



Universiteit  
Leiden  
The Netherlands

## **The ALPINE-ALMA [C II] survey: luminosity function of serendipitous [C II] line emitters at $z \sim 5$**

Loiacono, F.; Decarli, R.; Gruppioni, C.; Talia, M.; Cimatti, A.; Zamorani, G.; ... ; Zucca, E.

### **Citation**

Loiacono, F., Decarli, R., Gruppioni, C., Talia, M., Cimatti, A., Zamorani, G., ... Zucca, E. (2021). The ALPINE-ALMA [C II] survey: luminosity function of serendipitous [C II] line emitters at  $z \sim 5$ . *Astronomy & Astrophysics*, 646, 1-18. doi:10.1051/0004-6361/202038607

Version: Submitted Manuscript (under Review)

License: [Leiden University Non-exclusive license](#)

Downloaded from: <https://hdl.handle.net/1887/3273964>

**Note:** To cite this publication please use the final published version (if applicable).

# The ALPINE–ALMA [C II] survey: the luminosity function of serendipitous [C II] line emitters at $z \sim 5$

Federica Loiacono<sup>1,2,10,11,\*</sup>, Roberto Decarli<sup>2</sup>, Carlotta Gruppioni<sup>2</sup>, Margherita Talia<sup>1,2</sup>, Andrea Cimatti<sup>2,3</sup>, Gianni Zamorani<sup>2</sup>, Francesca Pozzi<sup>1</sup>, Lin Yan<sup>4</sup>, Brian C. Lemaux<sup>20</sup>, Dominik A. Riechers<sup>5,6</sup>, Olivier Le Fèvre<sup>19</sup>, Matthieu Béthermin<sup>19</sup>, Peter Capak<sup>7</sup>, Paolo Cassata<sup>14,21</sup>, Andreas Faisst<sup>7</sup>, Daniel Schaerer<sup>15</sup>, John D. Silverman<sup>22,23</sup>, Sandro Bardelli<sup>2</sup>, Médéric Boquien<sup>18</sup>, Sandra Burkutean<sup>17</sup>, Miroslava Dessauges-Zavadsky<sup>15</sup>, Yoshinobu Fudamoto<sup>15</sup>, Seiji Fujimoto<sup>8,9</sup>, Michele Ginolfi<sup>15</sup>, Nimish P. Hathi<sup>12</sup>, Gareth C. Jones<sup>10,11</sup>, Yana Khusanova<sup>19</sup>, Anton M. Koekemoer<sup>12</sup>, Guilaine Lagache<sup>19</sup>, Marcella Massardi<sup>17</sup>, Pascal Oesch<sup>15</sup>, Michael Romano<sup>14,21</sup>, Livia Vallini<sup>13</sup>, Daniela Vergani<sup>2</sup> and Elena Zucca<sup>2</sup>

(Affiliations can be found after the references)

submitted to A & A

## ABSTRACT

We present the first [C II] 158  $\mu\text{m}$  luminosity function (LF) at  $z \sim 5$  from a sample of serendipitous lines detected in the ALMA Large Program to INvestigate [C II] at Early times (ALPINE). A search performed over the 118 ALPINE pointings revealed several serendipitous lines. Based on their fidelity, we selected 14 lines for the final catalog. According to the redshift of their counterparts, we identified eight out of 14 detections as [C II] lines at  $z \sim 5$ , and two as CO transitions at lower redshifts. The remaining four lines have an elusive identification in the available catalogs and we considered them as [C II] candidates. We used the eight confirmed [C II] and the four [C II] candidates to build one of the first [C II] LFs at  $z \sim 5$ . We found that 11 out of these 12 sources have a redshift very similar to that of the ALPINE target in the same pointing, suggesting the presence of overdensities around the targets. Therefore, we split the sample in two (a “clustered” and “field” sub-sample) according to their redshift separation and built two separate LFs. Our estimates suggest that there could be an evolution of the [C II] LF between  $z \sim 5$  and  $z \sim 0$ . By converting the [C II] luminosity to star formation rate we evaluated the cosmic star formation rate density (SFRD) at  $z \sim 5$ . The clustered sample results in a SFRD  $\sim 10$  times higher than previous measurements from UV-selected galaxies. On the other hand, from the field sample (likely representing the average galaxy population) we derived a SFRD  $\sim 1.6$  higher compared to current estimates from UV surveys but compatible within the errors. Because of the large uncertainties, observations of larger samples are necessary to better constrain the SFRD at  $z \sim 5$ . This study represents one of the first efforts aimed at characterizing the demography of [C II] emitters at  $z \sim 5$  using a mm-selection of galaxies.

**Key words.** galaxies: evolution – galaxies: high-redshift – galaxies: ISM – submillimeter: galaxies – galaxies: luminosity function, mass function

## 1. Introduction

Our quest on the early phases of galaxy evolution cannot prescind from the study of cold gas. High-redshift galaxies are indeed more gas-rich than present day objects with gas fractions up to unity, as witnessed by large observing campaigns (e.g. Tacconi et al. 2018). The rate at which the Universe form stars varies significantly across cosmic time (Madau & Dickinson 2014); however, the drivers of this trend are still poorly known. Up to  $z \sim 3$ , we have a robust understanding of the star formation history, thanks to more than twenty years of multi-wavelength investigations (e.g., Takeuchi et al. 2003; Schiminovich et al. 2005; Cucciati et al. 2012; Gruppioni et al. 2013; Magnelli et al. 2013; Bouwens et al. 2015; Finkelstein 2016; Oesch et al. 2018; Bowler et al. 2020). Nevertheless at  $z > 3$  our constraints are almost exclusively based on observations sampling the rest-frame ultraviolet (UV) emission, which is very sensitive to dust reddening. Studies at longer wavelengths (e.g., Karim et al. 2013; Rowan-Robinson et al. 2016; Novak et al. 2017; Maniyar et al. 2018) hint to the presence of a population of gas- and dust-rich galaxies that may be missed by the UV selection. However, the demography of such dusty galaxies, and therefore their role in shaping the cosmic star formation rate density at  $z > 3$ , is still

very uncertain.

Over the last few years we have been witnessing a true revolution, with the Atacama Large Millimeter/submillimeter Array (ALMA) opening a window on the high- $z$  obscured Universe. Thanks to its unprecedented sensitivity, ALMA allows us to detect for the first time the dust continuum and the bright infrared (IR) lines in normal galaxies at  $z > 3$  and constrain the cosmic star formation history (Bouwens et al. 2016; Scoville et al. 2017; Liu et al. 2019). In particular, the [C II] 158  $\mu\text{m}$  line can easily be detected because it is one of brightest galaxy lines in the IR, radiating up to a hundredth of the entire far-infrared luminosity of a galaxy (Díaz-Santos et al. 2013), and it is conveniently redshifted into atmospheric relatively transparent windows. This line is mainly excited by collision with neutral hydrogen atoms in the so-called photo-dissociation regions (PDRs; Hollenbach & Tielens 1999) and in the neutral diffuse gas (Wolfire et al. 2003). Nevertheless, it can also trace diffuse ionized gas where it is excited by collisions with free electrons (e.g. Cormier et al. 2012). Thanks to its brightness, [C II] is a powerful tool to derive accurate redshifts of distant galaxies (e.g. Walter et al. 2012; Riechers et al. 2013; Capak et al. 2015). Spatially resolved observations of this line can be used to characterize the kinematics of the cold interstellar medium (ISM; Smit et al. 2018; Kohandel et al. 2019). Besides, when other lines are also available, flux

\* e-mail: federica.loiacono2@unibo.it

ratios can be used to study the physical properties of the ISM in terms of gas density, strength of the radiation field and excitation source (e.g., Pavese et al. 2016, 2018; Novak et al. 2019). Finally, [C II] has also been found to be a star-formation rate (SFR) indicator at low and possibly at high- $z$  by observations (De Looze et al. 2014; Magdis et al. 2014; Carniani et al. 2018; Matthee et al. 2019; Schaerer et al. 2020) and models predictions (Vallini et al. 2015; Lagache et al. 2018).

The main limitation of mm interferometers such as ALMA is their relatively small field of view, which makes surveys of blank fields expensive in terms of telescope time. Most of the studies at high- $z$  have therefore focused on the exploration of properties of “targeted” galaxies that were pre-selected based on their stellar mass, SFR and/or IR luminosity (e.g. Daddi et al. (2015); Tacconi et al. (2018)). These kind of studies have been instrumental to shape our understanding of the connection between the inner gas reservoirs and the build-up of galaxies. However, the pre-selection may introduce biases associated to our prior knowledge of the emitting systems. On the other hand, “blind” surveys, as well as serendipitous discoveries in observations targeting other sources, aid to circumvent selection biases, thus enabling a proper census of the cold gas properties in a volume-limited region of the universe (Decarli et al. 2016; Riechers et al. 2019). In particular, blind selections of lines in the mm-domain are sensitive to heavily obscured galaxies that can be missed in the UV surveys. Properly accounting for these objects is crucial when estimating global quantities such as the cosmic SFRD and building luminosity functions.

Recently, the ALMA Large Program to INvestigate [C II] at Early times (ALPINE) has been completed (Le Fèvre et al. 2019; Bethermin et al. 2020; Faisst et al. 2019). This project aims at studying the [C II] emission in 118 spectroscopically confirmed and UV-selected star-forming galaxies at  $4 < z < 6$ . A search for spectral lines in the 118 ALPINE pointings unveiled a wealth of unexpected lines, i.e. serendipitous discoveries in a wide redshift range. Most of the lines are due to [C II] emission. We use these lines to build the [C II] luminosity function at  $z \sim 5$ . This is the first [C II] LF based on galaxies purely selected for their [C II] emission. On the other hand, the companion paper of Yan et al. (2020) presents the [C II] LF from the UV-selected central targets. Despite being well constrained at  $z \sim 0$  from statistical samples, at high-redshift the number density of [C II] emitters represents an uncharted territory. A knowledge of their LF is crucial to constrain the semi-analytical models and cosmological zoom-in simulations (e.g. Pallottini et al. 2019). Furthermore, it is also pivotal for quantifying the SFRD at high-redshift with an unbiased tracer, not affected by obscuration.

The paper is organized as follows. In Sect. 2 we briefly describe the ALPINE data and the ancillary photometry. In Sect. 3 we present the search for the serendipitous lines and the fidelity and completeness assessment. Sect. 4 is devoted to the identification of the lines. In Sect. 5 we show the [C II] luminosity function and compare it with other observational studies and models predictions. Sect. 6 deals with the cosmic star formation rate density. We finally summarize the main results in Sect. 7.

We adopt a  $\Lambda$ CDM cosmology using  $\Omega_\Lambda = 0.7$ ,  $\Omega_M = 0.3$  and  $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$ . We assumed a Chabrier (2003) initial mass function (IMF).

## 2. ALPINE in a nutshell

In this section we briefly describe the ALPINE project and the ALMA and ancillary data used in this work. Rather than being used to study the main UV-selected targets, the ALPINE dat-

acubes were employed to look for serendipitous sources, as will be exhaustively described in Sect. 3. We call “serendipitous” every line that is detected at a distance larger than  $1''$  from the targeted UV-galaxies (see also Bethermin et al. 2020).

### 2.1. Data description

The primary goal of ALPINE is to study the [C II] emission in a statistical sample of galaxies (Le Fèvre et al. 2019). The targets are 118 UV-selected star-forming galaxies, placed on the SFR- $M_\star$  “main sequence” (e.g. Rodighiero et al. 2011; Speagle et al. 2014; Faisst et al. 2019). Their redshifts are robustly constrained by UV-optical spectroscopy. The galaxies are located in well-studied sky regions, i.e. the Cosmic Evolution Survey field (COSMOS; Scoville et al. 2007) and the Extended Chandra Deep Field-South (ECDFS; Giavalisco et al. 2004; Cardamone et al. 2010). For 75 out of 118 galaxies (64 % of the sample) the [C II] emission was successfully detected while only 23 sources show significant continuum emission (20 % of the sample). For a comprehensive description of the targets catalogs see Bethermin et al. 2020.

The observations were carried out using ALMA band 7 during Cycles 5 and 6. Two frequency settings were adopted to observe two redshift windows at  $4.40 < z < 4.58$  and  $5.13 < z < 5.85$ . The achieved noise is, on average,  $0.14 \text{ Jy/beam km s}^{-1}$  over a line width of  $235 \text{ km s}^{-1}$  and  $39 \mu\text{Jy/beam}$  over the continuum. The data reduction and processing was handled with the software CASA (see Bethermin et al. 2020 for a full description). The visibilities were imaged using a natural weighting of the uv-plane, as the best compromise between spatial resolution and sensitivity. We used a pixel size of  $0.15''$  and an image size of  $256 \times 256$  pixels in order to properly sample the primary beam ( $\sim 21''$  at 300 GHz). The final 118 datacubes have a channel width varying from  $26 \text{ km s}^{-1}$  (highest frequency setting) to  $33 \text{ km s}^{-1}$  (lowest frequency setting). The average spatial resolution is  $0.85'' \times 1.13''$ . The total area covered by each pointing is  $0.41 \text{ arcmin}^2$ . However, in order to guarantee an adequate sensitivity, we limited the search of the serendipitous lines to a smaller area (see Sect. 3.2 for the details). We also excluded a circle of  $1''$  radius around the phase center to avoid the emission due to the central UV-targets. This entails a final effective sky area of  $27.42 \text{ arcmin}^2$  ( $0.23 \text{ arcmin}^2$  per pointing) where the serendipitous sources can be detected<sup>1</sup>.

### 2.2. Ancillary photometry

Since ALPINE observed extensively studied fields, all the sources located in the 118 pointings benefit from a wealth of multi-wavelength ancillary data (see Faisst et al. 2019 for a comprehensive description). The UV to near-infrared photometry is widely covered by the COSMOS15 and 3D-*HST* catalogs (Laigle et al. 2016; Brammer et al. 2012) and *HST* imaging (Koekemoer et al. 2007, 2011). These catalogs also contain estimates of the photometric redshifts, which were used to guide the line identification (see Sect. 4). In addition, for the two fields there are also *Spitzer*-IRAC images at 3.6, 4.5, 5.8 and  $8 \mu\text{m}$  (Capak et al. 2012; Ashby et al. 2013; Guo et al. 2013; Sanders et al. 2007; Laigle et al. 2016), MIPS (Dickinson et al. 2003; Le Floc’h et al. 2009) and *Herschel* data (Lutz et al. 2011; Elbaz

<sup>1</sup> We note that this survey area is higher than the area reported by Bethermin et al. 2020 since we included in our estimate a region where the primary beam attenuation reaches the 90% while in Bethermin et al. 2020 the 80% region has been considered.

et al. 2011). Also *Chandra* data are available for the sources located in the COSMOS field (Marchesi et al. 2016). Finally, at the longest wavelengths, deep JVLA observations at 3 GHz provide estimate of the radio continuum (Smolčić et al. 2017). The study of the spectral energy distribution (SEDs) of the serendipitous sources based on their photometric properties will be presented in another paper (Loiacono et al., in prep.).

### 3. Search for the serendipitous emission lines

#### 3.1. Code description

We performed the search for the serendipitous lines using *find-clumps* (see Decarli et al. 2016; Walter et al. 2016 for an exhaustive description), a code designed to look for sources without any prior knowledge of their frequency and spatial position, and which has been already exploited in the ASPECS survey (Walter et al. 2016; Decarli et al. 2019). In short, the algorithm performs a floating average of the channels over a range of kernels (number of channels) and searches for peaks exceeding a given signal-to-noise ratio (S/N). The latter is defined as the peak flux density as measured in the averaged map divided by the RMS computed within the entire map.

