes  $> 4 \text{ mJy}$ ) agree with ALMA fluxes within a factor of two (i.e., in the shaded region). In the case of the super-deblended photometry, 88% of the matches (with ALMA fluxes  $> 4 \text{ mJy}$ ) agree with ALMA fluxes within a factor of two. There are a few sources with much higher ALMA fluxes than the corresponding super-deblended fluxes. Two sources even have  $> 10$  times brighter ALMA fluxes, suggesting that their super-deblended fluxes are significantly underestimated.

As a final and complementary test, we compare with measurements from the A3COSMOS project (Liu et al. 2019), which processes ALMA archival observations in COSMOS every six months. The newest data release, including all data which were made public before 10/03/2020, covers  $622.8 \text{ arcmin}^2$ . The A3COSMOS sample is much larger than AS2COSMOS, but has a more heterogeneous selection and mix of data with different depths and resolutions. The A3COSMOS data release includes two catalogues. One is a catalogue of blindly extracted ALMA sources with peak  $S/N > 5.4$ , with a  $\sim 50\%$  spurious rate at  $S/N = 5.4$  and a  $\sim 12\%$  cumulative spurious rate at  $S/N > 5.4$ . The other is a prior photometry catalogue containing fitting of known optical/near-IR/mid-IR/radio sources with  $S/N$  peak  $> 4.35$  and  $\sim 50\%$  spurious rate at  $S/N = 4.35$  and a  $\sim 8\%$  cumulative spurious rate for  $S/N > 4.35$ . As A3COSMOS includes data at various bands, we selected measurements with observed wavelengths between 850 and 890  $\mu\text{m}$ , comparable to the Band 7 measurements from AS2COSMOS.

First, we matched our XID+ deblended 850  $\mu\text{m}$  catalogue with the A3COSMOS blind catalogue by finding the closest match within  $0.5''$ , resulting in 354 matches. In comparison, there are 457 matches between the super-deblended catalogue and the blind catalogue. The top-left panel of Fig. 25 compares the deblended fluxes with the blind ALMA photometry. For the super-deblended fluxes, 87% of the matches (with ALMA fluxes  $> 4 \text{ mJy}$ ) agree with the ALMA blind photometry within a factor of two (i.e. in the shaded region). For the XID+ deblended fluxes, 90% of the matches (with ALMA fluxes  $> 4 \text{ mJy}$ ) agree with the ALMA blind photometry within a factor of two, demonstrating a higher level of flux accuracy. Compared to the XID+ deblended fluxes, more super-deblended fluxes show significant underestimation relative to ALMA. In the bottom-left panel of Fig. 25, we show that the probability distributions of the flux difference normalised by the  $1\sigma$  flux uncertainty estimate, for sources with ALMA measured fluxes  $> 4 \text{ mJy}$ .

We now compare with the ALMA prior photometry catalogue. We matched our XID+ deblended catalogue with the ALMA prior catalogue by finding the closest match within  $0.5''$ , resulting in 788 matches. In comparison, there are 810 matches between the super-deblended catalogue and the ALMA prior



**Fig. 25.** Testing deblended 850  $\mu\text{m}$  flux densities using A3COSMOS. Top left: comparison between deblended 850  $\mu\text{m}$  flux and ALMA 870  $\mu\text{m}$  photometry from the blind ALMA catalogue (red: XID+ deblended fluxes; blue: super-deblended fluxes from Jin et al. 2018). The blue vertical dotted line corresponds to the rms noise level of S2CLS-COSMOS. Top right: similar to the top left panel but for comparison with the ALMA prior photometry measurements. Bottom left: normalised distributions of flux difference (between deblended fluxes and ALMA blind photometry) divided by the  $1\sigma$  flux uncertainty, for sources with ALMA measured fluxes  $>4$  mJy. The vertical dotted line corresponds to no difference between the deblended flux and the ALMA flux. Bottom right: similar to the bottom left panel but for comparison with the ALMA prior photometry catalogue.

catalogue. The right column of Fig. 25 compares the deblended fluxes with the fluxes from the ALMA prior catalogue. Qualitatively, we see similar patterns as in the left column when comparing with the ALMA blind catalogue. For the super-deblended fluxes, 75% of the matches (with ALMA fluxes  $>4$  mJy) agree with the ALMA prior photometry within a factor of two. For the XID+ deblended fluxes, 77% of the matches (with ALMA fluxes  $>4$  mJy) agree with the ALMA prior photometry within a factor of two. Our deblended fluxes clearly extend to fainter sources than the super-deblended fluxes, which could be mostly due to the deeper 850  $\mu\text{m}$  map from S2COSMOS compared to S2CLS-COSMOS. Other differences in the prior catalogues and deblending methodologies might also play a role.

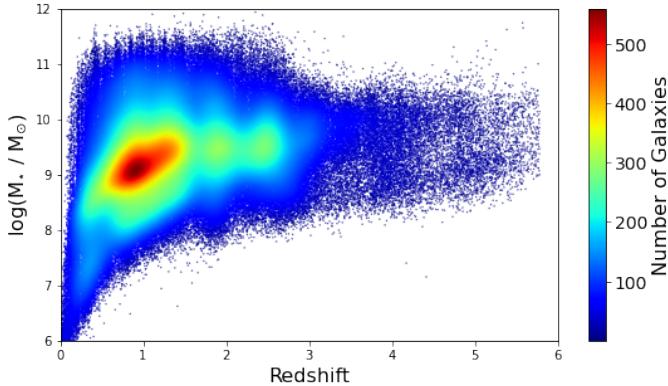
## 6. Example science applications

Our deblended far-IR and sub-mm photometry catalogue can be used in a wide range of studies related to dusty galaxy evolution. For example, one could investigate the number counts, the monochromatic luminosity functions as well as the integrated IR luminosity functions, and the contribution from dusty galaxies to

the cosmic star-formation history. These long-wavelength data are also critical for studying the SMF of dusty sources and how they compare with the general SMFs, the fraction of DSFGs as a function of stellar mass and redshift, and the fraction of dust-obscured SFR as a function of stellar mass and redshift. Here we present two example science applications, which are the galaxy SFMS and the FIRCS.

### 6.1. The galaxy star-formation main sequence

The SFMS is a tight correlation between stellar mass ( $M_\star$ ) and SFR from the local Universe to high redshift (Brinchmann et al. 2004; Noeske et al. 2007; Speagle et al. 2014; Popesso et al. 2023). To estimate  $M_\star$  and SFR, we used CIGALE and the setup of Pearson et al. (2018), with a delayed exponentially declining star-formation history plus an exponentially declining burst, Bruzual & Charlot (2003) stellar population, a double power law dust attenuation model, the Draine et al. (2014) dust emission model, and the Fritz et al. (2006) AGN templates. We then selected galaxies for which the reduced  $\chi^2$  of the best fit model is  $< 5$ . Fig. 26 shows the distribution of the sources in



**Fig. 26.** Stellar mass  $M_*$  vs. redshift for all galaxies in our deblended catalogue, with colour-coding corresponding to number of sources.

our deblended catalogue in stellar mass vs. redshift. Using the derived estimates of stellar mass and SFR, we plot the correlation between the two properties in six redshift bins from  $z \sim 0.2$  to  $z \sim 2.3$  in Fig. 27 and compare with the SFMS relations from Pearson et al. (2018), which were derived using COSMOS2015 (Laigle et al. 2016) and a less advanced deblending method for only the SPIRE data (Pearson et al. 2017, 2018). Thus, Pearson et al. (2018) is an ideal comparison to check if there are any differences our updated deblending makes to the SFMS. We also make a comparison with Speagle et al. (2014), which compiled a large number of SFMS studies made before their publication, and Popesso et al. (2023), who also made a compilation of earlier works but used a SFMS with a high mass turnover. Our galaxies in Fig. 27 are selected as star-forming using a UVJ colour-colour cut (Whitaker et al. 2011; Pearson et al. 2018).