We executed the search on the 118 ALPINE datacubes adopting a S/N threshold of 3. For each pointing the search was repeated on datacubes of different channel width, from  $\sim 90 \text{ km s}^{-1}$  to  $550 \text{ km s}^{-1}$ , since these values are compatible with the typical widths of mm-lines at high- $z$  (Capak et al. 2015; Aravena et al. 2019). The probability of a detection is indeed maximized when the channel width is of the order of the full-width at half-maximum (FWHM) of the line, while it is lowered when the channel width is larger/narrower.

After the search, we removed the double detections from the output list, i.e. all the peaks at a distance lower than the beam size and in contiguous channels for each detection. We also repeated the search after subtracting the continuum for those lines for which also its emission was detected, in order to obtain S/N referring on the line emission only (for the continuum source detection method see Bethermin et al. 2020). Moreover, for those lines detected in datacubes with different channel widths, we considered the detection with the highest S/N as the final entry for our catalog. We obtained in this way the final list of the line candidates, where a mixture of real lines and spurious detections (i.e. noise peaks exceeding the S/N threshold) is expected.

#### 3.2. Fidelity

In order to disentangle the genuine lines from the noise peaks in the output list, we compared the number of the positive peaks detected in the datacubes (i.e. real lines and noise peaks) with the number of negative peaks above the threshold, as a function of the S/N. Unlike the positive ones, the negative peaks provide indeed the distribution of the pure noise of our data. This comparison provides the fidelity, i.e. the probability that one detection is a genuine line. Following the approach of Decarli et al. (2016) we defined the fidelity  $f$  as

$$f(S/N) = 1 - \frac{N_{\text{neg}}(S/N)}{N_{\text{pos}}(S/N)} \quad (1)$$

where  $N_{\text{neg}}$  and  $N_{\text{pos}}$  are the number of negative and positive peaks respectively. Defined in this way, the fidelity looks like a function of the S/N only of a detection. We note that in principle there are other factors that could influence it. For instance,

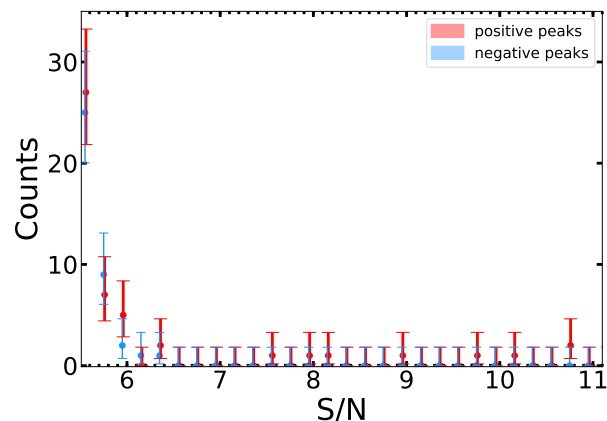


Fig. 1: Number of positive (red) and negative (i.e. noise; blue) peaks detected in the 118 ALPINE pointings as a function of the S/N. The errorbars are the Poissonian uncertainties. We can see that for  $S/N > 5.8$  the number of positive peaks becomes higher than the number of the negative ones as the number of genuine detections wrt spurious sources increases.

the fidelity could be also a function of the line width as, for two detections of equal S/N, a larger line has a higher fidelity than a narrower one (see González-López et al. 2019). Moreover, the fidelity can also depend on the line location in the field of view (FOV), since the sensitivity within the primary beam is not uniform. However, because of the low statistics of the positive/negative peaks above  $S/N = 5.6$  (below ten counts per bin even considering the 118 pointings; see Figure 1), it was not possible to split the peaks in sub-samples based on their distance from the pointing center and their width. This S/N range is indeed crucial to assess the fidelity, as the number of genuine detections starts to be significant compared to the noise peaks at  $S/N \sim 5.8$  (Figure 1). We thus consider only one fidelity curve, which is valid for the entire sample (Figure 2). We note that the curve was computed after having excluded the peaks located in the regions with a primary beam attenuation larger than 90% as we do not expect sources at those radii. We excluded also the region within  $1''$  from the phase center to remove the positive peaks due to the central targets. The inclusion of the central targets would bias indeed the fidelity to higher values. We note that the fidelity is very steep, jumping from 0.2 to 0.8 in a narrow range of S/N.

We used the fidelity to define the final catalog of the serendipitous lines. We included in it the lines with a fidelity higher than 85% (corresponding to a  $S/N = 6.30$  cutoff). This sample includes 12 line detections. We added two more lines with lower fidelity ( $\sim 50\%$ , corresponding to  $S/N \sim 5.98$ ) based on the fact that they present an optical-NIR counterpart (see Sect. 4). This provides a final catalog of 14 serendipitous line detections over the entire ALPINE pointings.

We note that the adopted fidelity cut certainly excludes some genuine detections with low S/N from our catalog. Indeed, if we push down the fidelity to 20% ( $S/N = 5.69$ ), we find 10 more sources. According to their fidelity we expect that the fraction of true sources is low ( $\sim 30\%$ ). However, their exclusion could have an impact on the derivation of the luminosity function (see Sect. 5). We address this point in Sect. 5 and in Appendix C.

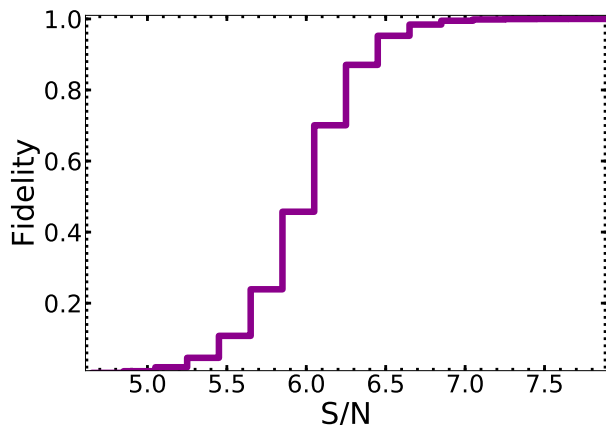


Fig. 2: Fidelity curve for the serendipitous lines detected in ALPINE. The fidelity was computed by comparing the number of positive (genuine lines and noise peaks) and negative (only noise) peaks detected in the 118 ALPINE pointings. We see that the fidelity is a very steep function of the S/N. We adopted a fidelity threshold of 85% (corresponding to a S/N cutoff of 6.3) for the final catalog of the serendipitous lines.

### 3.3. Completeness

Given the purpose of the present work, we need to estimate also the completeness of the sample, i.e. the fraction of recovered lines with respect to the underlying population. We assessed the completeness by simulating  $\sim 50000$  Gaussian-like lines with various peak flux  $F$  and FWHM and by injecting them in datacubes containing pure noise representative of the survey ( $0.14 \text{ Jy/beam km s}^{-1}$  over a line width of  $235 \text{ km s}^{-1}$ ). We injected the lines in random locations in the FOV and along the spectral axis, splitting them in groups of 15 lines per datacube in order to not artificially increase the source confusion. We simulated point sources ( $1.16'' \times 0.78''$ ) since the sources in our catalog are point-like or marginally resolved. However, we note that recent studies reported the existence of extended [C II] structures (e.g., Fujimoto et al. 2019, 2020; Ginolfi et al. 2020a,b), which may cause incompleteness for some faint objects (see Fig. 5 in Fujimoto et al. 2017). The simulated FWHM range is between 50 and 550  $\text{km s}^{-1}$ , while the peak flux varies between 1.0 mJy/beam and 12 mJy/beam in order to widely sample the parameter space of the detected lines (see Figure 3). In particular, for each line the primary beam attenuation is taken into account, i.e. its peak flux is lowered based on the primary beam response depending on its spatial position. We hence derived the completeness  $C$  in the  $j$ -th cell of the (FWHM,  $F$ ) grid as

$$C^j(\text{FWHM}, F) = \frac{N_{\text{rec}}^j(\text{FWHM}, F)}{N_{\text{inj}}^j(\text{FWHM}, F)} \quad (2)$$

where  $N_{\text{inj}}^j$  and  $N_{\text{rec}}^j$  are the number of injected lines and recovered lines by *findclumps* in the cell. We considered cells of 50  $\text{km s}^{-1}$  and 0.5 mJy/beam width. This cell size allows us to accurately evaluate the completeness, with an average number of 60 lines in each cell. We note that completeness is a strong function of the line location in the FOV since the sensitivity decreases significantly as the distance from the phase center increases. We thus evaluated it locally, splitting the lines in four regions based on the primary beam response. In particular, we defined four rings of radii  $R_{<30}$ ,  $R_{30-50}$ ,  $R_{50-70}$ ,  $R_{70-90}$ , in which the primary beam attenuation goes from zero to the 30% (distance from the

phase center  $R_{<30} \leq 7.1''$ ), from 30% to 50% ( $7.1'' < R_{30-50} \leq 10.4''$ ), from 50% to 70% ( $10.4'' < R_{50-70} \leq 13.1''$ ) and from 70% to 90% ( $13.1'' < R_{70-90} \leq 16.4''$ ) and computed the completeness for each of these regions. We avoided the separation in narrower rings since it would have implied a poor statistics of fake sources to adequately sample the completeness.

The diagrams showing the completeness in the four rings are presented in Figure 3. It seems clear from the plots that for equal FWHM and peak flux, lines that are easily detected close to the phase center though become tricky to be detected when observed in the outskirts of the FOV. We show also the location of the lines used to build the [C II] luminosity function (see Sect. 5.1 and Table 1) in the parameter space (FWHM,  $F$ ). All the lines have a completeness higher than 95% in the two most internal regions except for two cases that have completeness between 90% and 70%. In the remaining less sensitive rings the completeness is still higher than 65% in all the cases except for three sources with completeness values below 50%. This fact guaranties that we applied small completeness corrections to our lines when evaluating the luminosity function (see Sect. 5). We present also the completeness curves as a function of the flux peak for fixed FWHM in Appendix B. We can see that at fixed flux peak the completeness is obviously higher for larger lines.

## 4. Identification and sources properties

In order to identify the detected lines we cross-matched their spatial position with the entries in the COSMOS and 3D-HST photometric catalogs (Laigle et al. 2016; Brammer et al. 2012). The astrometry offsets between these catalogs and the ALMA maps are of the order of  $0.1''$  (see Faisst et al. 2020). In addition to this, we checked for counterparts also in the SPLASH (Capak et al. 2012), UltraVista-DR4 (McCracken et al. 2012), 24  $\mu\text{m}$ -selected (LeFloch et al. 2009) and 3 GHz-selected JVLA catalogs (Smolcic et al. 2017). Moreover, we also visually inspected the images from UV to MIR wavelengths in order to look for faint emissions not reported in the catalogs. We classified a galaxy as a physical counterpart of a serendipitous line if their spatial distance is less than  $1''$ . The choice of this value derived from the distance distribution between the serendipitous lines and all the galaxies lying within  $10''$ , which clearly presents a minimum for a distance  $\sim 1''$  for all the catalogs.

Based on the photometric or spectroscopic redshift available, we identified eight lines as [C II] and two lines as  $\text{CO}(J_{\text{up}} = 7, 5)$  transitions. The remaining four detections have an ambiguous identification because of the lack of an optical/NIR or uncertain photometric redshift from ancillary data<sup>2</sup>. All the images and spectra of the serendipitous lines are reported in Appendix A. We refer to a future paper for an analysis of the CO emitting galaxies (Loiacono et al., in prep.), while hereafter we will focus on the [C II] emitters and on the ambiguous lines (i.e. 12 objects in total).

### 4.1. [C II] serendipitous emitters at $4.3 < z < 5.4$

We identified eight lines as [C II] based on the photometric or spectroscopic redshift of the optical/NIR infrared counterpart available from ancillary data. Namely, four out of eight detec-

<sup>2</sup> We note that if we consider the sources with a fidelity down to 20% nine out of the ten new sources do not show any optical/NIR counterpart. There is only one detection associated to a galaxy with a photometric redshift that makes the line emission compatible with a CO(4-3), CO(5-4) or [C I] transition.



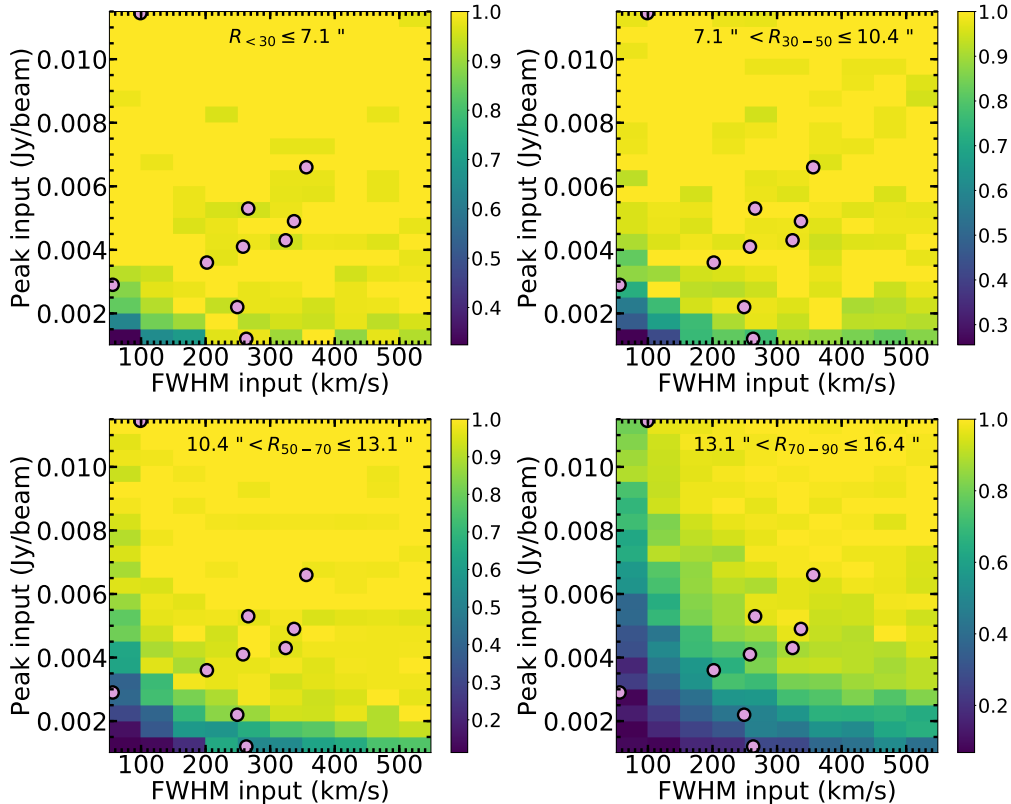


Fig. 3: Completeness (color scale) as a function of the flux peak and the FWHM of a line. The four diagrams correspond to the  $R_{<30}$ ,  $R_{30-50}$ ,  $R_{50-70}$ , and  $R_{70-90}$  regions respectively. As it is evident from their comparison, the completeness is a strong function of the line location in the FOV because of the degrading sensitivity from the phase center to larger radii. The lines used to build the [C II] luminosity function (see Sect. 5) are also shown (filled circles), except for the two brightes ones (i.e., S848185 and S842313) that are located outside the plotted ranges and have completeness equal to one everywhere in the FOV. We show the [C II] serendipitous detections in all the panels since we computed their completeness in each ring when building the luminosity function, independently from the line location in the FOV (see Eq. 3).

tions have an UV–optical spectroscopic redshift (M. Salvato, private communication; Capak et al. 2008, 2011). The remaining four sources have photometric redshifts compatible with [C II] emission (Laigle et al. 2016). The sources have redshift  $4.3 < z < 5.4$ , as expected due to the spectral coverage of ALPINE. We note that among the serendipitous [C II] emitters we recovered the well–studied sub–mm galaxies AzTEC–C17 (here called S842313; Laigle et al. 2016; Schinnerer et al. 2008; Jones et al. 2017) and AzTEC–3 (S848185; Capak et al. 2011; Riechers et al. 2010, 2014).