Since this exercise was only meant to be a simple demonstration of what kind of scientific studies would be enabled by our data, we did not attempt a detailed analysis which would demand carefully taking into account various selection effects. With these caveats in mind and based on a simple visual inspection, we find that our new deblended data appear to give rise to a steeper SFMS than Pearson et al. (2018) at  $z < 1.8$  and more in line with Speagle et al. (2014) and Popesso et al. (2023). This work also appears to show that the normalisations (the value of the SFMS at  $\log(M_*/M_\odot) = 10.5$ ) of the SFMS are now more consistent with Speagle et al. (2014) and Popesso et al. (2023), at least at  $z < 1.4$ , partially resolving the known lower normalisation of the Pearson et al. (2018) SFMS described in Popesso et al. (2019).

## 6.2. The far-IR to radio correlation

The correlation between the far-IR and radio emission in star-forming galaxies is one of the tightest known in astronomy. Although the FIRC has been studied extensively at various IR and radio frequencies, the exact physical origin is not yet known and it is still under debate whether it evolves with redshift and other physical parameters such as stellar mass. As a second example, we briefly explore the FIRC using our deblended fluxes of the MIGHTEE sources at 1.3 GHz. Here we focus on the 1038 MIGHTEE sources that are not included in COSMOS2020. We emphasise again that in deblending the FIR/sub-mm maps for the additional radio sources, only positional priors are used (i.e., no information regarding their radio fluxes are used in the deblending process).

It is common to use a dimensionless parameter  $q$ , defined as the logarithmic ratio of the far-IR luminosity to the radio luminosity in the rest-frame, to describe the FIRC. Different studies often also use slight variations of  $q$ , e.g., by using either integrated or monochromatic (far-)IR luminosity. Also, the choice of the radio frequency can vary. Here we compute a monochromatic  $q_{250}$  as below,

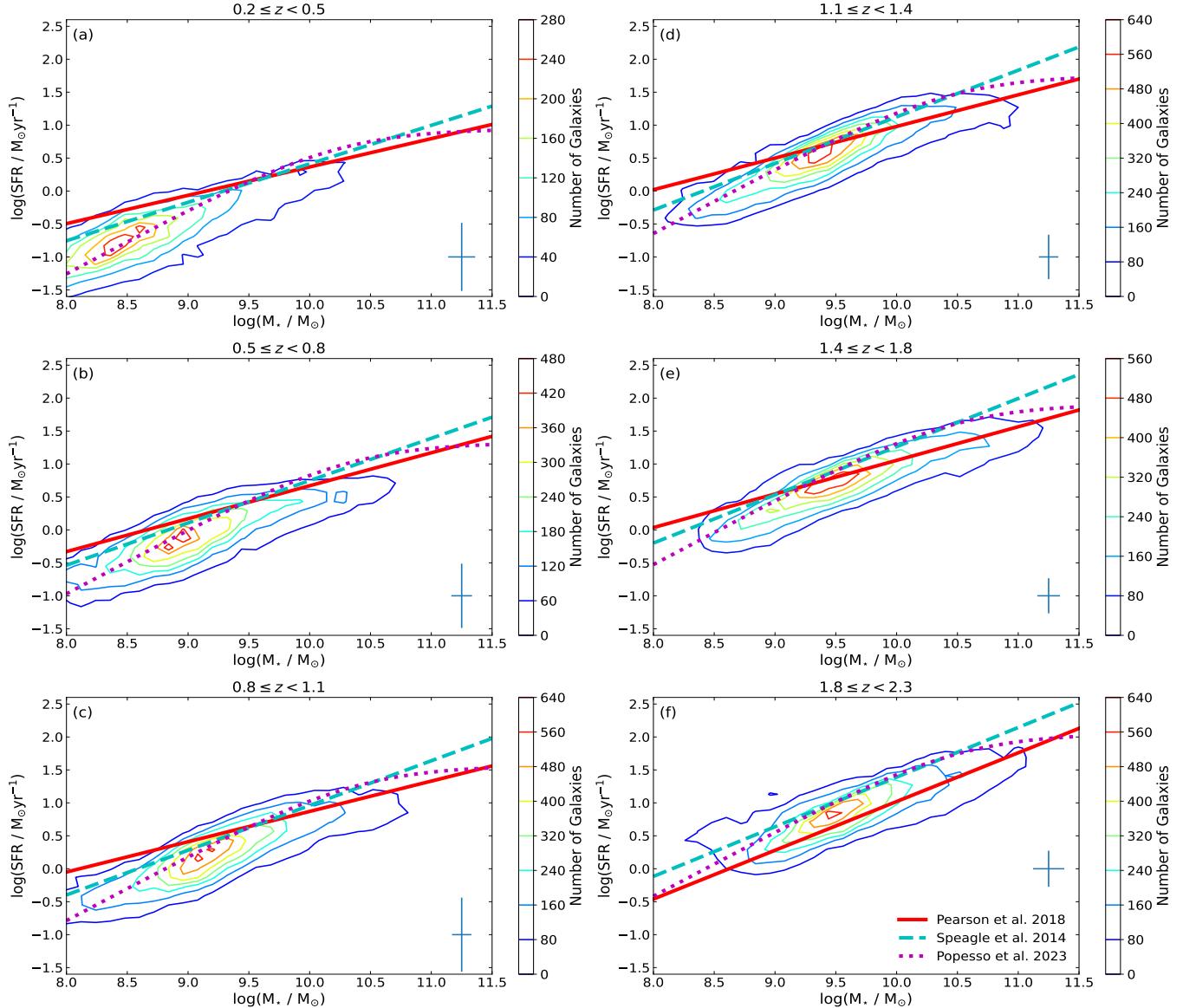
$$q_{250} = \log_{10} \left( \frac{S_{250}}{S_{1.3\text{GHz}}} \right). \quad (7)$$

Ideally, one need to apply  $k$ -correction to derive the rest-frame 250  $\mu\text{m}$  and 1.3 GHz flux densities, particularly if the sample extends a wide redshift range. However, as our sample consists of the 1038 MIGHTEE sources that are not present in COSMOS2020, we do not have photo- $z$  information for these sources. Therefore, we simply calculate the  $q_{250}$  value in the observed frame.

Figure 28 shows the distribution of our measured monochromatic  $q_{250}$  values for the MIGHTEE sources that do not have a match in COSMOS2020. We also compare with similar measurements from previous studies of the FIRC at 1.4 GHz. For example, Ivison et al. (2010) used multi-wavelength data across the redshift range  $0 < z < 3$  in the extended Chandra Deep Field South field, including 250, 350 and 500  $\mu\text{m}$  data from the Balloon-borne Large Aperture Submillimetre Telescope (BLAST; Devlin et al. 2009) and 1.4 GHz data from the VLA (Miller et al. 2008). For their 250  $\mu\text{m}$  selected sample, they measured  $q_{250} = 2.26 \pm 0.35$ , without applying  $k$ -correction. Using a stacking analysis on their 24  $\mu\text{m}$  selected sample, they measured  $q_{250} = 2.70 \pm 0.08$  without applying  $k$ -correction. Jarvis et al. (2010) studied the FIRC at  $z < 0.5$  using the *Herschel* Astrophysical TeraHertz Large Area Survey (H-ATLAS; Eales et al. 2010) Science Demonstration Phase (SDP) data and radio data from the NRAO VLA Northern Sky Survey (NVSS; Condon et al. 1998) and the Faint Images of the Radio Sky at Twenty-one centimetres (FIRST; Becker et al. 1995) and found a mean value of  $2.01 \pm 0.04$  for  $q_{250}$ . Smith et al. (2014) used a sample of 250  $\mu\text{m}$ -selected galaxies at  $z < 0.5$  from H-ATLAS and cross-matched with the 1.4 GHz flux density measurements from FIRST. They reported a value around 2.5 for  $q_{250}$  based on a stacking analysis. Finally, Gürkan et al. (2018) selected a sample of galaxies from the Sloan Digital Sky Survey (SDSS DR7; Abazajian et al. 2009). Matching the sample to the far-IR data from the H-ATLAS survey in the North Galactic Pole (NGP) field and the 1.4 GHz data from FIRST, they measured a best-fit value of  $\sim 1.95 \pm 0.2$  for  $q_{250}$ . Although it is difficult to make proper comparisons between our measured  $q_{250}$  values and the previous results due to various differences (such as the probed redshift range, sample selection, the adopted radio frequency and whether  $k$ -correction is applied), we see a reasonably good agreement. This demonstrates yet again the high quality of our deblended far-IR and sub-mm photometry.

## 7. Conclusions

Far-IR and sub-mm data are key to measuring the obscured star-formation activity. The star-formation properties of very dusty galaxies (for which the commonly adopted energy balance assumption could also break down) cannot be constrained reliably without far-IR/sub-mm data. The dusty galaxy population is also essential in understanding the formation and evolution of massive galaxies, as most of the massive population are shown to be dusty at high redshifts. However, there is currently no



**Fig. 27.** SFR as a function of  $M_*$  for (a)  $0.2 \leq z < 0.5$ , (b)  $0.5 \leq z < 0.8$ , (c)  $0.8 \leq z < 1.1$ , (d)  $1.1 \leq z < 1.4$ , (e)  $1.4 \leq z < 1.8$ , (f)  $1.8 \leq z < 2.3$ . Red lines are the results from Pearson et al. (2018), dashed cyan lines are the results from Speagle et al. (2014), and dotted magenta lines are the results from Popesso et al. (2023). Blue cross indicates the typical uncertainty in  $M_*$  and SFR. Colour-coding correspond to number of sources from low (dark blue) to high (red). Data are binned into the same redshift bins as in Pearson et al. (2018).

planned IR mission until the 2030s at the earliest. In this paper, we present a new deblended far-IR and sub-mm point source catalogue derived using a probabilistic and progressive deblending approach. We have combined the strengths of previous efforts and also improved on several aspects including the construction of the prior catalogue and taking into account the correlations between bands. Crucially, the construction of the prior benefits from an improved multi-wavelength catalogue in COSMOS and from a deep learning based emulator to speed up the flux prediction step which needs to be carried out progressively.

Based on testing using realistic simulations of the far-IR and sub-mm sky, we have demonstrated the good performance of our deblending methodology. For the *Spitzer*/MIPS 24  $\mu\text{m}$  map and the *Herschel*/PACS 100 and 160  $\mu\text{m}$  maps which are dominated by instrument noise, we can deblend fluxes which are median unbiased and have approximately Gaussian uncertainties all the way down to the  $1\sigma$  instrument noise. For the *Herschel*/SPIRE maps which are dominated by confusion noise, we show that

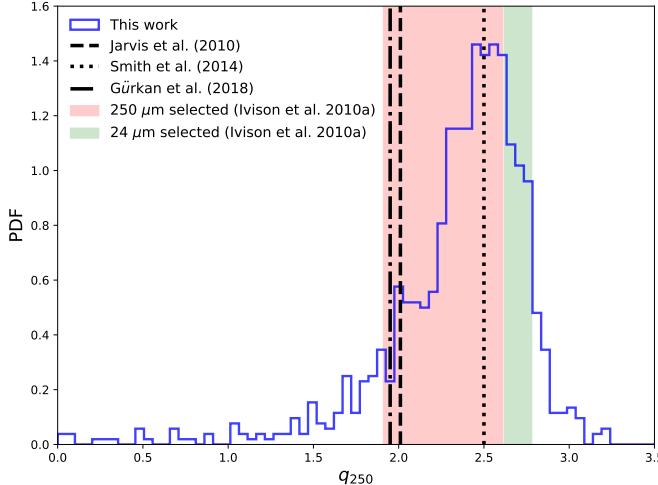
with our deblending methodology we can deblend fluxes down to roughly the  $1\sigma$  confusion noise. There is a slight systematic flux underestimation towards faint flux levels, which reaches approximately 10%, 15% and 25% at 250, 350 and 500  $\mu\text{m}$ , respectively, for the faintest sources. After testing with simulations, we deblend the real data and compare with previous blindly extracted catalogues and the super-deblended catalogue from Jin et al. (2018). As an additional test, we deblend the SCUBA-2 850  $\mu\text{m}$  and test our deblended photometry using ALMA Band 7 continuum measurements. We publicly release our XID+ deblended far-IR and sub-mm point source catalogues through the Strasbourg astronomical Data Center (CDS) and the *Herschel* Extragalactic Legacy Project database<sup>14</sup>. In Table 4, we provide an explanation of the columns contained in the catalogue.