In addition to these eight detections we found four lines whose identification based on the available photometry is ambiguous. Two of them (S818760 and S859732) do not present any counterpart in the available catalogs and also in the multi–wavelength images (from UV to MIR). The lack of counterparts suggests that these emissions are produced from highly dusty/high- $z$  sources or from gas–rich galaxies with low stellar masses. The most likely associations are thus [C II] at  $4 < z < 6$  or CO transitions at lower redshifts. However, S818760 is located within  $3''$  and has a velocity separation  $< 300 \text{ km s}^{-1}$  from the central target in the same pointing (see Figure 4). As a consequence, it is produced very likely by a companion/interacting source with the UV–target emitting also [C II] but optically faint (see also Jones et al. 2020). A similar argument applies to S5100822662<sup>3</sup>. Differently from

<sup>3</sup> We note that this particular source was included also in the luminosity function of Yan et al. 2020, as they used the total flux of the central

S818760, this source is detected in the available images (see Figure 4) and based on the Laigle et al. (2016) catalog it has a photometric redshift of 0.69. However, the strict association with the ALPINE target in the same pointing (Figure 4) favours a high- $z$  interpretation for this source with the ALMA emission likely due to the [C II] line. It is also possible that the emission in the photometric images is produced by a foreground source that is not related to the ALMA detection (see also Pavesi et al. 2018). Finally, S665626 does not present any counterpart in all but K–band UltraVista image (Romano et al. 2020). This source was studied in detail by Romano et al. (2020) and their modelling seems to favour a [C II] interpretation than a CO line. Follow-up observations are necessary to unambiguously confirm the nature of these four sources.

In the rest of this work we assume that the four unidentified lines are due to [C II] emission. We used both them and the confirmed [C II] to build the luminosity function (Sect. 5). We note that the exclusion of the unidentified lines from it does not alter significantly any of the results. The optical/NIR images of the 12 serendipitous [C II] lines (confirmed and candidates) are shown in Figure 4.

target vuds cosmos 5100822662 (see also Figure 4), which was not deblended from the serendipitous companion (see Bethermin et al. 2020). However, since the latter respects the criterion to be a non–target source (i.e., distance from the phase higher than  $1''$ ) we included it in our calculations. We note that its exclusion from our sample does not alter the result significantly.

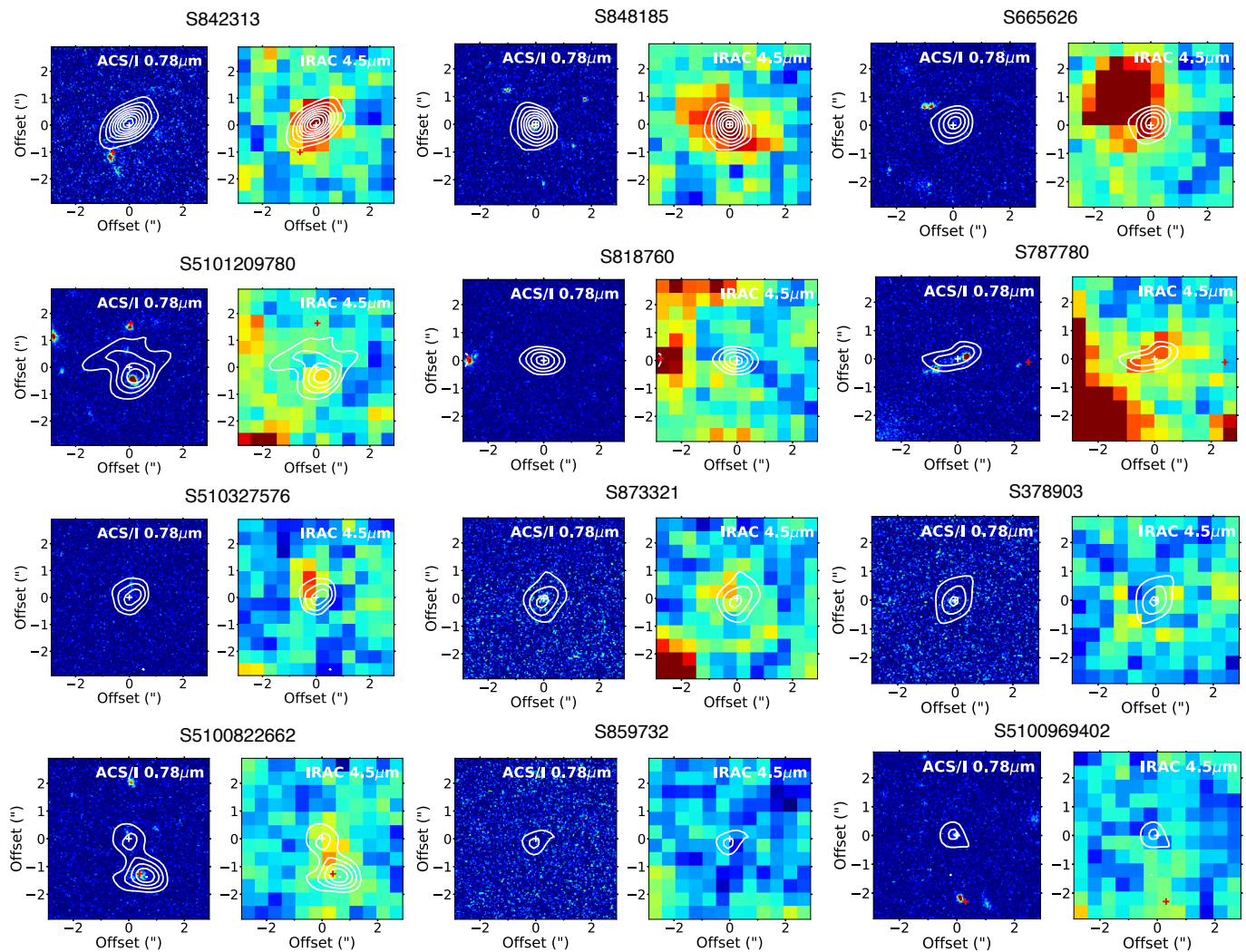


Fig. 4: Images cutouts of the 12 serendipitous lines used to build the [C II] luminosity function. The *HST*-ACS  $0.78 \mu\text{m}$  (Koekemoer et al. 2007, 2011) and *Spitzer*-IRAC  $4.5 \mu\text{m}$  (Capak et al. 2012) are reported. The white contour shows the [C II] emission (lowest level at  $3\sigma$ ). We indicate with a white cross the location of the serendipitous detection while the red cross shows the position of the central target. We can see that for 6 out of 12 lines the distance between the central target and the serendipitous line is  $< 3''$  hence we are possibly witnessing interacting systems. For S5100822662 the [C II] emission is blended with that of the central target.

We estimated the main properties of the [C II] lines (i.e. frequency, FWHM, total fluxes) by performing a single-component Gaussian fit to the continuum subtracted spectrum, with the exception of source S842313 where two Gaussians were adopted to model the line profile as it shows signs of rotation (Jones et al. 2017). To compute the line flux we used the peak flux if the source size is comparable with the beam or we extracted it from a  $3\sigma$  aperture in case the emission is resolved. To distinguish between resolved and unresolved sources, we compare the number of pixels within a  $3\sigma$  aperture with the beam size in pixels. In case the number of pixels exceeds the beam size we labelled the source as resolved. Otherwise we considered the source as not resolved. Then we evaluated the deconvolved sizes of the resolved sources using the 2D fitting tool of *CASA*. We also measured the line fluxes on the moment zero maps but we do not report them since they show consistent results. All the fitted values are reported in Table 1.

#### 4.2. Overdensities around the central targets

The detection of eight confirmed [C II] lines in targeted [C II] observations of  $4 < z < 6$  galaxies suggests that we are witnessing possible overdensities around the central UV-selected galaxies. This is highlighted from the velocity separation  $\Delta v$  between the central target and the serendipitous line in the same pointing. Seven out of eight [C II] have indeed  $|\Delta v| < 750 \text{ km s}^{-1}$ , corresponding to a redshift separation  $|\Delta z| < 0.0154$ . Such a velocity difference suggests that the two galaxies in the same pointing could be physically connected and/or associated to the same large-scale structure. An extended protocluster at  $z \sim 4.5$  (PCI J1001+0220) in the COSMOS field was discovered by Lemaux et al. (2018). Capak et al. (2011) found another protocluster of galaxies in COSMOS at higher redshift ( $z \sim 5.3$ ). In fact some of the serendipitous lines in our sample (e.g. S848185) are well known members of these protoclusters. However there are other detections in our catalog that could constitute potential new members of these overdense regions. This is likely valid for one confirmed [C II] emitter and two [C II] candidates that lie in

Table 1: Catalog of the serendipitous emitters in ALPINE (confirmed and candidates; the latter are marked with a \*). The sources names are labelled according to the ID number of the UV target in the same pointings, preceded by letter "S" that stands for "serendipitous". The reported parameters were estimated using a Gaussian fit (see Appendix A) to the line emission. Also the de-convolved sizes and the distance from the central target (in arcsec) are reported. We show also the redshift separation  $\Delta z$  between the central target and the serendipitous [C II] line in the same pointing. The only galaxy in the "field" sample is S510327576 (see Sect 5). The continuum flux density was measured by Bethermin et al. (2020). We report also the two CO line detections.

ID	Line	S/N	Frequency (GHz)	FWHM ( $\text{km s}^{-1}$ )	Line flux ( $\text{Jy km s}^{-1}$ )	Continuum flux (mJy)	Optical-NIR counterpart	Ancillary redshift	$z_{\text{line}}$	$ \Delta z $	$\log L_{\text{IR}}$ ( $L_{\odot}$ )	Size (arcsec)	Fidelity	Distance (arcsec)
S842313	[C II]	28.18	343.124	$889 \pm 35$	$8.45 \pm 0.29$	$8.24 \pm 0.09$	yes	spec-z	$4.5389 \pm 0.0001$	0.0148	9.72	$0.89'' \times 0.45''$	1.00	1.17
S848185 <sup>a</sup>	[C II]	15.97	301.839	$472 \pm 20$	$11.57 \pm 0.65$	$5.983 \pm 0.227$	yes	spec-z	$5.2965 \pm 0.0002$	0.0034	9.96	$0.91'' \times 0.61''$	1.00	15.17
S665626	[C II]*	10.76	340.752	$324 \pm 19$	$1.47 \pm 0.12$	$0.392 \pm 0.087$	no	...	$4.5775 \pm 0.0001$	0.00020	8.96	$0.66'' \pm 0.46$	1.00	6.35
S5101209780	[C II]	10.66	341.275	$356 \pm 19$	$2.50 \pm 0.18$	...	yes	photo-z	$4.5686 \pm 0.0001$	0.0014	9.19	$1.65'' \times 1.26''$	1.00	1.64
S818760	[C II]*	10.25	341.450	$202 \pm 12$	$0.78 \pm 0.06$	$0.425 \pm 0.104$	no	...	$4.56609 \pm 0.00008$	0.0048	8.69	not resolved	1.00	2.73
S787780	[C II]	9.02	344.866	$258 \pm 14$	$1.13 \pm 0.08$	$0.398 \pm 0.106$	yes	spec-z	$4.51095 \pm 0.00009$	0.00005	8.84	not resolved	1.00	2.49
S510327576	[C II]	8.14	355.894	$337 \pm 23$	$1.75 \pm 0.16$	...	yes	photo-z	$4.3405 \pm 0.0002$	0.2194	9.00	$1.1'' \times 0.84''$	1.00	7.15
S873321	[C II]	8.0	308.730	$266 \pm 39$	$1.50 \pm 0.29$	...	yes	spec-z	$5.1560 \pm 0.0003$	0.0018	9.05	$1.26'' \pm 0.44''$	1.00	12.69
S378903	[C II]	7.5	295.858	$249 \pm 26$	$0.58 \pm 0.08$	...	yes	photo-z	$5.4238 \pm 0.0002$	0.0059	8.67	not resolved	1.00	6.50
S5100822662	[C II]*	6.39	344.256	$56 \pm 7$	$0.17 \pm 0.03$	...	no	...	$4.52071 \pm 0.00004$	0.00021	8.02	not resolved	0.89	1.32
S859732	[C II]*	6.34	343.096	$99 \pm 15$	$1.21 \pm 0.24$	...	no	...	$4.5393 \pm 0.0001$	0.0075	8.87	not resolved	0.86	12.07
S5100969402	[C II]	5.99	340.402	$263 \pm 38$	$0.32 \pm 0.06$	...	yes	photo-z	$4.5832 \pm 0.0002$	0.0047	8.30	not resolved	0.51	2.31
S5110377875	CO(7-6)	9.85	354.109	$183 \pm 9$	$1.35 \pm 0.09$	$3.512 \pm 0.163$	yes	photo-z	$1.27793 \pm 0.00002$	...	7.60	$0.88'' \pm 0.59''$	1.00	6.53
S460378	CO(5-4)	5.97	295.935	$855 \pm 102$	$1.11 \pm 0.18$	$0.680 \pm 0.117$	yes	photo-z	$0.9472 \pm 0.0001$	...	7.12	not resolved	0.48	7.99

**Notes.** <sup>(a)</sup> We note that the difference between the line flux of S848185 reported in this work and in Riechers et al. (2014) is due the use of different apertures. This difference has a negligible impact on the luminosity function.

the spatial region corresponding to PCI J1001+0220 and have a redshift in the range  $4.53 < z < 4.6$  (Lemaux et al. 2018) while other three [C II] lines (two candidates and one confirmed) are possibly located in the outskirts of the same protocluster.

Besides the low velocity/redshift separation, we also see that for four out of eight confirmed [C II] the spatial separation from the central target is less than  $3''$ , corresponding to a physical distance  $< 20$  kpc at  $z \sim 5$  (Figure 4). The number of sources increases to six if we include also the [C II] candidates S818760 and S5100822662. These sources are galaxies likely interacting with the central targets. This is also suggested from the [C II] morphologies, which appear irregular in some cases (see for example S5101209780, Figure 4; Ginolfi et al. 2020a). Therefore, we could be in presence of two kind of overdensities: one on very small scale ( $< 20$  kpc) due to galaxy pairs and/or mergers and another on a larger scale (up to  $\sim 90$  kpc, i.e. the maximum distance allowed by the size of our pointings), related to a more extended structure. We will analyze in detail the overdense environment in a future paper (Loiacono et al., in prep.). The effect of clustering was taken into account when building the [C II] luminosity function (see Sect. 5).