<sup>14</sup> [https://hedam.lam.fr/HELP/dataproducts/dmu26/dmu26\\_XID+COSMOS2024/](https://hedam.lam.fr/HELP/dataproducts/dmu26/dmu26_XID+COSMOS2024/)

**Table 4.** Columns contained in our XID+ deblended far-IR and sub-mm point source catalogue.

Name	Unit	Description
ID	–	The COSMOS2020 ID (negative numbers for radio sources)
RA	–	Right Ascension from COSMOS2020 (or radio positions)
Dec	–	Declination from COSMOS2020 (or radio positions)
F_24	mJy	24 $\mu\text{m}$ flux density (median)
FErr_24_u	mJy	24 $\mu\text{m}$ flux density (84th Percentile); $\sigma^+ = \text{FErr\_24\_u} - \text{FErr\_24}$
FErr_24_l	mJy	24 $\mu\text{m}$ flux density (16th Percentile); $\sigma^- = \text{FErr\_24} - \text{FErr\_24\_l}$
FErr_24_1 $\sigma$	mJy	maximum of $\sigma^+$ and $\sigma^-$
Bkg_24	mJy/Beam	Fitted Background of 24 $\mu\text{m}$ map (median)
Sig_conf_24	mJy/Beam	Fitted residual noise component due to confusion (median)
Sig_tot_24	mJy/Beam	total error = $\sqrt{(\text{Sig\_conf\_24})^2 + (\text{FErr\_24\_1}\sigma)^2}$
Rhat_24	–	Convergence Statistic (ideally < 1.2)
n_eff_24	–	Number of effective samples (ideally > 40)
Post_24	mJy	3000 samplings from the posterior PDF of the 24 $\mu\text{m}$ flux density
tile_MIPS	–	tile number
F_100/160	mJy	100/160 $\mu\text{m}$ flux density (median)
FErr_100/160_u	mJy	100/160 $\mu\text{m}$ flux density (84th Percentile)
FErr_100/160_l	mJy	100/160 $\mu\text{m}$ flux density (16th Percentile)
FErr_100/160_1 $\sigma$	mJy	maximum of $\sigma^+$ and $\sigma^-$
Bkg_100/160	mJy/Beam	Fitted Background of 100/160 $\mu\text{m}$ map (median)
Sig_conf_100/160	mJy/Beam	Fitted residual noise component due to confusion (median)
Sig_tot_100/160	mJy/Beam	total error = $\sqrt{(\text{Sig\_conf\_100/160})^2 + (\text{FErr\_100/160\_1}\sigma)^2}$
Rhat_100/160	–	Convergence Statistic (ideally < 1.2)
n_eff_100/160	–	Number of effective samples (ideally > 40)
Post_100/160	mJy	3000 samplings from the posterior PDF of the 100/160 $\mu\text{m}$ flux density
tile_PACS	–	tile number
F_250/350/500	mJy	250/350/500 $\mu\text{m}$ flux density (median)
FErr_250/350/500_u	mJy	250/350/500 $\mu\text{m}$ flux density (84th Percentile)
FErr_250/350/500_l	mJy	250/350/500 $\mu\text{m}$ flux density (16th Percentile)
FErr_250/350/500_1 $\sigma$	mJy	maximum of $\sigma^+$ and $\sigma^-$
Bkg_250/350/500	mJy/Beam	Fitted Background of 250/350/500 $\mu\text{m}$ map (median)
Sig_conf_250/350/500	mJy/Beam	Fitted residual noise component due to confusion (median)
Sig_tot_250/350/500	mJy/Beam	total error = $\sqrt{(\text{Sig\_conf\_250/350/500})^2 + (\text{FErr\_250/350/500\_1}\sigma)^2}$
Rhat_250/350/500	–	Convergence Statistic (ideally < 1.2)
n_eff_250/350/500	–	Number of effective samples (ideally > 40)
Post_250/350/500	mJy	3000 samplings from the posterior PDF of the 250/350/500 $\mu\text{m}$ flux density
tile_SPIRE	–	tile number
F_850	mJy	850 $\mu\text{m}$ flux density (median)
FErr_850_u	mJy	850 $\mu\text{m}$ flux density (84th Percentile)
FErr_850_l	mJy	850 $\mu\text{m}$ flux density (16th Percentile)
FErr_850_1 $\sigma$	mJy	maximum of $\sigma^+$ and $\sigma^-$
Bkg_850	mJy/Beam	Fitted Background of 850 $\mu\text{m}$ map (median)
Sig_conf_850	mJy/Beam	Fitted residual noise component due to confusion (median)
Sig_tot_850	mJy/Beam	total error = $\sqrt{(\text{Sig\_conf\_850})^2 + (\text{FErr\_850\_1}\sigma)^2}$
Rhat_850	–	Convergence Statistic (ideally < 1.2)
n_eff_850	–	Number of effective samples (ideally > 40)
Post_850	mJy	3000 samplings from the posterior PDF of the 850 $\mu\text{m}$ flux density
tile_SCUBA	–	tile number

**Notes.**  $\sigma^+$  can be calculated from the difference of the 84th percentile and the median.  $\sigma^-$  can be calculated from the difference between the median and the 16th percentile. The  $1\sigma$  uncertainty of the source flux density can be derived from the maximum of  $\sigma^+$  and  $\sigma^-$ . For a final estimate of the flux uncertainty, one can use the total error derived from combining the  $1\sigma$  uncertainty and the residual confusion noise in quadrature. For the MIPS and SPIRE bands, this is equivalent to scaling the  $1\sigma$  uncertainty by a factor of two. For the PACS and SCUBA-2 bands, the residual confusion noise is much smaller than the instrument noise level. Consequently, the total error is very close to the  $1\sigma$  uncertainty.



**Fig. 28.** Distribution of the monochromatic  $q_{250}$  value of the MIGHTEE sources that do not have a match in COSMOS2020.  $k$ -correction is not applied due to the lack of photo- $z$  estimates. The various vertical lines and the shaded regions correspond to similar measurements from previous studies.

We use the COSMOS field as the first application in this work. In the future, we plan to apply our deblending methodology to other premier extragalactic survey fields. On the observational front, a lot of progress has been made over the past few years and exciting new data are expected to arrive soon which can be used to further improve the prior information used to deblend the long-wavelength data. For example, the COSMOS-Web (Casey et al. 2023; PIs: Kartaltepe & Casey, ID=1727) treasury program using JWST will cover  $0.54 \text{ deg}^2$  the COSMOS field with NIRCam imaging and  $0.19 \text{ deg}^2$  with MIRI imaging. The recently launched *Euclid* satellite will also observe the COSMOS field as one of its calibration fields. The deep optical and near-IR *Euclid* data, together with the rich multi-wavelength data such as deep IRAC mid-IR observations will be crucial for more accurate photometric redshift and stellar mass estimates, thereby providing better prior catalogue for future deblending work. On the other hand, new deep sub-mm observations from the ground can be used to directly validate any deblending approach. For example, the upcoming A-MKID sub-mm camera on the APEX telescope based on the latest microwave kinetic inductance detector (MKID) technologies will survey the sky at 350 and 850  $\mu\text{m}$  simultaneously. Thanks to its sensitivity, large format, and field of view ( $15 \times 15 \text{ arcmin}^2$ ), A-MKID will be an unprecedented sub-mm survey machine, capable of imaging large areas down to confusion limits in a reasonable time. Benefiting hugely from the 12-metre dish size, the resulting beam size at 350  $\mu\text{m}$  will be around  $7''$  (a factor of 3.4 smaller than *Herschel*).

**Acknowledgements.** We thank the anonymous referee for the constructive comments and suggestions. We thank Ian Smail and Emanuele Daddi for providing many very helpful comments and suggestions which significantly improved the quality of the paper. This publication is part of the project ‘Clash of the titans: deciphering the enigmatic role of cosmic collisions’ (with project number VI.Vidi.193.113 of the research programme Vidi which is (partly) financed by the Dutch Research Council (NWO). Using data from the public data release of the PACS Evolutionary Probe PEP (Lutz et al. 2011). W.J.P. has been supported by the Polish National Science Center project UMO-2020/37/B/ST9/00466. IHW, MJJ and CLH acknowledge generous support from the Hintze Family Charitable Foundation through the Oxford Hintze Centre for Astrophysical Surveys. CLH acknowledges support from the Leverhulme Trust through an Early Career Research Fellowship. MV 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 and the University of the Western Cape, and from the South African Department of Science and Innovation’s National Research Foundation under the ISARP RADIOMAP+ Joint Research Scheme (DSI-NRF Grant Number 150551) and the CPRR HIPPO Project (DSI-NRF Grant Number SRUG22031677). LM acknowledges support from the Italian Ministry of Foreign Affairs and International Cooperation (MAECI Grant Number ZA18GR02) and the South African Department of Science and Innovation’s National Research Foundation (DSI-NRF Grant Number 113121) under the ISARP RADIOSKY2020 Joint Research Scheme. Based on observations collected at the European Southern Observatory under ESO programme ID 179.A-2005 and on data products produced by CALET and the Cambridge Astronomy Survey Unit on behalf of the UltraVISTA consortium. The MeerKAT telescope is operated by the South African Radio Astronomy Observatory, which is a facility of the National Research Foundation, an agency of the Department of Science and Innovation. We acknowledge the use of the ilifu cloud computing facility – [www.ilifu.ac.za](http://www.ilifu.ac.za), a partnership between the University of Cape Town, the University of the Western Cape, Stellenbosch University, Sol Plaatje University and the Cape Peninsula University of Technology. The Ilifu facility is supported by contributions from the Inter-University Institute for Data Intensive Astronomy (IDIA – a partnership between the University of Cape Town, the University of Pretoria and the University of the Western Cape, the Computational Biology division at UCT and the Data Intensive Research Initiative of South Africa (DIRISA). The authors acknowledge the Centre for High Performance Computing (CHPC), South Africa, for providing computational resources to this research project. The James Clerk Maxwell Telescope is operated by the East Asian Observatory on behalf of The National Astronomical Observatory of Japan; Academia Sinica Institute of Astronomy and Astrophysics; the Korea Astronomy and Space Science Institute; the Operation, Maintenance and Upgrading Fund for Astronomical Telescopes and Facility Instruments, budgeted from the Ministry of Finance (MOF) of China and administrated by the Chinese Academy of Sciences (CAS), as well as the National Key R&D Program of China (No. 2017YFA0402700). Additional funding support is provided by the Science and Technology Facilities Council of the United Kingdom and participating universities in the United Kingdom and Canada (ST/M007634/1, ST/M003019/1, ST/N005856/1). The James Clerk Maxwell Telescope has historically been operated by the Joint Astronomy Centre on behalf of the Science and Technology Facilities Council of the United Kingdom, the National Research Council of Canada and the Netherlands Organization for Scientific Research and data from observations undertaken during this period of operation is used in this manuscript. This research used the JCMT project IDs M16AL002, M17BL006 (S2COSMOS) and S2CLS (MJLCS01) and the facilities of the Canadian Astronomy Data Centre operated by the National Research Council of Canada with the support of the Canadian Space Agency.

## References

- Abadi, M., Barham, P., Chen, J., et al. 2016, arXiv e-prints [arXiv:[1605.08695](https://arxiv.org/abs/1605.08695)]
- Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, *ApJS*, **182**, 543
- Aihara, H., AlSayyad, Y., Ando, M., et al. 2019, *PASJ*, **71**, 114
- Algera, H. S. B., Inami, H., Oesch, P. A., et al. 2023, *MNRAS*, **518**, 6142
- An, F. X., Stach, S. M., Smail, I., et al. 2018, *ApJ*, **862**, 101
- An, F. X., Simpson, J. M., Smail, I., et al. 2019, *ApJ*, **886**, 48
- Arnouts, S., Moscardini, L., Vanzella, E., et al. 2002, *MNRAS*, **329**, 355
- Becker, R. H., White, R. L., & Helfand, D. J. 1995, *ApJ*, **450**, 559
- Berta, S., Magnelli, B., Nordon, R., et al. 2011, *A&A*, **532**, A49
- Béthermin, M., Dole, H., Cousin, M., & Boulanger, N. 2010, *A&A*, **516**, A43
- Béthermin, M., Daddi, E., Magdis, G., et al. 2015, *A&A*, **573**, A113
- Béthermin, M., Wu, H.-Y., Lagache, G., et al. 2017, *A&A*, **607**, A89
- Béthermin, M., Van Cuyck, M., Beelen, A., & Gkogkou, A. 2022, pySIDES: Simulated Infrared Dusty Extragalactic Sky in Python, Astrophysics Source Code Library, [[record ascl:2204.016](https://doi.org/10.5281/2204.016)]
- Bourne, N., Dunlop, J. S., Merlin, E., et al. 2017, *MNRAS*, **467**, 1360
- Braak, C. J. F., Boer, M. P., Totir, L. R., et al. 2010, *Genetics*, **185**, 1045
- Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2008, *ApJ*, **686**, 1503
- Brinchmann, J., Charlot, S., White, S. D. M., et al. 2004, *MNRAS*, **351**, 1151
- Bruzual, G., & Charlot, S. 2003, *MNRAS*, **344**, 1000
- Bussmann, R. S., Riechers, D., Fialkov, A., et al. 2015, *ApJ*, **812**, 43
- Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, *ApJ*, **533**, 682
- Casey, C. M. 2012, *MNRAS*, **425**, 3094
- Casey, C. M., Narayanan, D., & Cooray, A. 2014, *Phys. Rep.*, **541**, 45
- Casey, C. M., Kartaltepe, J. S., Drakos, N. E., et al. 2023, *ApJ*, **954**, 31
- Chabrier, G. 2003, *PASP*, **115**, 763
- Chapin, E. L., Chapman, S. C., Coppin, K. E., et al. 2011, *MNRAS*, **411**, 505
- Charlot, S., & Fall, S. M. 2000, *ApJ*, **539**, 718
- Condon, J. J., Cotton, W. D., Greisen, E. W., et al. 1998, *AJ*, **115**, 1693
- Dale, D. A., Helou, G., Magdis, G. E., et al. 2014, *ApJ*, **784**, 83