#### 4.3. Relation between [C II] luminosity and SFR

Within the sample of the serendipitous [C II] lines there are five sources for which also the continuum has been detected (Bethermin et al. 2020; Gruppioni et al. 2020). It is well known that there is a correlation between the [C II] luminosity and the SFR (De Looze et al. 2014). Since the latter is well-traced by the total IR luminosity ( $8\text{--}1000 \mu\text{m}$ ), we used the five lines to test if this relation is valid also at  $z \sim 5$ . We included also the two unconfirmed [C II] for which the ALMA continuum has been detected (S818760 and S665626).

The [C II] fluxes were converted to luminosities using Eq. 1 of Solomon et al. (1992) and we propagated the errors from the fitted quantities in Table 1. The total IR luminosity  $L_{\text{IR}}$  was estimated from a SED fitting of the galaxies. We assumed the template of a star-forming galaxy that reproduces most of the

*Herschel* galaxies at  $z \sim 2 - 3$  (Gruppioni et al. 2013). We note that the uncertainty on the total IR luminosity can be up to a factor of 5 depending on the assumed dust temperature (see Faisst et al. 2017; Fudamoto et al. 2020). This uncertainty accounts for about a factor  $\sim 2.5$  on the derived SFR, which we assumed as the typical error of this quantity. Then the  $L_{\text{IR}}$  was converted to SFR using the Kennicutt (1998) relation.

If we compare our values with the local relation of De Looze et al. (2014) (see their Table 3, case HII/starburst) we can see that they are broadly consistent within the  $1\sigma$  errorbars (Figure 5). On the other hand, our points suggest a slightly different slope compared to the model predictions at  $z = 5$  of Lagache et al. (2018). The same trend is also shown by the ALPINE targets (Schaefer et al. 2020; Bethermin et al. 2020), which do not present any evidence of evolution of the SFR- $L_{\text{[C II]}}$  relation between  $z \sim 0$  and  $z \sim 5$ . The only difference is that, compared to the central UV-galaxies, the serendipitous sources sample a different space in the SFR- $L_{\text{[C II]}}$  plane, shifted to higher SFRs and [C II] luminosities. Besides, we note that the SFR of the ALPINE targets takes into account both UV and IR estimates, otherwise the UV-targets would not lie on the De Looze et al. (2014) relation (see Schaefer et al. 2020 for the details). On the other hand we considered the IR-derived SFR only for the serendipitous galaxies. This means that the serendipitous sources detected both in line and continuum have SFRs dominated by the IR emission, with little or negligible contribution from UV.

Therefore both the ALPINE targets and the serendipitous sources seem to independently suggest that there is no significant variation of the relation of De Looze et al. (2014) up to  $z \sim 5$ .

## 5. The [C II] luminosity function at $z \sim 5$

### 5.1. Building of the luminosity function

The 14 [C II] lines (eight confirmed and four candidates) were used to build the luminosity function (LF). We populated the  $i$ -th



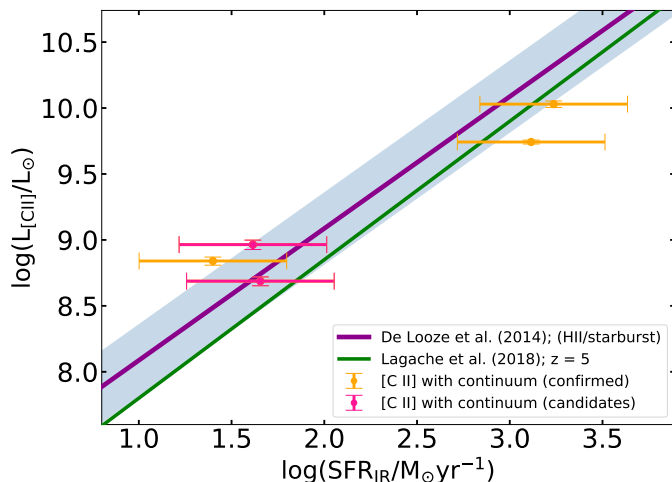


Fig. 5: SFR– $L_{[\text{C II}]}$  relation for the 5 serendipitous lines detected in continuum. The SFR was computed from the IR luminosity of the sources. We can see that the sample is quite consistent at  $1\sigma$  (colored area) with the De Looze et al. (2014) relation (purple line), suggesting that the contribution from the UV–traced SFR is negligible. We compare our results also with the models of Lagache et al. (2018) at  $z = 5$  that suggest a slightly different slope.

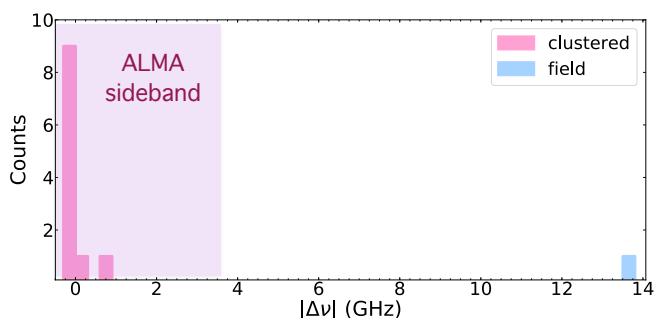


Fig. 6: Offset in frequency between the central UV target and the serendipitous [C II] in the same pointing. We see that the distribution is non–uniform, with several sources lying at a frequency (and hence a redshift) close to that of the central target. We thus defined two sub–samples (named “clustered” and “field” respectively) and evaluated two distinct LFs in order to account for any bias due to overdense regions. The separation between the two sample relies on the frequency width of one ALMA sideband (3.6 GHz), corresponding to a velocity separation  $\geq 2000 \text{ km s}^{-1}$ .

luminosity bin  $d \log L_i$  according to the relation

$$\Phi(L_i) d \log L_i = \sum_j \frac{F_j}{\sum_k C_j^k V^k} \quad (3)$$

where  $\Phi(L_i) d \log L_i$  is the number density of [C II] emitters,  $F_j$  and  $C_j^k$  are the fidelity and completeness of the  $j$ -th source associated to the comoving volume  $V^k$ . The latter was evaluated for the regions  $R_{<30}$ ,  $R_{30-50}$ ,  $R_{50-70}$ ,  $R_{70-90}$  in order to take into account the completeness variation in the FOV, hence the  $k$  index goes by the four rings. Only the sources with completeness and fidelity equal to unity everywhere in the FOV would have been indeed observable within the total comoving volume  $V_{\text{TOT}}$  covered by the 118 ALPINE pointings. This volume was evaluated as  $V_{\text{TOT}} = \sum_i^{118} A_i \Delta D_c(z_i) = 9810 \text{ Mpc}^3$  where  $A_i$  is the area with a primary beam attenuation  $< 90\%$  covered by each

ALPINE pointing and  $\Delta D_c(z_i)$  is the difference between the comoving distances of the [C II] line at the beginning and at the end of the ALMA sidebands for the  $i$ -th pointing. This difference was computed after having excluded 3–4 channels at the beginning and at the end of each sideband to account for border effect (i.e. noisy channels). We note that we excluded the central  $R < 1''$  region from each pointing in the computation of the volume. As the luminosity bin size we considered 0.5 dex in order to have at least one source per bin. The adopted bin spacing is 0.25 dex in luminosity. Although the bins are not independent this choice gives the advantage to better highlight the luminosity distribution of the sample. We point out that we did not split the [C II] lines in different redshift bins because of the poor statistics, hence our LF refers to an average redshift  $z \sim 5$ . As done in Sect. 4.3, we evaluated the [C II] luminosities following Solomon et al. (1992) (see Table 1 for the values). The errorbars associated to each luminosity bin are computed as the Poissonian uncertainties corresponding to  $1\sigma$  since the source number in each bin is small (Gehrels 1986) and thus constitutes the major uncertainty.

Before computing the LF, we splitted the [C II] lines in two sub–samples. As we saw in Sect. 4, seven out of eight confirmed [C II] have a redshift separation from the central targets in the same pointings  $|\Delta z| < 0.015$  (corresponding to a velocity separation  $< 750 \text{ km s}^{-1}$ ). This number increases to 11 out of 12 if we include also the four unconfirmed [C II]. This means that their LF could be not representative of the field galaxy population since it is likely biased by the presence of overdensities around the UV–selected targets. The only exception is S510327576, which has a redshift separation  $|\Delta z| = 0.2195$  ( $|\Delta v| \sim 1.2 \times 10^4 \text{ km s}^{-1}$ ) and thus is not related to the central target. This could be the only [C II] line not associated to clustered structures, i.e. the only genuine field source in our sample.

In order to study the effect of clustering on the LF, we thus considered two separate sub–samples, each of them containing the lines with a frequency offset from the central target lower/higher than one ALMA sideband ( $\Delta \nu \sim 3.6 \text{ GHz}$ ; see Figure 6). This separation corresponds to a redshift difference  $\geq 0.04$  and to a velocity separation  $\geq 2000 \text{ km s}^{-1}$  (see also Henawi et al. 2010 who used a similar velocity separation in a study on quasars pairs). In this way we defined the “clustered” and “field” sub–samples, containing 11 and one sources respectively. Also the survey volume was splitted consistently, obtaining a total comoving volume of  $5026 \text{ Mpc}^3$  for the clustered sub–sample and  $4784 \text{ Mpc}^3$  for the field one. Thus we built a separate [C II] LF for each sub–sample (Figure 7 and Table 2). The median luminosity of the clustered sample ( $\log(L/L_\odot) = 8.96 \pm 0.14$ ) is very similar to the luminosity of the field one ( $\log(L/L_\odot) \sim 9.0$ ). However, we remind that the field LF is based on one object only and therefore it could present also galaxies at higher luminosity that we do not detect for the limited survey volume. Despite the similar median luminosity, the clustered LF shows objects with luminosity of about one order of magnitude higher than the field. If this trend were confirmed by a larger sample of galaxies it would highlight a dependence between clustering and the [C II] luminosity, as already shown based on other tracers (e.g. Hawkins et al. 2001).

In Appendix C we report also the LF computed including sources with fidelity as low as 20% (see Figure C.1). In Sect. 3.2 we cut indeed our catalog of serendipitous detections at a fidelity of 85%, with only one [C II] line (S5100969402) having a fidelity  $\sim 50\%$ , in order to study a very robust sample. However this sample is obviously incomplete at low luminosity. We thus calculated the clustered and field LFs for two new

Table 2: Luminosity functions for the clustered and field sample considering the eight confirmed and four candidates [C II]. We reported also the number of sources in each luminosity bin. We indicated with the bold font the values corresponding to independent luminosity bins.

$\log(L/L_{\odot})$	$\log(\Phi_{\text{clust}}/\text{Mpc}^{-3}\text{dex}^{-1})$	$N_{\text{clust}}$	$\log(\Phi_{\text{field}}/\text{Mpc}^{-3}\text{dex}^{-1})$	$N_{\text{field}}$
8.25	<b><math>-2.94^{+0.36}_{-0.45}</math></b>	<b>2</b>	...	...
8.50	$-2.87^{+0.29}_{-0.34}$	3	...	...
8.75	<b><math>-2.65^{+0.22}_{-0.24}</math></b>	<b>5</b>	...	...
9.00	$-2.69^{+0.22}_{-0.24}$	5	$-3.37^{+0.52}_{-0.77}$	1
9.25	<b><math>-3.09^{+0.36}_{-0.45}</math></b>	<b>2</b>	<b><math>-3.37^{+0.52}_{-0.77}</math></b>	<b>1</b>
9.50	$-3.40^{+0.52}_{-0.77}$	1	...	...
9.75	<b><math>-3.10^{+0.36}_{-0.45}</math></b>	<b>2</b>	...	...
10.00	$-3.40^{+0.52}_{-0.77}$	1	...	...

sub-samples, in which we included also low fidelity (i.e. low luminosity) sources. The fidelity cut of 20% adds nine lines to our catalog of [C II] candidates (we excluded the only source possibly associated to CO emission; see Sect. 4). We note that none of these lines presents an optical/NIR counterpart hence their redshift is unconstrained from ancillary data. Therefore, for these sources we can only assume that their emission is due to [C II]. We see that the shape of both clustered/field LFs remains quite unchanged, with the field LF sampled by more sources now. Also in this case the field sources lie at lower luminosities compared to the clustered sample. However, this is not surprising since, as the fidelity lowers, also the line flux decreases and hence we expect the population of the low luminosity bins only for both field and clustered sources.

Finally, in Sect. 4.2 we saw that the clustered sources are possibly part of two different types of overdensity, one associated to interactions/mergers (scale  $< 20$  kpc) with the central UV-selected galaxy and the other associated to a more extended structure (up to  $\sim 90$  kpc). In order to overcome the bias introduced by the interacting systems, we excluded from the LF the 6 sources with a spatial distance  $< 3''$  from the central target (Figure 4). We report the derived LF in Appendix C. The new points are consistent within the errors with the LF computed using all the clustered sources. Overall the faint end of the LF results lower compared to the case in which all the 11 clustered [C II] are considered. However, this has a negligible effect on the derivation of quantities like the fitted parameters of the Schechter function and the star formation rate density (see Sect. 5.3 and Sect. 6).

## 5.2. Comparison with observations and models

### 5.2.1. Luminosity functions from ALPINE

In this section we discuss our LFs in relation to those from other works (Figure 7).

We start by comparing our results with the other  $z \sim 5$  LFs based on the ALPINE data. First of all, we consider the [C II] LFs presented in the companion paper of Yan et al. (2020). These LFs were built using the 75 [C II] central UV-targets in the two redshift ranges  $4.40 < z < 4.58$  and  $5.13 < z < 5.85$ . Globally, we see that the clustered LF predicts more [C II] emitters than the Yan et al. (2020) sample. This was expected due to clustering effects and also because the LF of the central targets is based on UV-selected galaxies, hence it is likely missing the most obscured galaxies. On the other hand the field LF is quite consistent with the targets LFs, showing a slight excess in the highest lumi-

nosity bin.

Then we compare our sample with the LF based on the sources serendipitously detected in the rest-frame FIR continuum (Gruppioni et al. 2020, Bethermin et al. 2020). The 118 ALPINE pointings revealed indeed a wealth of serendipitous continuum emitters in a wide range of redshifts. These sources were used to build a rest-frame  $250 \mu\text{m}$  LF and a total IR LF from  $z = 0.5$  to  $z = 6$  (see Gruppioni et al. 2020 for the details). For our comparison we considered the IR LF in the highest redshift interval  $4.5 < z < 6$ , where the companions of the central targets have been removed (green water hexagons; see Table 2 of Gruppioni et al. 2020). The IR luminosities ( $8\text{--}1000 \mu\text{m}$ ) were first converted to SFRs according to the Kennicutt (1998) relation. We note that the computed SFRs do not include the UV contribution, therefore they can be considered as lower limits. However we do not expect the UV contribution to be significant since the sources are selected to be dusty (i.e., FIR/sub-mm emitters). The SFRs were then used to derive the [C II] luminosities following the De Looze et al. (2014) relation (case HII/starburst), scaled for a Chabrier (2003) IMF. Globally, the clustered LF presents a higher number density (up to about 1 dex) and higher luminosity objects than the IR-derived [C II] LF of Gruppioni et al. (2020). The difference in the lower luminosity bins is however enhanced by the fact that these bins are strongly incomplete in the continuum survey (see Bethermin et al. 2020). On the other hand, there is perfect agreement between the field LF and the LF derived from Gruppioni et al. (2020). However, if we show the IR-derived [C II] LF that includes also the companions of the central targets for  $L_{[\text{C II}]} > 10^{0.5}L_{\odot}$  (magenta hexagons; see Gruppioni et al. 2020) we find that, in this luminosity range, the clustered [C II] LF and the IR-derived [C II] LF are nicely consistent within the errorbars. This is due to the fact that these luminosity bins include the same sources, clustered around the central targets, detected both in line and in continuum.