- Delhaize, J., Smolčić, V., Delvecchio, I., et al. 2017, [A&A](#), **602**, A4  
 Devlin, M. J., Ade, P. A. R., Aretxaga, I., et al. 2009, [Nature](#), **458**, 737  
 Dole, H., Rieke, G. H., Lagache, G., et al. 2004, [ApJS](#), **154**, 93  
 Draine, B. T., Aniano, G., Krause, O., et al. 2014, [ApJ](#), **780**, 172  
 Driver, S. P., Popescu, C. C., Tuffs, R. J., et al. 2008, [ApJ](#), **678**, L101  
 Dudzevičiūtė, U., Smail, I., Swinbank, A. M., et al. 2020, [MNRAS](#), **494**, 3828  
 Eales, S., Dunne, L., Clements, D., et al. 2010, [PASP](#), **122**, 499  
 Elbaz, D., Dickinson, M., Hwang, H. S., et al. 2011, [A&A](#), **533**, A119  
 Euclid Collaboration (Moneti, A., et al.) 2022, [A&A](#), **658**, A126  
 Euclid Collaboration (Bisigello, L., et al.) 2023, [MNRAS](#), **520**, 3529  
 Fixsen, D. J., Dwek, E., Mather, J. C., Bennett, C. L., & Shafer, R. A. 1998, [ApJ](#), **508**, 123  
 Fritz, J., Franceschini, A., & Hatziminaoglou, E. 2006, [MNRAS](#), **366**, 767  
 Gao, F., Wang, L., Efstathiou, A., et al. 2021, [A&A](#), **654**, A117  
 Gardner, J. P., Mather, J. C., Clampin, M., et al. 2006, [Space Sci. Rev.](#), **123**, 485  
 Geach, J. E., Dunlop, J. S., Halpern, M., et al. 2017, [MNRAS](#), **465**, 1789  
 Gkogkou, A., Béthermin, M., Lagache, G., et al. 2023, [A&A](#), **670**, A16  
 Gruppioni, C., Béthermin, M., Loiacono, F., et al. 2020, [A&A](#), **643**, A8  
 Gürkan, G., Hardcastle, M. J., Smith, D. J. B., et al. 2018, [MNRAS](#), **475**, 3010  
 Hatziminaoglou, E., Farrah, D., Humphreys, E., et al. 2018, [MNRAS](#), **480**, 4974  
 Hauser, M. G., & Dwek, E. 2001, [ARA&A](#), **39**, 249  
 Heywood, I., Jarvis, M. J., Hale, C. L., et al. 2022, [MNRAS](#), **509**, 2150  
 Hodge, J. A., Karim, A., Smail, I., et al. 2013, [ApJ](#), **768**, 91  
 Holland, W. S., Bintley, D., Chapin, E. L., et al. 2013, [MNRAS](#), **430**, 2513  
 Hurley, P. D., Oliver, S., Betancourt, M., et al. 2017, [MNRAS](#), **464**, 885  
 Ilbert, O., Arnouts, S., McCracken, H. J., et al. 2006, [A&A](#), **457**, 841  
 Ivison, R. J., Alexander, D. M., Biggs, A. D., et al. 2010, [MNRAS](#), **402**, 245  
 Jarvis, M. J., Smith, D. J. B., Bonfield, D. G., et al. 2010, [MNRAS](#), **409**, 92  
 Jarvis, M., Taylor, R., Agudo, I., et al. 2016, in [MeerKAT Science: On the Pathway to the SKA](#), 6  
 Jin, S., Daddi, E., Liu, D., et al. 2018, [ApJ](#), **864**, 56  
 Jonas, J., & MeerKAT Team 2016, in [MeerKAT Science: On the Pathway to the SKA](#), 1  
 Kingma, D. P., & Ba, J. 2014, arXiv e-prints [[arXiv:1412.6980](#)]  
 Komatsu, E., Smith, K. M., Dunkley, J., et al. 2011, [ApJS](#), **192**, 18  
 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](#), [[record ascl:1604.008](#)]  
 Larson, D., Dunkley, J., Hinshaw, G., et al. 2011, [ApJS](#), **192**, 16  
 Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, arXiv e-prints [[arXiv:1110.3193](#)]  
 Le Floc'h, E., Aussel, H., Ilbert, O., et al. 2009, [ApJ](#), **703**, 222  
 Lee, N., Sanders, D. B., Casey, C. M., et al. 2013, [ApJ](#), **778**, 131  
 Liu, D., Daddi, E., Dickinson, M., et al. 2018, [ApJ](#), **853**, 172  
 Liu, D., Lang, P., Magnelli, B., et al. 2019, [ApJS](#), **244**, 40  
 Long, A. S., Casey, C. M., del P. Lagos, C., et al. 2023, [ApJ](#), **953**, 11  
 Lutz, D., Poglitsch, A., Altieri, B., et al. 2011, [A&A](#), **532**, A90  
 Madau, P., & Dickinson, M. 2014, [ARA&A](#), **52**, 415  
 Magnelli, B., Lutz, D., Berta, S., et al. 2010, [A&A](#), **518**, A28  
 Magnelli, B., Lutz, D., Santini, P., et al. 2012, [A&A](#), **539**, A155  
 Magnelli, B., Popesso, P., Berta, S., et al. 2013, [A&A](#), **553**, A132  
 Martis, N. S., Marchesini, D., Brammer, G. B., et al. 2016, [ApJ](#), **827**, L25  
 McCracken, H. J., Milvang-Jensen, B., Dunlop, J., et al. 2012, [A&A](#), **544**, A156  
 Miller, N. A., Fomalont, E. B., Kellermann, K. I., et al. 2008, [ApJS](#), **179**, 114  
 Moneti, A., McCracken, H. J., Hudelot, W., et al. 2023, [VizieR Online Data Catalog](#), **II/373**  
 Nair, V., & Hinton, G. E. 2010, in [Proceedings of the 27th International Conference on Machine Learning \(ICML-10\)](#), 807  
 Nguyen, H. T., Schulz, B., Levenson, L., et al. 2010, [A&A](#), **518**, L5  
 Noeske, K. G., Weiner, B. J., Faber, S. M., et al. 2007, [ApJ](#), **660**, L43  
 Noll, S., Burgarella, D., Giovannoli, E., et al. 2009, [A&A](#), **507**, 1793  
 Oliver, S. J., Wang, L., Smith, A. J., et al. 2010, [A&A](#), **518**, A21  
 Oliver, S. J., Bock, J., Altieri, B., et al. 2012, [MNRAS](#), **424**, 1614  
 Pannella, M., Carilli, C. L., Daddi, E., et al. 2009, [ApJ](#), **698**, L116  
 Pearson, W. J., Wang, L., van der Tak, F. F. S., et al. 2017, [A&A](#), **603**, A102  
 Pearson, W. J., Wang, L., Hurley, P. D., et al. 2018, [A&A](#), **615**, A146  
 Popesso, P., Morselli, L., Concas, A., et al. 2019, [MNRAS](#), **490**, 5285  
 Popesso, P., Concas, A., Cresci, G., et al. 2023, [MNRAS](#), **519**, 1526  
 Puget, J. L., Abergel, A., Bernard, J. P., et al. 1996, [A&A](#), **308**, L5  
 Rieke, G. H., Alonso-Herrero, A., Weiner, B. J., et al. 2009, [ApJ](#), **692**, 556  
 Rodighiero, G., Lari, C., Pozzi, F., et al. 2006, [MNRAS](#), **371**, 1891  
 Roseboom, I. G., Oliver, S. J., Kunz, M., et al. 2010, [MNRAS](#), **409**, 48  
 Roseboom, I. G., Ivison, R. J., Greve, T. R., et al. 2012, [MNRAS](#), **419**, 2758  
 Sanders, D. B., Salvato, M., Aussel, H., et al. 2007, [ApJS](#), **172**, 86  
 Sawicki, M., Arnouts, S., Huang, J., et al. 2019, [MNRAS](#), **489**, 5202  
 Schinnerer, E., Sargent, M. T., Bondi, M., et al. 2010, [ApJS](#), **188**, 384  
 Scoville, N., Aussel, H., Brusa, M., et al. 2007, [ApJS](#), **172**, 1  
 Scudder, J. M., Oliver, S., Hurley, P. D., et al. 2016, [MNRAS](#), **460**, 1119  
 Serra, P., Amblard, A., Temi, P., et al. 2011, [ApJ](#), **740**, 22  
 Shirley, R., Duncan, K., Campos Varillas, M. C., et al. 2021, [MNRAS](#), **507**, 129  
 Simpson, J. M., Smail, I., Swinbank, A. M., et al. 2019, [ApJ](#), **880**, 43  
 Simpson, J. M., Smail, I., Dudzevičiūtė, U., et al. 2020, [MNRAS](#), **495**, 3409  
 Smith, A. J., Wang, L., Oliver, S. J., et al. 2012, [MNRAS](#), **419**, 377  
 Smith, D. J. B., Jarvis, M. J., Hardcastle, M. J., et al. 2014, [MNRAS](#), **445**, 2232  
 Smolčić, V., Novak, M., Bondi, M., et al. 2017, [A&A](#), **602**, A1  
 Speagle, J. S., Steinhardt, C. L., Capak, P. L., & Silverman, J. D. 2014, [ApJS](#), **214**, 15  
 Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. 2014, [J. Mach. Learn. Res.](#), **15**, 1929  
 Stalevski, M., Fritz, J., Baes, M., Nakos, T., & Popović, L. Č. 2012, [MNRAS](#), **420**, 2756  
 Stan Development Team 2015, [Stan: A C++ library for probability and sampling](#), version 2.8. 0  
 Stan Development Team 2018, [PyStan: the Python interface to Stan](#), version 2.19, 1.1  
 Tibshirani, R. 1996, [J. Roy. Statist. Soc. Ser. B: Statist. Methodol.](#), **58**, 267  
 Wang, L., Viero, M., Clarke, C., et al. 2014, [MNRAS](#), **444**, 2870  
 Wang, L., Norberg, P., Gunawardhana, M. L. P., et al. 2016, [MNRAS](#), **461**, 1898  
 Wang, L., Pearson, W. J., Cowley, W., et al. 2019, [A&A](#), **624**, A98  
 Wang, L., Gao, F., Best, P. N., et al. 2021, [A&A](#), **648**, A8  
 Weaver, J. R., Kauffmann, O. B., Ilbert, O., et al. 2022, [ApJS](#), **258**, 11  
 Weaver, J. R., Davidzon, I., Toft, S., et al. 2023, [A&A](#), **677**, A184  
 Whitaker, K. E., Labb  , I., van Dokkum, P. G., et al. 2011, [ApJ](#), **735**, 86  
 Whitaker, K. E., van Dokkum, P. G., Brammer, G., & Franx, M. 2012, [ApJ](#), **754**, L29  
 Whitaker, K. E., Pope, A., Cybulski, R., et al. 2017, [ApJ](#), **850**, 208  
 Wuyts, S., F  rster Schreiber, N. M., van der Wel, A., et al. 2011, [ApJ](#), **742**, 96  
 Zavala, J. A., Casey, C. M., Manning, S. M., et al. 2021, [ApJ](#), **909**, 165  
 Zou, H. 2006, [J. Am. Statist. Assoc.](#), **101**, 1418