### 5.2.2. Observed luminosity function at high and low- $z$

Now we can move on to comparing our results to other works in the literature, at high and low- $z$ . We see that our LFs are consistent with previous estimates at  $z = 4.4$  and  $z \sim 5$  from Swinbank et al. (2012) and Capak et al. (2015). Swinbank et al. (2012) started from an original  $870 \mu\text{m}$  selection of galaxies with LABOCA (Weiß et al. 2009) and considered the only two galaxies for which the [C II] line was detected in a subsequent ALMA follow-up. However, the low continuum detection rate of the ALPINE targets (20%; Bethermin et al. 2020) compared to the line detection rate (64%) suggests that a considerable fraction of [C II] emitting galaxies can be missed when starting from continuum pre-selected samples, hence the LF of Swinbank et al. (2012) likely provides a lower limit to the number density of the [C II] emitters. In case of the estimate from Capak et al. (2015) we use the value reported in Hemmati et al. (2017). Also in this case the data likely provide a lower limit to the true distribution, since the targets of Capak et al. (2015) are Lyman break galaxies, i.e. UV-selected objects, and hence [C II]-bright but optically-faint objects are not taken into account in this LF. Moreover, in this estimate the [C II] serendipitous emitters in the ten pointings of Capak et al. (2015) are not considered (e.g. AzTEC-3, Riechers et al. 2010; CRLE, Riechers et al. 2010).

Our values are also above the LF of Cooke et al. (2018). This study considers [C II] emitting galaxies pre-selected based on their SCUBA2  $850 \mu\text{m}$  flux density (Geach et al. 2017), hence also this estimate provides a lower limit.

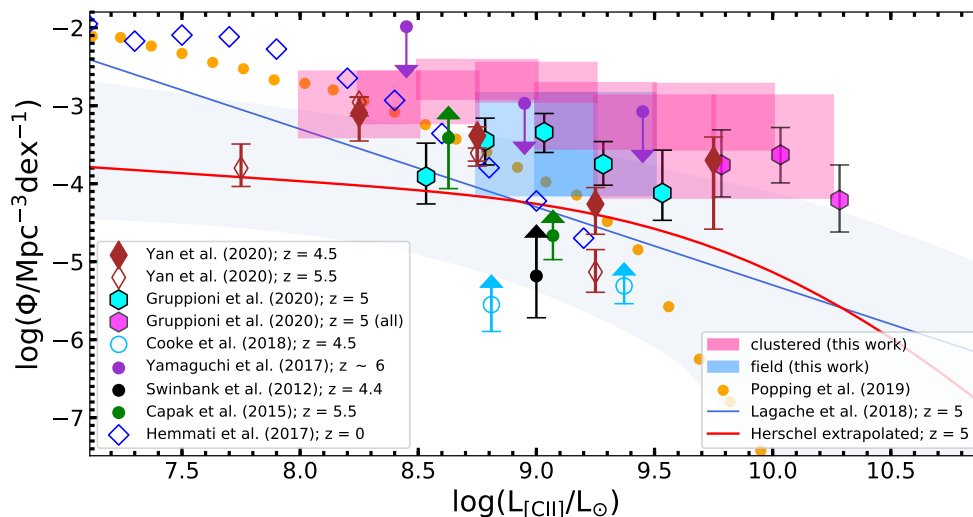


Fig. 7: [C II] luminosity functions (LFs) at  $z \sim 5$  from the serendipitous sources in ALPINE compared to other works in the literature. We splitted the lines in two sub-samples, called “clustered” (pink) and “field” (azure) respectively, and built two separate LFs. Compared to the clustered LF, the field one lies at lower luminosities. We compare our [C II] LFs at  $z \sim 5$  with other [C II] LFs at high and low- $z$ . Overall, the estimates from the clustered sample lie above the LFs of the ALPINE targets (Yan et al. 2020) likely because they include also UV-dark galaxies and because of the clustering effect. On the other hand the field LF seems to be quite consistent with the targets ones except for the highest luminosity bin. There is a nice agreement between the field [C II] LF and the IR-derived [C II] LF based on the ALPINE serendipitous sources detected in continuum (Gruppioni et al. 2020). The agreement persists at  $L_{[\text{C II}]} > 10^{9.5}L_{\odot}$  for the clustered sample if the companions of the central targets are included in the IR-derived [C II] LF of Gruppioni et al. (2020). The clustered LF is up to  $> 1$  dex higher than the local [C II] LF (Hemmati et al. 2017). Also the field LF predicts an excess of [C II] emitters at  $L_{[\text{C II}]} > 10^9L_{\odot}$ , suggesting a possible evolution of the [C II] LF between  $z \sim 5$  to  $z \sim 0$ . The field LF appears in agreement with the models predictions of Popping et al. (2019).

We also compared our estimates with measurements at higher redshift (Yamaguchi et al. 2017). The points of Yamaguchi et al. (2017) represent upper limits to the [C II] LF at  $z \sim 6$ . We can see that the field LF is well consistent with the upper limits. On the other hand the clustered LF seems to predict more [C II] emitters than Yamaguchi et al. (2017) at  $L_{[\text{C II}]} = 10^{8.75}L_{\odot}$  probably because it is biased to an overdense environment.

It is interesting to compare our work also with an extrapolation of the *Herschel* LF at  $z \sim 5$  (Gruppioni et al. 2013; Gruppioni et al., in prep.). The extrapolation was performed using the SCUBA2 number counts (Geach et al. 2017) to constrain the evolution at high-redshift (Gruppioni & Pozzi 2019). The IR luminosities were thus converted to SFRs using the Kennicutt (1998) relation and the SFRs were transformed in [C II] luminosities following De Looze et al. (2014). We note that the same approach has been already used for deriving the CO LF in Vallini et al. (2016), which successfully reproduces the observed CO LF of ASPECS (Decarli et al. 2019). Interestingly, we see that the global shapes of the clustered LF and the *Herschel*-derived one are in good agreement, with both LFs predicting [C II] emitters with very high luminosities ( $L_{[\text{C II}]} > 10^9L_{\odot}$ ), with at least some of the discrepancy coming from the fact that the *Herschel* extrapolation was not intended to account for the clustering inherent in the ALPINE serendipitous sample.

Finally, we discuss how the  $z \sim 5$  [C II] LF compares with the  $z \sim 0$  values (Hemmati et al. 2017) to underline potential evolutionary effects. We can see that the clustered LF shows a strong evolution both in number density (up to  $> 1$  dex) and in luminosity between  $z \sim 0$  and  $z \sim 5$ . The field LF suggests also a possible excess of objects at  $L_{[\text{C II}]} > 10^9L_{\odot}$  compared to the local value. The two LFs are however consistent within

$2\sigma$ . A higher statistics for the field sample is necessary to draw robust conclusions about any evolutionary trend that is independent from clustering.

### 5.2.3. Theoretical predictions

We also compare our results with model predictions for the early Universe. First of all, we considered the models for the [C II] LF by Popping et al. (2019). These are semi-analytical models that include radiative transfer modelling. We can see that the clustered [C II] LF predicts a higher number of objects than the models expectations at  $z \sim 5$ , with a disagreement that rises with increasing luminosity. A similar disagreement with models predictions is seen also for the CO LFs at high- $z$  (Riechers et al. 2019) and for the IR LF at  $z \sim 2$  (Gruppioni et al. 2015). On the other hand, the field LF appears quite consistent with the models. Further statistics would be useful to constrain the bright end of the field LF and disentangle if it remains flat at  $L_{[\text{C II}]} > 10^9L_{\odot}$  (as for the clustered sample) or if it declines as shown by models.

Then we examine the predictions at  $z \sim 5$  by Lagache et al. (2018). This is also a semi-analytic model combined with a photoionization code. We note that at luminosities between  $10^9L_{\odot}$  and  $10^{10.5}L_{\odot}$  the Lagache et al. (2018) curve is not very different from the *Herschel* extrapolation. Compared to Popping et al. (2019), this model predicts more [C II] emitters at  $L_{[\text{C II}]} > 10^{9.5}L_{\odot}$ , with luminosities consistent with the observed values for the clustered sample. However, we see that our observed LFs (especially the clustered one) show a higher number density of objects ( $> 1$  dex), which is not predicted by this model.

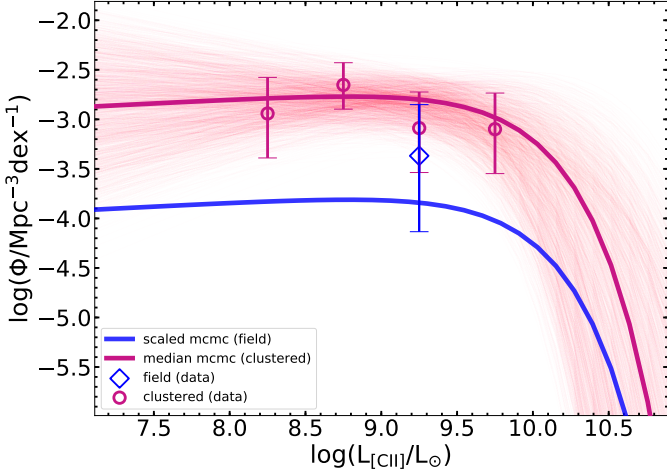


Fig. 8: Schechter functions for the clustered (pink) and field (azure) [C II] luminosity functions (LFs). Also the observed LFs corresponding to the independent luminosity bins are indicated (same color code). We fitted  $\log \Phi^*$ ,  $\log L^*$ , and  $\alpha$  for the clustered LF using a Markov Chain Monte Carlo (MCMC) method. We assumed for the field sample the same  $\alpha$  and  $\log L^*$  of the clustered LF and we scaled the normalization of the clustered LF by a factor of 11 (corresponding to the ratio between the number of clustered/field sources). The shaded area (pink; clustered sample) shows the MCMC realizations within the 16th and 84th percentile, hence it corresponds approximately to  $1\sigma$  errorbars. In case of the scaled field LF the  $1\sigma$  errors (blue area) were computed from the uncertainties of  $\log \Phi^*$  of the clustered sample and the Poissonian uncertainty (at  $1\sigma$ ) on 11 counts.

Table 3: Schechter parameters for the clustered and field sample. We report the uncertainties corresponding to the 16th and 84th percentile ( $\sim 1\sigma$ ).

Parameter	Clustered sample	Field sample
$\log(L^*/L_\odot)$	$9.88^{+0.54}_{-0.55}$	9.88 (fixed)
$\log(\Phi^*/\text{Mpc}^{-3}\text{dex}^{-1})$	$-3.01^{+0.44}_{-0.61}$	$-4.05^{+0.62}_{-0.72}$
$\alpha$	$-0.92^{+0.56}_{-0.44}$	-0.92 (fixed)

### 5.3. Fitting with a Schechter function

We performed a fit to the [C II] LFs with the Schechter (1976) function written in logarithmic form (Figure 8). Given the element of luminosity  $d\log L$ , the number of objects  $\Phi(L)d\log L$  falling in the bin is:

$$\phi(L)d\log L = \ln 10 \Phi^* \left(\frac{L}{L^*}\right)^{\alpha+1} \exp^{-\frac{L}{L^*}} d\log L \quad (4)$$

where  $\alpha$  is the faint-end slope and  $L^*$  and  $\Phi^*$  are the luminosity and the value of the LF at the “knee” respectively. For simplicity, we fitted the  $\log \Phi(L)$  and thus also the logarithms of  $L^*$  and  $\Phi^*$ . We fitted the clustered LF only because of the low statistics of the field LF and the only one independent bin. Before performing the fit, we rebinned the clustered and field LF adopting a bin spacing of 0.5 dex instead of 0.25 dex (see sec 5.1), i.e. equal to the bin width. This ensures that the number counts in the bins are independent as well as the uncertainties on the fitted points.

To derive a first estimate of the fitted parameters, we performed a fit based on the maximum likelihood criterion. The

best-fit values were used as initial guesses for a Markov Chain Monte Carlo (MCMC) method with the Python package *emcee* (Foreman-Mackey et al. 2013). We assumed for  $\alpha$ ,  $\log L^*$ ,  $\log \Phi^*$  uniform priors. We preferred uniform than Gaussian priors as they represent the simplest possible choice, since the probability distribution of these parameters is not known a-priori. In case of  $\log L^*$  we limited the upper boundary for the luminosity prior to 10.5, corresponding to an IR luminosity of  $10^{13.5}L_\odot$ , assuming a fiducial ratio between [C II] and IR luminosity of  $10^{-3}$  (Díaz-Santos et al. 2013). This is a reasonable upper boundary to the IR luminosity motivated by pre-existing IR LFs at lower redshifts (Gruppioni et al. 2013; Vallini et al. 2016). The validity of the  $L_{[\text{C II}]}$ -SFR relation, where the latter quantity is derived from continuum estimates, for our sample (see Sect. 4.3) suggests that this is a trustworthy assumption.

The best values for  $\alpha$ ,  $\log L^*$ ,  $\log \Phi^*$  for the clustered LF are reported in Table 3. These values were evaluated as the medians of the posterior probability distributions. The reported uncertainties correspond to the 16th and 84th percentile of the posteriors (equivalent to about  $1\sigma$  in case of Gaussian posteriors).

We then computed the Schechter function also for the field LF. Since it was not possible to directly fit the data, we scaled  $\Phi^*$  by a factor 1/11 (i.e. the ratio between the number of field and clustered sources), under the assumption that the shape of the two LFs is similar. In this way the integration of the luminosity function over the accessible volume and luminosity predicts a number of sources equal to the observed one (i.e., one source). See also Marshall et al. (1983). Moreover, this approach has the advantage to be independent from the binning of the LF. We obtained a value of  $\log(\Phi^*/\text{Mpc}^{-3}\text{dex}^{-1}) = -4.05^{+0.62}_{-0.72}$  where the errors were propagated from the uncertainty on  $\log \Phi^*$  of the clustered sample and the Poissonian error on the ratio 11:1. We note that the normalization determined in this way results consistent with the normalization that would be obtained by performing a Schechter fit to the field LF in which  $\alpha$  and  $\log L^*$  are fixed to the clustered values.

The Schechter functions of the clustered and field samples were used to estimate the cosmic star formation rate density (Sect. 6).

## 6. Star formation rate density at $z \sim 5$

We know that the [C II] line is a SFR indicator (De Looze et al. 2014). Therefore the [C II] LF, providing the total [C II] luminosity budget, can be used to estimate the cosmic star formation density (SFRD). First, we integrated the Schechter functions for the field and clustered sample in order to obtain the [C II] luminosity density  $\rho_{L_{[\text{C II}]}} = \int \Phi(L')L'd\log L'$ . We considered in the integration all the luminosities higher than  $10^7L_\odot$ . However, integrating from lower luminosities does not alter significantly the final estimates because the LFs are quite flat. In case of the clustered sample, the integration was performed for all the realizations of the MCMC. On the other hand, for the field sample, we integrated the best curve and the curves corresponding to the  $1\sigma$  errorbars. Then, we converted the luminosity densities to SFRDs using the relation (see Table 3 of De Looze et al. 2014; case HII/starburst)

$$\log \dot{\rho}_\star = -7.06 + 1.00 \log \rho_{L_{[\text{C II}]}} + \log 0.94, \quad (5)$$

where  $\dot{\rho}_\star$  is the SFRD and the last term accounts for scaling the De Looze et al. (2014) relation from Kroupa (2001) to a Chabrier (2003) IMF. We note that the working assumption of a non-evolving  $L_{[\text{C II}]}$ -SFR relation is not trivial (Vallini et al.



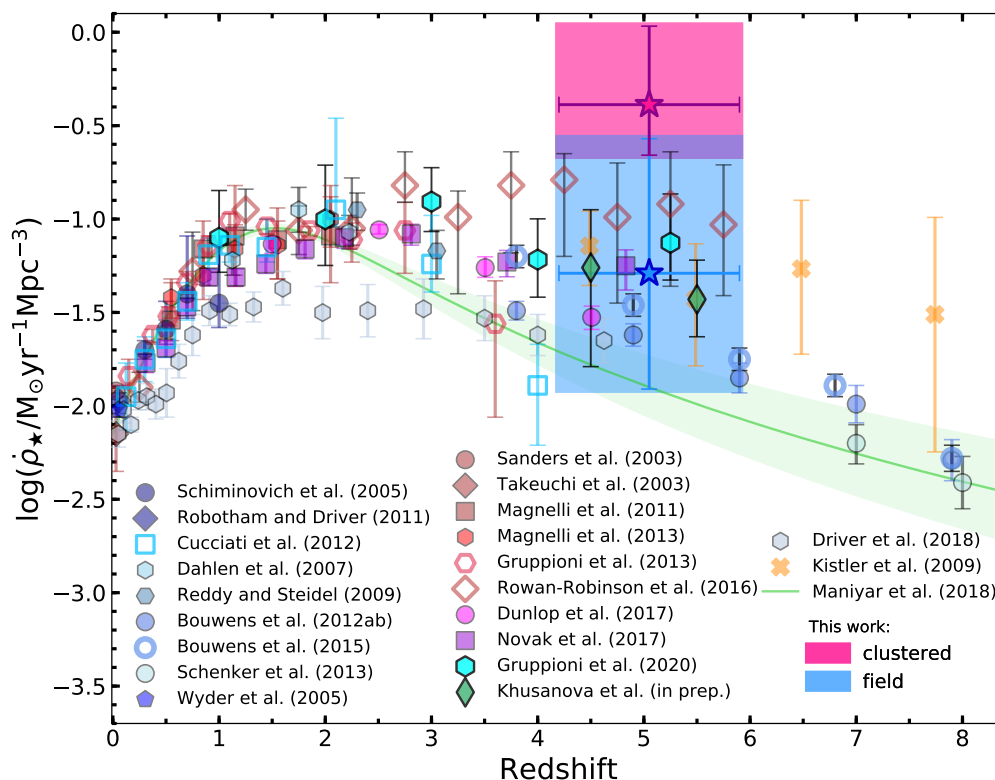


Fig. 9: Cosmic star formation rate density (SFRD) across cosmic time. Both the estimates from the “field” and “clustered” sample are shown (azure and pink box respectively). We compare our measurements with estimates available from the literature based on multiwavelength observations. The SFRD derived from the clustered [C II] LF at  $z \sim 5$  is about one order of magnitude higher than the current measurements at that redshift. On the other hand, the SFRD of the field sample spans values compatible with both UV and IR-derived estimates, with an average value a factor  $\sim 1.6$  higher than the estimates based on UV-surveys. We consider the SFRD from the field sample as representative of the overall galaxy population since the clustered estimate is biased by overdensities around the targeted [C II].

2015; Carniani et al. 2018). However, we mentioned in Sect. 4.3 that it seems to work at least for the serendipitous [C II] detected in continuum. Furthermore, the validity of this conversion is independently confirmed by the ALPINE targets which, as discussed in Bethermin et al. (2020) and Schaerer et al. (2020), lie within  $1\sigma$  on the De Looze et al. (2014) relation. In this way, we obtained for the clustered sample a SFRD probability distribution based on all the MCMC realizations. We considered the median value of the distribution as the best estimate of the SFRD from the clustered sample while, as done before, we reported the uncertainties corresponding to the 16th and 84th percentile (Table 4). On the other hand, for the field sample, we considered the SFRD value corresponding to the integration of the best curve with the associated errorbars (see Figure 8).

In Figure 9 we compare our results with previous estimates from the literature<sup>4</sup>, based on UV surveys (Schiminovich et al. 2005; Wyder et al. 2005; Dahlen et al. 2007; Reddy & Steidel 2009; Robotham & Driver 2011; Bouwens et al. 2012a,b; Cucchiati et al. 2012; Schenker et al. 2013; Bouwens et al. 2015) and IR, mm and radio selections of galaxies (Sanders et al. 2003; Takeuchi et al. 2003; Magnelli et al. 2011, 2013; Gruppioni et al.

2013; Rowan-Robinson et al. 2016; Dunlop et al. 2017; Novak et al. 2017). We show also the measurements derived from optical–NIR observations (Driver et al. 2018) and gamma-ray bursts (Kistler et al. 2009). We plot also the models predictions of Maniyar et al. (2018) based on the cosmic microwave background. Finally, we compare our results with other independent measurements of the SFRD based on the ALPINE data. In particular, we show the results derived from the serendipitous sources detected in continuum (Gruppioni et al. 2020) and the SFRD inferred from the ALPINE central targets (Khusanova et al., in prep.).

We can see that the SFRD derived from the clustered sample is almost  $10\times$  higher than the field value and the current estimates of the SFRD at  $z \sim 5$  from the literature. We consider the SFRD computed using the field sample as the most likely estimate of the cosmic star formation activity at  $z \sim 5$ . The measurement based on the clustered LF could be indeed biased by companions around the targeted [C II], which might not have been observed if we had started from a pure “blind” survey. Therefore, the clustered estimate may not be representative of the overall population of galaxies.

We know that a relevant question deals with the relative contribution of the unobscured wrt obscured star formation across cosmic time. The former is well sampled by UV surveys from

<sup>4</sup> For the works before 2014 we show the values reported in Table 1 of Madau & Dickinson (2014), except for Kistler et al. (2009) that is not included in the table.



Table 4: Cosmic star formation rate density (SFRD) from the clustered and field [C II] LFs.

	Clustered	Field
$\log\left(\frac{\dot{\rho}_\star}{M_\odot \text{yr}^{-1} \text{Mpc}^{-3}}\right)$	$-0.39^{+0.42}_{-0.27}$	$-1.29^{+0.72}_{-0.62}$

$z \sim 0$  up to  $z \sim 10$  (Bouwens et al. 2015; Oesch et al. 2018). On the other hand, the latter is captured by surveys at longer wavelengths, typically IR and sub-mm. At the moment the obscured star formation is well constrained by statistically robust samples up to  $z \sim 3$  while at higher redshift its contribution to the total budget of star formation is quite uncertain. If we look at the average value of the SFRD based on the field LF we can see that it is a factor  $\sim 1.6$  higher than the measurement based on UV surveys (Bouwens et al. 2015). This means that it might be a fraction of (obscured) star formation that is not captured by UV surveys. However, when looking at the errors, we see that our estimate varies between values that are completely consistent with the UV estimates (i.e., negligible obscured star formation) to values that are about ten times higher than the UV measurements. A scenario consisting in a significant fraction of dust obscured star formation already in place at  $z > 4$  is suggested by IR, mm and radio selections of galaxies (Bouwens et al. 2015; Novak et al. 2017; Gruppioni et al. 2020). Because of the large uncertainties our measurement does not allow us to assess the importance of obscured wrt unobscured star formation at  $z \sim 5$ . Further observations of larger volumes in the sky are thus necessary to better constrain the [C II]–derived SFRD.

## 7. Summary and conclusions

We summarize the main results of this work:

1. We built the [C II] luminosity function (LF) using the lines serendipitously discovered in the ALMA ALPINE large program. This is the first LF at  $z \sim 5$  based on galaxies purely selected based on their [C II] line emission. First of all, we performed a blind search in the 118 ALPINE pointings that revealed several unexpected lines. We assessed the fidelity and the completeness of the detections. The final catalog of the serendipitous sources includes 14 line emitters with high fidelity ( $> 85\%$  for 12 out of 14 detections). We identified the line emission by comparing its spatial position with the available photometric catalogs and multi-wavelength images. Out of the 14 lines, eight are [C II] lines at  $4.3 < z < 5.4$ , supported by a spectroscopic or photometric redshift from ancillary data. Two out of 14 lines are CO transitions at lower redshift. Finally four out of 14 lines have a more tricky nature because they are not associated to any optical/NIR counterpart or have an uncertain photometric redshift. However, three of them are very likely [C II] emitters based on the strict association with the central target or individual SED modelling. Observational follow-up are necessary to unambiguously confirm the nature of these sources.
2. The eight [C II] emitters and the four lines with an ambiguous identification were used to build the [C II] LF. We found that 11 out of 12 sources are strongly clustered around the central target in the same pointing since they are located at very similar redshifts ( $|\Delta z| < 0.0154$ , corresponding to  $|\Delta v| < 750 \text{ km s}^{-1}$ ). The discovery of these sources could be
3. We compared our LFs with other works, both observational and theoretical. We found that, globally, the clustered LF suggests an excess of sources compared to the LFs of the ALPINE targets (Yan et al. 2020). This could be due both to clustering effect and to the fact that the targets LFs are based on UV-selected galaxies, hence they do not include highly dusty objects. On the other hand the field LF is quite consistent with the targets LFs. Our measurements, especially the field one, are quite in agreement with the estimates from the serendipitous continuum sources found in ALPINE (Gruppioni et al. 2020). The estimates from the field LF are also in agreement with the semi-analytical models of Popping et al. (2019) at  $L_{[\text{C II}]} \sim 10^9 L_\odot$ . Observations of more extended volumes will be useful to assess if this agreement persists also at higher luminosities. Finally, both the clustered and field LFs suggest a possible evolution of the [C II] LF from  $z \sim 5$  to  $z \sim 0$ . Also in this case observations of larger samples are necessary to confirm this trend.
4. We performed a Schechter fit to the clustered LF. Then, we scaled the fitted normalization function for a factor 11 (corresponding to the ratio between the number of clustered and field sources) to reproduce the field LF, under the assumption that they have the same shape. From the Schechter fits we estimated the [C II] luminosity density. This value was then converted to star formation rate density (SFRD) using the relation of De Looze et al. (2014) (case HII/starburst). We found that the clustered sample shows values that are up to a factor  $\sim 10$  higher than the current estimates from the literature. We considered the estimate obtained for the field sample as representative of the average galaxy population, since the clustered estimate could be biased to high-density environment. The average SFRD results a factor  $\sim 1.6$  on average higher than the current estimates from UV surveys. However, because of the large errorbars, it is not possible to say if this value could be indicative of significant fraction of obscured star formation at  $z \sim 5$ , as suggested by IR and mm selections of galaxies. Observations of larger samples are necessary to better constrain the SFRD of the Universe.

This work represents the first effort to study the luminosity function of [C II] emitters at  $z \sim 5$  based on sources purely selected due to their [C II] emission and provides new insight onto the properties of the star-forming medium in a poorly known cosmic epoch.

*Acknowledgements.* This paper is based on data obtained with the ALMA Observatory, under Large Program 2017.1.00428.L. ALMA is a partnership of ESO (representing its member states), NSF (USA) and NINS (Japan), together with