## Appendix A: CIGALE parameters

Table A.1 shows the choices of the SED model components and parameters in the 16 CIGALE runs, with all possible combinations of two star-formation history models, two dust attenuation models, two dust emission models, and two AGN models.

## Appendix B: Scaling the flux uncertainty

To make the computational time manageable, XID+ ignores the fact that residual confusion noise is correlated between pixels. Instead, it is assumed to be constant in each tile and is modelled as Gaussian fluctuations around a global background,  $N(B, \Sigma_{\text{confusion}}^{\text{residual}})$ , as described in Eq. (5). If we add this residual confusion noise  $\Sigma_{\text{confusion}}^{\text{residual}}$  in quadrature to the  $1\sigma$  flux uncertainty estimate to form a "total" noise, then the "total" flux uncertainties behave as expected in the sense that the distribution of the flux differences normalised by the "total" flux uncertainties follows the standard Gaussian.

In the MIPS and SPIRE bands, scaling the  $1\sigma$  flux uncertainty by a factor of two also makes the distribution of the normalised flux difference to follow the standard Gaussian. In other words, the effect of scaling by a factor of two is roughly the same as combining the  $1\sigma$  flux uncertainty with the residual confusion noise in quadrature, as shown in Fig. B.1. To understand why they coincide, consider using XID+ to fit a map with  $n_1 \times n_2 = M$  pixels. XID+ assumes that the map  $\mathbf{d}$  (a flattened vector of  $M$  pixels) is formed from the contribution of the known sources and two independent noise terms (instrument noise and residual confusion noise). The contribution of the known sources  $\sum_{i=1}^S \mathbf{P} f_i$ , where  $\mathbf{P}$  is the point response function (PRF) and  $f_i$  is the flux of a given source, can be written in a linear form,

$$\mathbf{d} = \mathbf{Af}. \quad (\text{B.1})$$

Here  $\mathbf{f}$  is a vector of  $S$  elements (corresponding to the number of known sources) and  $\mathbf{A}$  is a  $M \times S$  pointing matrix. Therefore, the maximum-likelihood solution can be found as

$$\mathbf{f} = (\mathbf{A}^T \mathbf{N}_d^{-1} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{N}_d^{-1} \mathbf{d}, \quad (\text{B.2})$$

where  $\mathbf{N}_d$  is a diagonal matrix with  $\mathbf{N}_{d,ii} = \sigma_{\text{inst},ii}^2 + \sigma_{\text{conf}}^2$ . If we make an assumption that the sources do not overlap with each other, then the Fisher information matrix can be simplified to a  $S \times S$  diagonal matrix with each entry equal to

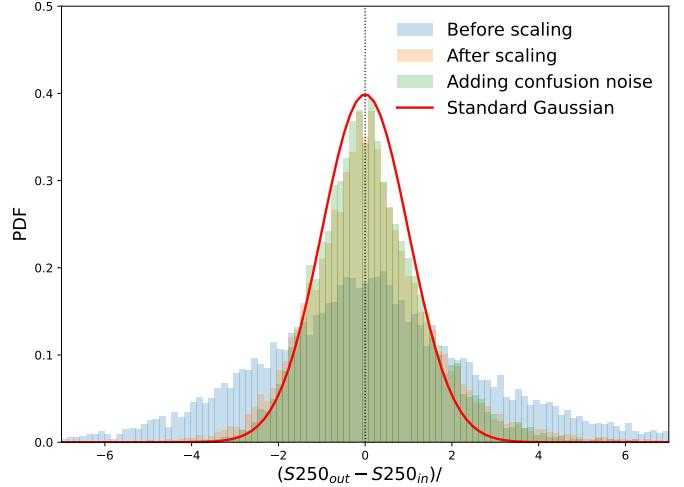
$$\Sigma \text{PRF}_i^2 / < \sigma_{\text{inst},i}^2 + \sigma_{\text{conf}}^2 >. \quad (\text{B.3})$$