- NRC (Canada), MOST and ASIAA (Taiwan), and KASI (Republic of Korea), in cooperation with the Republic of Chile. The Joint ALMA Observatory is operated by ESO, AUI/NRAO and NAOJ. Based on data products from observations made with ESO Telescopes at the La Silla Paranal Observatory under ESO programme ID 179.A-2005 and on data products produced by TERAPIX and the Cambridge Astronomy Survey Unit on behalf of the UltraVISTA consortium. This program is supported by the national program Cosmology and Galaxies from the CNRS in France. F.L., C.G., A.C., F.P. and M.T. acknowledge the support from the grant PRIN MIUR 2017. D.R. acknowledges support from the National Science Foundation under grant numbers AST-1614213 and AST-1910107 and from the Alexander von Humboldt Foundation through a Humboldt Research Fellowship for Experienced Researchers. G.C.J. acknowledges ERC Advanced Grant 695671 “QUENCH” and support by the Science and Technology Facilities Council (STFC). GL acknowledges support from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (project CONCERTO, grant agreement No 788212) and from the Excellence Initiative of Aix-Marseille University-A\*Midex, a French “Investissements d’Avenir” programme.
- ## References
- Aravena, M., Decarli, R., González-López, J., et al. 2019, *ApJ*, 882, 136  
 Ashby, M. L. N., Willner, S. P., Fazio, G. G., et al. 2013, *ApJ*, 769, 80  
 Béthermin, M., Fudamoto, Y., Ginolfi, M., et al. 2020, *arXiv e-prints*, arXiv:2002.00962  
 Bouwens, R. J., Aravena, M., Decarli, R., et al. 2016, *ApJ*, 833, 72  
 Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2012a, *ApJ*, 754, 83  
 Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2015, *ApJ*, 803, 34  
 Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2012b, *ApJ*, 752, L5  
 Bowler, R. A. A., Jarvis, M. J., Dunlop, J. S., et al. 2020, *MNRAS*, 493, 2059  
 Brammer, G. B., van Dokkum, P. G., Franx, M., et al. 2012, *ApJS*, 200, 13  
 Capak, P., Aussel, H., Bundy, K., et al. 2012, SPLASH: Spitzer Large Area Survey with Hyper-Suprime-Cam, Spitzer Proposal  
 Capak, P., Carilli, C. L., Lee, N., et al. 2008, *ApJ*, 681, L53  
 Capak, P. L., Carilli, C., Jones, G., et al. 2015, *Nature*, 522, 455  
 Capak, P. L., Riechers, D., Scoville, N. Z., et al. 2011, *Nature*, 470, 233  
 Cardamone, C. N., van Dokkum, P. G., Urry, C. M., et al. 2010, *ApJS*, 189, 270  
 Carniani, S., Maiolino, R., Amorin, R., et al. 2018, *MNRAS*, 478, 1170  
 Chabrier, G. 2003, *PASP*, 115, 763  
 Cooke, E. A., Smail, I., Swinbank, A. M., et al. 2018, *ApJ*, 861, 100  
 Cormier, D., Leboutteiller, V., Madden, S. C., et al. 2012, *A&A*, 548, A20  
 Cucciati, O., Tresse, L., Ilbert, O., et al. 2012, *A&A*, 539, A31  
 Daddi, E., Dannerbauer, H., Liu, D., et al. 2015, *A&A*, 577, A46  
 Dahlen, T., Mobasher, B., Dickinson, M., et al. 2007, *ApJ*, 654, 172  
 De Looze, I., Cormier, D., Leboutteiller, V., et al. 2014, *A&A*, 568, A62  
 Decarli, R., Walter, F., Aravena, M., et al. 2016, *ApJ*, 833, 69  
 Decarli, R., Walter, F., González-López, J., et al. 2019, *ApJ*, 882, 138  
 Díaz-Santos, T., Armus, L., Charmandaris, V., et al. 2013, *ApJ*, 774, 68  
 Dickinson, M., Giavalisco, M., & GOODS Team. 2003, in *The Mass of Galaxies at Low and High Redshift*, ed. R. Bender & A. Renzini, 324  
 Driver, S. P., Andrews, S. K., da Cunha, E., et al. 2018, *MNRAS*, 475, 2891  
 Dunlop, J. S., McLure, R. J., Biggs, A. D., et al. 2017, *MNRAS*, 466, 861  
 Elbaz, D., Dickinson, M., Hwang, H. S., et al. 2011, *A&A*, 533, A119  
 Faisst, A. L., Capak, P. L., Yan, L., et al. 2017, *ApJ*, 847, 21  
 Faisst, A. L., Schaerer, D., Lemaux, B. C., et al. 2019, *arXiv e-prints*, arXiv:1912.01621  
 Finkelstein, S. L. 2016, *PASA*, 33, e037  
 Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *PASP*, 125, 306  
 Fudamoto, Y., Oesch, P. A., Faisst, A., et al. 2020, *arXiv e-prints*, arXiv:2004.10760  
 Fujimoto, S., Ouchi, M., Ferrara, A., et al. 2019, *ApJ*, 887, 107  
 Fujimoto, S., Ouchi, M., Shibuya, T., & Nagai, H. 2017, *ApJ*, 850, 83  
 Fujimoto, S., Silverman, J. D., Béthermin, M., et al. 2020, *arXiv e-prints*, arXiv:2003.00013  
 Geach, J. E., Dunlop, J. S., Halpern, M., et al. 2017, *MNRAS*, 465, 1789  
 Gehrels, N. 1986, *ApJ*, 303, 336  
 Giavalisco, M., Ferguson, H. C., Koekemoer, A. M., et al. 2004, *ApJ*, 600, L93  
 Ginolfi, M., Jones, G. C., Béthermin, M., et al. 2020a, *arXiv e-prints*, arXiv:2004.13737  
 Ginolfi, M., Jones, G. C., Béthermin, M., et al. 2020b, *A&A*, 633, A90  
 González-López, J., Decarli, R., Pavesi, R., et al. 2019, *ApJ*, 882, 139  
 Grupponi, C., Calura, F., Pozzi, F., et al. 2015, *MNRAS*, 451, 3419  
 Grupponi, C. & Pozzi, F. 2019, *MNRAS*, 483, 1993  
 Grupponi, C., Pozzi, F., Rodighiero, G., et al. 2013, *MNRAS*, 432, 23  
 Guo, Y., Ferguson, H. C., Giavalisco, M., et al. 2013, *ApJS*, 207, 24  
 Hawkins, E., Maddox, S., Branchini, E., & Saunders, W. 2001, *MNRAS*, 325, 589  
 Hemmati, S., Yan, L., Diaz-Santos, T., et al. 2017, *ApJ*, 834, 36  
 Hennawi, J. F., Myers, A. D., Shen, Y., et al. 2010, *ApJ*, 719, 1672  
 Hollenbach, D. J. & Tielens, A. G. G. M. 1999, *Reviews of Modern Physics*, 71, 173  
 Jones, G. C., Béthermin, M., Fudamoto, Y., et al. 2020, *MNRAS*, 491, L18  
 Jones, G. C., Carilli, C. L., Shao, Y., et al. 2017, *ApJ*, 850, 180  
 Karim, A., Swinbank, A. M., Hodges, J. A., et al. 2013, *MNRAS*, 432, 2  
 Kennicutt, Robert C., J. 1998, *ApJ*, 498, 541  
 Kistler, M. D., Yüksel, H., Beacom, J. F., Hopkins, A. M., & Wyithe, J. S. B. 2009, *ApJ*, 705, L104  
 Koekemoer, A. M., Aussel, H., Calzetti, D., et al. 2007, *ApJS*, 172, 196  
 Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, *ApJS*, 197, 36  
 Kohandel, M., Pallottini, A., Ferrara, A., et al. 2019, *MNRAS*, 487, 3007  
 Kroupa, P. 2001, *MNRAS*, 322, 231  
 Lagache, G., Cousin, M., & Chatzikos, M. 2018, *A&A*, 609, A130  
 Laigle, C., McCracken, H. J., Ilbert, O., et al. 2016, *ApJS*, 224, 24  
 Le Fèvre, O., Béthermin, M., Faisst, A., et al. 2019, *arXiv e-prints*, arXiv:1910.09517  
 Le Floch, E., Aussel, H., Ilbert, O., et al. 2009, *ApJ*, 703, 222  
 Lemaux, B. C., Le Fèvre, O., Cucciati, O., et al. 2018, *A&A*, 615, A77  
 Liu, D., Schinnerer, E., Groves, B., et al. 2019, *ApJ*, 887, 235  
 Lutz, D., Poglitsch, A., Altieri, B., et al. 2011, *A&A*, 532, A90  
 Madau, P. & Dickinson, M. 2014, *ARA&A*, 52, 415  
 Magdis, G. E., Rigopoulou, D., Hopwood, R., et al. 2014, *ApJ*, 796, 63  
 Magnelli, B., Elbaz, D., Chary, R. R., et al. 2011, *A&A*, 528, A35  
 Magnelli, B., Popesso, P., Berta, S., et al. 2013, *A&A*, 553, A132  
 Maniyar, A. S., Béthermin, M., & Lagache, G. 2018, *A&A*, 614, A39  
 Marchesi, S., Civano, F., Elvis, M., et al. 2016, *ApJ*, 817, 34  
 Marshall, H. L., Tananbaum, H., Zamorani, G., et al. 1983, *ApJ*, 269, 42  
 Matthee, J., Sobral, D., Boogaard, L. A., et al. 2019, *ApJ*, 881, 124  
 Novak, M., Bañados, E., Decarli, R., et al. 2019, *ApJ*, 881, 63  
 Novak, M., Smolčić, V., Delhaize, J., et al. 2017, *A&A*, 602, A5  
 Oesch, P. A., Bouwens, R. J., Illingworth, G. D., Labbé, I., & Stefanon, M. 2018, *ApJ*, 855, 105  
 Pallottini, A., Ferrara, A., Decataldo, D., et al. 2019, *MNRAS*, 487, 1689  
 Pavesi, R., Riechers, D. A., Capak, P. L., et al. 2016, *ApJ*, 832, 151  
 Pavesi, R., Riechers, D. A., Sharon, C. E., et al. 2018, *ApJ*, 861, 43  
 Popping, G., Narayanan, D., Somerville, R. S., Faisst, A. L., & Krumholz, M. R. 2019, *MNRAS*, 482, 4906  
 Reddy, N. A. & Steidel, C. C. 2009, *ApJ*, 692, 778  
 Riechers, D. A., Bradford, C. M., Clements, D. L., et al. 2013, *Nature*, 496, 329  
 Riechers, D. A., Capak, P. L., Carilli, C. L., et al. 2010, *ApJ*, 720, L131  
 Riechers, D. A., Carilli, C. L., Capak, P. L., et al. 2014, *ApJ*, 796, 84  
 Riechers, D. A., Pavesi, R., Sharon, C. E., et al. 2019, *ApJ*, 872, 7  
 Robotham, A. S. G. & Driver, S. P. 2011, *MNRAS*, 413, 2570  
 Rodighiero, G., Daddi, E., Baronchelli, I., et al. 2011, *ApJ*, 739, L40  
 Romano, M., Cassata, P., Morselli, L., et al. 2020, *arXiv e-prints*, arXiv:2002.00961  
 Rowan-Robinson, M., Oliver, S., Wang, L., et al. 2016, *MNRAS*, 461, 1100  
 Sanders, D. B., Mazzarella, J. M., Kim, D. C., Surace, J. A., & Soifer, B. T. 2003, *AJ*, 126, 1607  
 Sanders, D. B., Salvato, M., Aussel, H., et al. 2007, *ApJS*, 172, 86  
 Schaerer, D., Ginolfi, M., Béthermin, M., et al. 2020, *arXiv e-prints*, arXiv:2002.00979  
 Schechter, P. 1976, *ApJ*, 203, 297  
 Schenker, M. A., Robertson, B. E., Ellis, R. S., et al. 2013, *ApJ*, 768, 196  
 Schiminovich, D., Ilbert, O., Arnouts, S., et al. 2005, *ApJ*, 619, L47  
 Schinnerer, E., Carilli, C. L., Capak, P., et al. 2008, *ApJ*, 689, L5  
 Scoville, N., Aussel, H., Brusa, M., et al. 2007, *ApJS*, 172, 1  
 Scoville, N., Lee, N., Vanden Bout, P., et al. 2017, *ApJ*, 837, 150  
 Smit, R., Bouwens, R. J., Carniani, S., et al. 2018, *Nature*, 553, 178  
 Smolčić, V., Novak, M., Bondi, M., et al. 2017, *A&A*, 602, A1  
 Solomon, P. M., Downes, D., & Radford, S. J. E. 1992, *ApJ*, 398, L29  
 Speagle, J. S., Steinhardt, C. L., Capak, P. L., & Silverman, J. D. 2014, *ApJS*, 214, 15  
 Swinbank, A. M., Karim, A., Smail, I., et al. 2012, *MNRAS*, 427, 1066  
 Tacconi, L. J., Genzel, R., Saintonge, A., et al. 2018, *ApJ*, 853, 179  
 Takeuchi, T. T., Yoshikawa, K., & Ishii, T. T. 2003, *ApJ*, 587, L89  
 Vallini, L., Gallerani, S., Ferrara, A., Pallottini, A., & Yue, B. 2015, *ApJ*, 813, 36  
 Vallini, L., Grupponi, C., Pozzi, F., Vignali, C., & Zamorani, G. 2016, *MNRAS*, 456, L40  
 Walter, F., Decarli, R., Aravena, M., et al. 2016, *ApJ*, 833, 67  
 Walter, F., Decarli, R., Carilli, C., et al. 2012, *Nature*, 486, 233  
 Weiß, A., Kovács, A., Coppin, K., et al. 2009, *ApJ*, 707, 1201  
 Wolfire, M. G., McKee, C. F., Hollenbach, D., & Tielens, A. G. G. M. 2003, *ApJ*, 587, 278  
 Wyder, T. K., Treyer, M. A., Milliard, B., et al. 2005, *ApJ*, 619, L15  
 Yamaguchi, Y., Kohno, K., Tamura, Y., et al. 2017, *ApJ*, 845, 108

- 
- <sup>1</sup> Università di Bologna, Dipartimento di Fisica e Astronomia (DIFA), via Gobetti 93/2, I-40129 Bologna, Italy
- <sup>2</sup> INAF – Osservatorio di Astrofisica e Scienza dello Spazio, via Gobetti 93/3, I-40129, Bologna, Italy
- <sup>3</sup> INAF – Osservatorio Astrofisico di Arcetri, Largo E. Fermi 5, I-50125, Firenze, Italy
- <sup>4</sup> The Caltech Optical Observatories, California Institute of Technology, Pasadena, CA 91125, USA
- <sup>5</sup> Department of Astronomy, Cornell University, Space Sciences Building, Ithaca, NY 14853, USA
- <sup>6</sup> Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany
- <sup>7</sup> IPAC, California Institute of Technology, 1200 East CaliforniaBoulevard, Pasadena, CA 91125, USA
- <sup>8</sup> The Cosmic Dawn Center, University of Copenhagen, Vibenshuset, Lyngbyvej 2, DK-2100 Copenhagen, Denmark
- <sup>9</sup> Niels Bohr Institute, University of Copenhagen, Lyngbyvej 2, DK-2100 Copenhagen, Denmark
- <sup>10</sup> Cavendish Laboratory, University of Cambridge, 19 J. J. Thomson Ave., Cambridge CB3 0HE, UK
- <sup>11</sup> Kavli Institute for Cosmology, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK
- <sup>12</sup> Space Telescope Science Institute, 3700 San Martin Dr., Baltimore, MD 21218, USA
- <sup>13</sup> Leiden Observatory, Leiden University, PO Box 9500, 2300 RA Leiden, The Netherlands
- <sup>14</sup> Dipartimento di Fisica e Astronomia, Università di Padova, Vicolo dell'Osservatorio 3, I-35122, Padova, Italy
- <sup>15</sup> Observatoire de Genève, Université de Genève, 51 Ch. des Maillettes, 1290 Versoix, Switzerland
- <sup>16</sup> INAF, Istituto di Radioastronomia, via Piero Gobetti 101, I-40129 Bologna, Italy
- <sup>17</sup> Centro de Astronomia (CITEVA), Universidad de Antofagasta, Avenida Angamos 601, Antofagasta, Chile
- <sup>18</sup> Aix Marseille Univ, CNRS, CNES, LAM, Marseille, France
- <sup>19</sup> Department of Physics, University of California, Davis, One Shields Ave., Davis, CA 95616, USA
- <sup>20</sup> INAF, Osservatorio Astronomico di Padova, vicolo dell'Osservatorio 5, I-35122 Padova, Italy
- <sup>21</sup> Kavli Institute for the Physics and Mathematics of the Universe, The University of Tokyo Kashiwa, Chiba 277-8583, Japan
- <sup>22</sup> Department of Astronomy, School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo, Tokyo 113-0033, Japan

Table C.1: Luminosity functions for the clustered and field sample considering the 21 sources with fidelity higher than 20%. We reported also the number of sources in each luminosity bin. We indicated with the bold font the values corresponding to the independent bins.

$\log(L/L_{\odot})$	$\log(\Phi_{\text{clust}}/\text{Mpc}^{-3}\text{dex}^{-1})$	$N_{\text{clust}}$	$\log(\Phi_{\text{field}}/\text{Mpc}^{-3}\text{dex}^{-1})$	$N_{\text{field}}$
8.25	<b>-2.87</b> <sup>+0.29</sup> <sub>-0.34</sub>	<b>3</b>	<b>-3.30</b> <sup>+0.36</sup> <sub>-0.45</sub>	<b>2</b>
8.50	-2.81 <sup>+0.25</sup> <sub>-0.28</sub>	4	-3.05 <sup>+0.25</sup> <sub>-0.28</sub>	4
8.75	<b>-2.64</b> <sup>+0.20</sup> <sub>-0.22</sub>	<b>6</b>	<b>-3.40</b> <sup>+0.36</sup> <sub>-0.45</sub>	<b>2</b>
9.00	-2.62 <sup>+0.17</sup> <sub>-0.18</sub>	8	-3.37 <sup>+0.52</sup> <sub>-0.77</sub>	1
9.25	<b>-2.96</b> <sup>+0.25</sup> <sub>-0.28</sub>	<b>4</b>	<b>-3.37</b> <sup>+0.52</sup> <sub>-0.77</sub>	<b>1</b>
9.50	-3.40 <sup>+0.52</sup> <sub>-0.77</sub>	1	...	...
9.75	<b>-3.10</b> <sup>+0.36</sup> <sub>-0.45</sub>	<b>2</b>	...	...
10.00	-3.40 <sup>+0.52</sup> <sub>-0.77</sub>	1	...	...

Table C.2: Luminosity functions for the clustered sample after having excluded the six sources interacting with the central targets. We reported also the number of sources in each luminosity bin. We indicated with the bold font the values corresponding to the independent bins.

$\log(L/L_{\odot})$	$\log(\Phi_{\text{clust}}/\text{Mpc}^{-3}\text{dex}^{-1})$	$N_{\text{clust}}$
8.25	...	...
8.50	-3.27 <sup>+0.52</sup> <sub>-0.77</sub>	1
8.75	<b>-2.88</b> <sup>+0.29</sup> <sub>-0.34</sub>	<b>3</b>
9.00	-2.92 <sup>+0.29</sup> <sub>-0.34</sub>	3
9.25	<b>-3.39</b> <sup>+0.52</sup> <sub>-0.77</sub>	<b>1</b>
9.50	...	...
9.75	<b>-3.40</b> <sup>+0.52</sup> <sub>-0.77</sub>	<b>1</b>
10.00	-3.40 <sup>+0.52</sup> <sub>-0.77</sub>	1

## Appendix A: Images and spectra

## Appendix B: Completeness

We show the completeness curves as a function of the flux peak for fixed FWHM in Appendix B. We can see that at fixed flux peak the completeness is higher for larger lines.

## Appendix C: Luminosity function

We show in this section the [C II] luminosity function (LF) computed including sources with a fidelity >20% (Figure C.1 and Table C.1). In Sect. 3.2 we included in our catalog only the detections with fidelity > 85%, with only one [C II] emitter (S5100969402) having a fidelity ~ 50%, in order to study a very robust sample. However this sample is obviously incomplete at the lowest luminosities. We thus derived the clustered and field LFs for two new sub-samples, in which we included also low fidelity (i.e. low luminosity) sources. The fidelity cut of 20% adds 10 lines to our catalog of serendipitous detection. Based on their individual fidelities we expect that only ~ 3 out of the 9 lines are genuine detections, thus this new sample is expected to contain a high fraction of spurious sources. One line is possibly associated to CO emission at  $z < 5$  (see Sect. 4). The remaining nine do not present any optical/NIR counterpart hence their redshift is unconstrained from ancillary data. Therefore, for these sources we can only assume that their emission is due to the [C II] line. We see that the shape of both clustered/field LFs remains quite unchanged, with the field LF sampled by more sources now. The field sources lie at lower luminosities compared to the clustered sample. However, this is not surprising since, as the fidelity lowers, also the lines flux decreases and hence we expect the population of the low luminosity bins only for both field and clustered sources.

We show also the [C II] LF (clustered sample) obtained excluding the six [C II] possibly associated to mergers/interactions (Figure C.2 and Table C.2). The new points are consistent within the errors with the old ones.

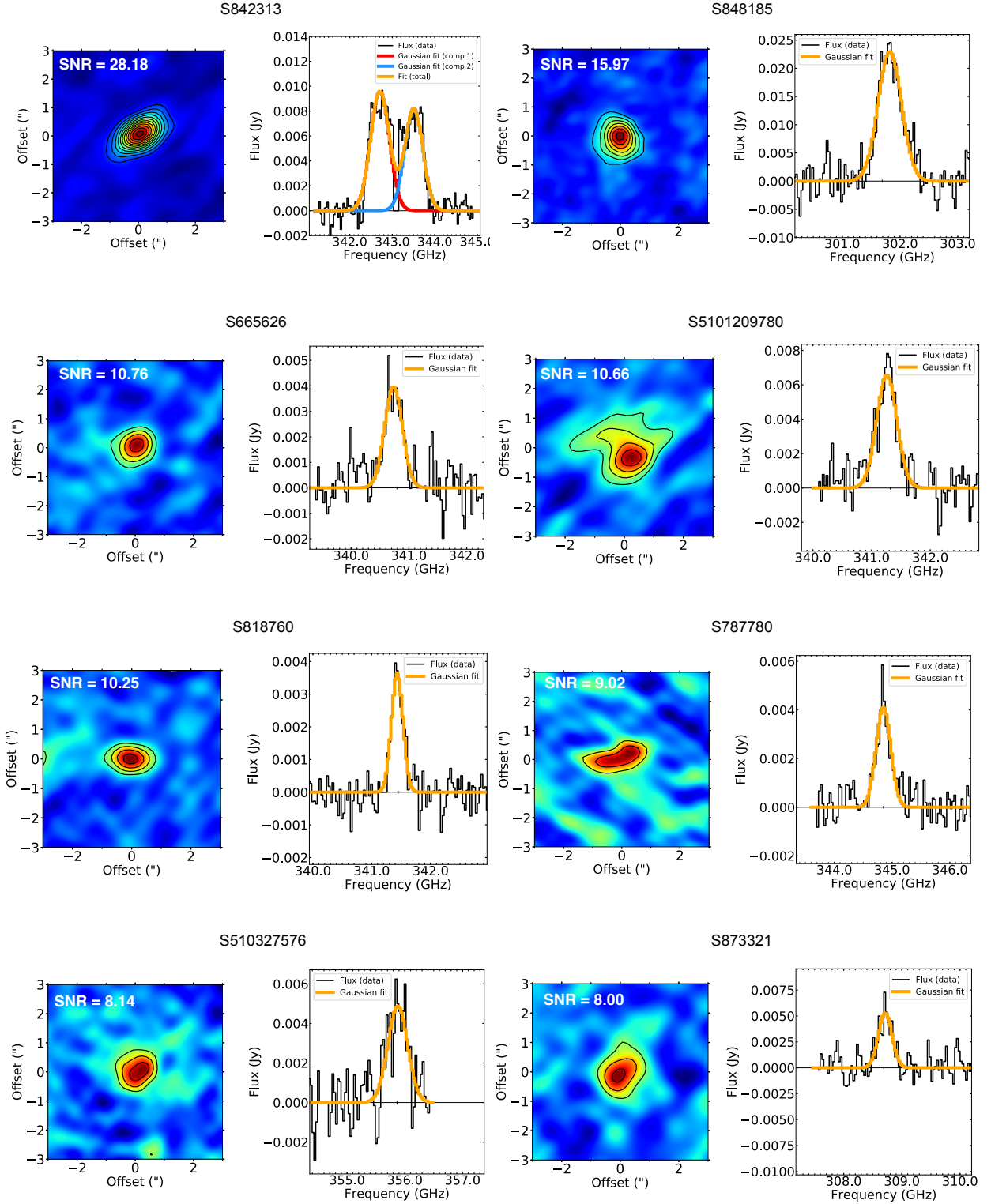


Fig. A.1: Continuum-subtracted ALMA maps and spectra of the 14 serendipitous lines found in the 118 ALPINE pointings. Each panel is labelled according to the number of the ALPINE source in the same pointing. The lowest contour level corresponds to  $3\sigma$ . We fitted the line emission (black) using a single Gaussian component (orange). In case of S842313 we fitted the line profile using two Gaussian components (cyan and red); the total model is shown in orange. For S5100822662 the serendipitous source is the small blob above the ALPINE. The blob is marked with a cross and the spectrum shows its emission.



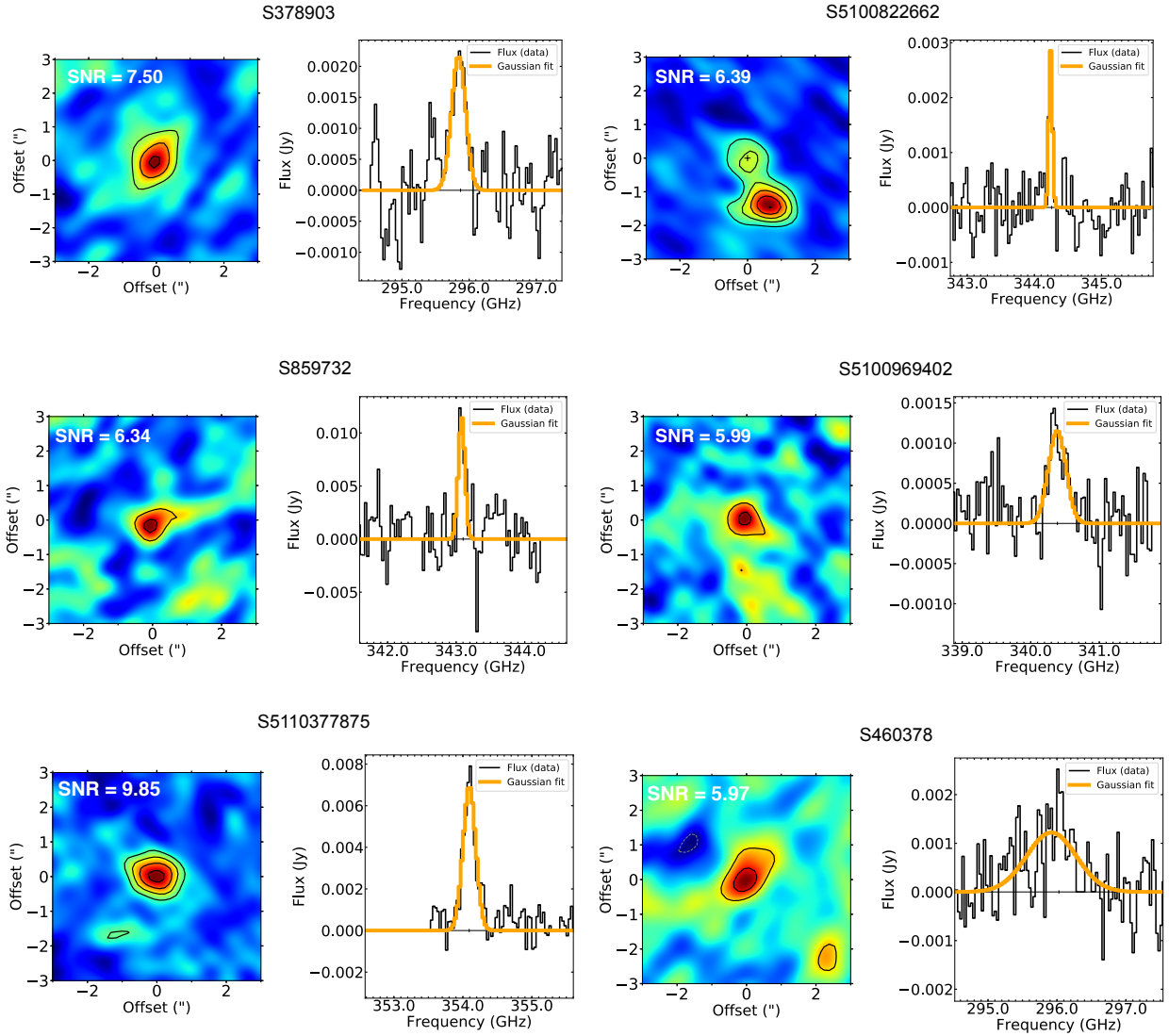


Fig. A.1: (continued) Continuum-subtracted ALMA maps and spectra of the 14 serendipitous lines found in the 118 ALPINE pointings.

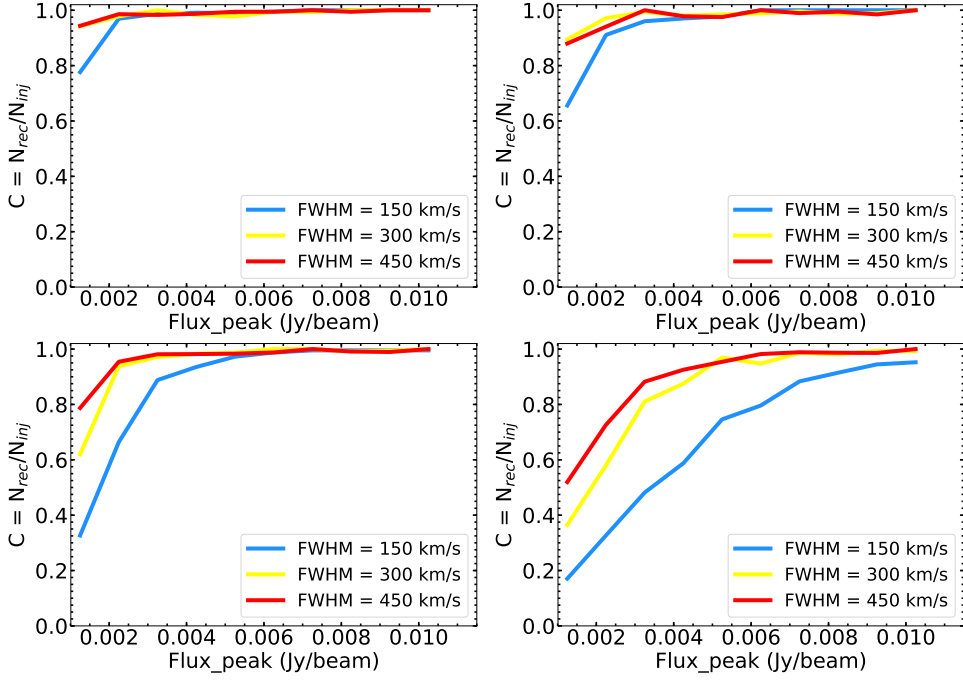


Fig. B.1: Completeness curves for fixed FWHM in four regions with radii  $R_{<30}$  (distance from the phase center  $R_{<30} \leq 7.1''$ ),  $R_{30-50}$  ( $7.1'' < R_{30-50} \leq 10.4''$ ),  $R_{50-70}$  ( $10.4'' < R_{50-70} \leq 13.1''$ ),  $R_{70-90}$  ( $13.1'' < R_{70-90} \leq 16.4''$ ), defined on the basis of their primary beam attenuation. The curves correspond to 150, 300, 450  $\text{km s}^{-1}$  (reported in blue, yellow and red respectively). We can see that completeness is a strong function of the line location and width.

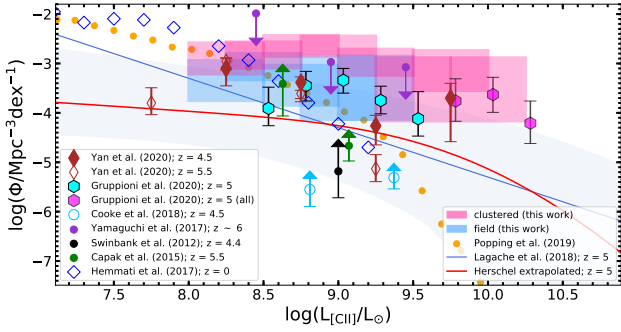


Fig. C.1: [C II] luminosity functions (LFs) at  $z \sim 5$  of the serendipitous sources in ALPINE including the sources with low fidelity. Compared to the clustered LF, the field one lies at lower luminosity. Both LFs result compatible with those based on the sources with higher fidelity. The main difference is at the low luminosity end since more sources are included in this sample.

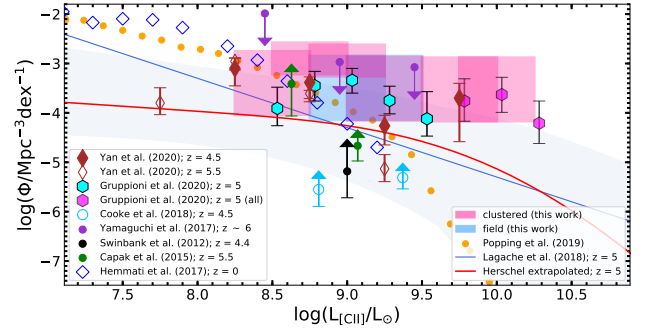


Fig. C.2: [C II] luminosity function (LF) at  $z \sim 5$  of the clustered sources, after removing the six [C II]-emitters associated to interacting galaxies with the central targets (see Figure 4). Compared to the LF presented in Sect. 5, this one results lower at the faint end.