By the Cramér-Rao lower bound, the inverse of Eq. (B.3) represents the lower limit of the variance of the estimated flux density. If we make a further simplifying assumption that the instrument noise does not vary much across the pixels, then the minimum of the variance would be

$$\text{var}_{\min} = < \sigma_{\text{inst}}^2 + \sigma_{\text{conf}}^2 > / \Sigma \text{PRF}_i^2. \quad (\text{B.4})$$

Take the deblended 250  $\mu\text{m}$  fluxes as an example. From Fig. B.2, the residual confusion noise is  $\sim 3$  mJy, which is a factor of 1.8 higher than the instrument noise (see Table 3). Given  $\Sigma \text{PRF}_i^2 = 5.2$  for the 250  $\mu\text{m}$  beam map, we can derive the minimum  $1\sigma$  flux uncertainty is  $\sim 2$  times smaller than the residual confusion noise, as

$$\text{var}_{\min}^{250} = < (\sigma_{\text{conf}}/1.8)^2 + \sigma_{\text{conf}}^2 > / 5.2 = 0.25 \sigma_{\text{conf}}^2. \quad (\text{B.5})$$

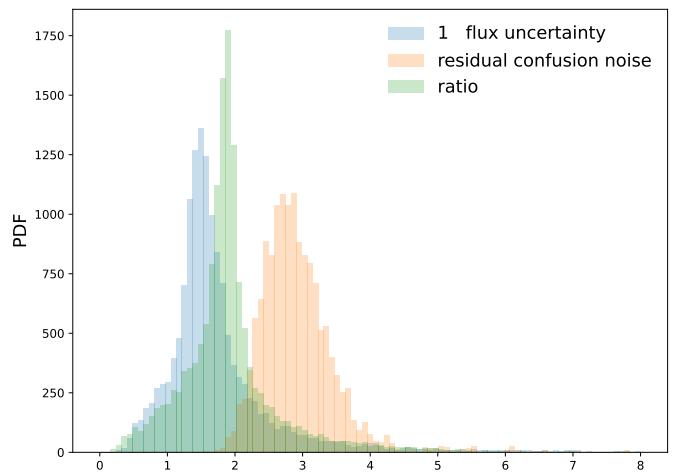


**Fig. B.1.** Distribution of flux difference normalised by the  $1\sigma$  flux uncertainty estimate, before and after scaling  $1\sigma$  by a factor of two. We also show the distribution after adding in the residual confusion noise in quadrature to the  $1\sigma$  flux uncertainty.

Therefore combining the  $1\sigma$  uncertainty with the residual noise in quadrature has roughly the same effect as scaling the  $1\sigma$  uncertainty by a factor of two,

$$\sigma_{\text{total}} = \sqrt{(1\sigma)^2 + \sigma_{\text{conf}}^2} = \sqrt{5}\sigma. \quad (\text{B.6})$$

Performing these calculations for the MIPS band shows the same conclusion. For the PACS and SCUBA-2 bands, because the instrument noise is much greater than the confusion noise, the  $1\sigma$  flux uncertainty is dominated by instrument noise and therefore does not need to be scaled up as for the other bands. In other words, the total error after including the residual confusion noise is almost the same as the  $1\sigma$  flux uncertainty.



**Fig. B.2.** Distribution of the  $1\sigma$  flux uncertainty at 250  $\mu\text{m}$  (derived from the maximum of  $\sigma^+$  and  $\sigma^-$ ) is shown by the blue histogram. The brown histogram shows the distribution of the residual confusion noise. The green histogram shows the ratio of the two.

**Table A.1.** Choices of the SED model components and parameters in the 16 CIGALE runs.

<b>star-formation history</b>		
delayed $\tau$ +starburst	e-folding time of the main population	500, 1000, 3000, 5000, 8000 Myrs
	age the main population	500, 1000-13 000 (step 1000) Myrs
	e-folding time of the late starburst population	9000, 13 000 Myrs
	age of the late starburst population	1, 30, 100, 500 Myrs
	mass fraction of the late starburst population	0.01, 0.05, 0.1, 0.2
double exponential	e-folding time of the main population	1000-9000 (step 2000) Myrs
	age the main population	500, 1000-13 000 (step 1000) Myrs
	e-folding time of the late starburst population	10000 Myrs
	age of the late starburst population	1, 30, 100, 500 Myrs
	mass fraction of the late starburst population	0.01, 0.05, 0.1, 0.2
<b>dust attenuation</b>		
modified Charlot and Fall 2000	V-band attenuation in the interstellar medium	0.1, 0.5-3.5 (step 0.5)
	$\text{Av}_{ISM}/\text{Av}_{BC} + \text{Av}_{ISM}$	0.25, 0.5, 0.75
	power-law slope of the attenuation in the birth clouds	-0.7
	power-law slope of the attenuation in the ISM	-0.7
modified Calzetti 2000	color excess	0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5
	reduction factor	0.25, 0.5, 0.75
	Slope of the power law modifying the attenuation curve	-0.5, -0.4, -0.3, -0.2, -0.1, 0
	Extinction law	Milky Way
<b>dust emission</b>		
Draine et al. 2014	Mass fraction of PAH	0.47
	minimum radiation field	5, 15, 25
	power-law slope $\alpha$ in $dM/dU \propto U^\alpha$	2
	Fraction illuminated from min to max radiation field	0.02
Dale et al. 2014	alpha slope	1.5-4 (step 0.5)
<b>AGN templates</b>		
Fritz et al. 2006	Full opening angle of the dust torus (degree)	100
	viewing angle (degree)	0.001, 50.1, 89.990
	AGN fraction	0, 0.1, 0.3, 0.5, 0.7
	optical depth at 9.7 $\mu\text{m}$	0.3, 1, 6
	linked to the radial dust distribution in the torus	-0.5
	linked to the angular dust distribution in the torus	0
SKIRTOR	optical depth at 9.7 $\mu\text{m}$	7
	torus density radial parameter $p$	1
	torus density angular parameter $q$	1
	half opening angle (degree)	40
	viewing angle (degree)	30, 70
	AGN fraction	0, 0.1, 0.3, 0.5, 0.7
	extinction law of polar dust	SMC
	E(B-V) of polar dust	0, 0.1, 0.2, 0.3, 0.4
	temperature of polar dust	100
	emissivity of polar dust	1.6

**Notes.** The 16 CIGALE runs involve all possible combinations of two star-formation history models (delayed  $\tau$  + starburst or double exponential), two dust attenuation models (modified [Charlot & Fall \(2000\)](#) or modified [Calzetti et al. \(2000\)](#) attenuation law), two dust emission models (IR SED templates from the [Draine et al. \(2014\)](#) model or the [Dale et al. \(2014\)](#) model), and two AGN models (smooth AGN templates from [Fritz et al. \(2006\)](#) or clumpy AGN templates from the SKIRTOR model in [Stalevski et al. \(2012\)](#)).