

A deep [CII] survey at $z = 6.6 - 7.2$: evidence for overdensity around a luminous Ly α emitter

Laurence Day¹[★], David Sobral^{1,2}, Jorryt Matthee², Emma Dodd¹, Lauren Draper^{1,3}

¹ Department of Physics, Lancaster University, Lancaster, LA1 4YB, UK

² Leiden Observatory, Leiden University, P.O. Box 9513, NL-2300 RA Leiden, The Netherlands

³ School of Physics and Astronomy, University of Manchester, Oxford Road, Manchester M13 9PL

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

Luminous Ly α emitters observed at $z > 6$ likely reside in early ionized bubbles, with sufficient ionizing flux produced within their environment to rapidly form and maintain the ionization state of the surrounding intergalactic medium (IGM). Due to the observational challenges encountered when studying the epoch of reionization, the nature of the environments most conducive to rapid reionization remain unclear. COSMOS Redshift 7 (CR7; Sobral et al. 2015) is the brightest Ly α emitter within the epoch of reionisation at $z = 6.6$, and has deep, high resolution ALMA data available ($\sigma = 5 \text{ mJy beam}^{-1} \text{ km s}^{-1}$). Here we explore deep [CII] observations of CR7 to search for other [CII] emitters around and in the line of sight within a region of $51.2'' \times 51.2''$ ($0.277\text{Mpc} \times 0.277\text{Mpc}$ proper distance) and spanning frequency range 249 - 251 GHz ($z \sim 6.57 - 6.62$ in [CII]). We undertake a blind survey of the immediate surroundings of CR7, as well as three volumes beyond CR7 ($z \sim 6.6 - 7.2$), utilizing the [CII] emission line which bypasses the attenuation due to neutral intergalactic hydrogen that is characteristic of typical observations. We identify 12 candidates for [CII] emitters from analysis of line spectra, and we perform statistical analysis to reveal the overdensity in which CR7 likely resides, obtaining CR7 to be overdense by a factor of $3.95^{+0.91}_{-0.76}$. We then place observational constraints on the faint end of the high- z [CII] luminosity function, and compare with predictions. The results of our analysis suggest that over-dense environments are potentially conducive to the rapid reionization of the surrounding IGM for $z > 6$.

Key words: keyword1 – keyword2 – keyword3

1 INTRODUCTION

The first generations of galaxies to form in the early Universe likely triggered the phase transition of intergalactic hydrogen that characterises the end of the cosmic ‘dark ages’, in the process of reionization (i.e. Robertson et al. 2015). Ionizing radiation produced by the first generation of stars that form from the pristine intergalactic gas, followed by metal enriched Pop II stars and the first active galactic nuclei (AGN) as these galaxies evolved, begins to reionize the intergalactic hydrogen immediately surrounding the host galaxy (e.g. Zahn et al. 2007). The process of reionization is expected to take place with the growth of a series of ionized ‘bubbles’ distributed throughout the neutral Intergalactic Medium (IGM hereafter), continuing until the bubbles overlap and the reionization of the Universe is complete in a process known as patchy reionization (e.g. Natarajan et al. 2013; Furlanetto & Oh 2005; Alvarez & Abel 2012; Sobacchi & Mesinger 2015). Galaxies are considered to be likely the predominant source of ionizing flux

(e.g. Robertson et al. 2015; Faisst 2016) that drove the reionization of the IGM with an observed high number density compared to AGN, which are expected to play a more minor role (e.g. Onoue et al. 2017; Parsa et al. 2018).

There are a variety of observational consequences associated with the presence of intervening neutral gas in the IGM. Photons with $E_\gamma > 13.6 \text{ eV}$ (Lyman Continuum radiation; LyC) emitted from the first galaxies are absorbed as they ionize the surrounding gas (e.g. Inoue et al. 2014), due to the large cross-section and high fraction of neutral hydrogen (HI) in the IGM. This negates detection of Lyman continuum radiation. Furthermore, the Ly α emission line, one of the brightest emission lines in star forming galaxies, undergoes resonant scattering processes when encountering intervening neutral hydrogen (e.g. Dijkstra et al. 2014). This causes significant deviation in the path of a photon from its point of origin, reducing brightness and hence its detectability in the line of sight.

The study of high redshift galaxy spectra can uncover clues into the neutrality of the intervening IGM with the study of the

* E-mail: l.day1@lancaster.ac.uk

$\text{Ly}\alpha$ forest (e.g. Inoue et al. 2014), and at higher redshift, the Gunn-Peterson trough (e.g. Gunn & Peterson 1965). Studies of the Gunn-Peterson trough of quasars have enabled the determination of the epoch at which reionization is almost complete, as well as provide evidence for a patchy progression of reionization. At present, observational constraints place the end of the epoch of reionization at $z \sim 6$, supported by studies of high redshift quasars (e.g. Becker et al. 2001; McGreer et al. 2015; Mortlock et al. 2011; Fan et al. 2006), and of the decline in $\text{Ly}\alpha$ fraction of LBGs at $z \geq 6$ (e.g. Ono et al. 2012; Caruana et al. 2014; Tilvi et al. 2014; Mason et al. 2018). Less well understood, however, is the onset time and the duration of the epoch of reionization. The most stringent constraints at present have been provided by study of the integrated optical depth of Thomson scattering, τ , to the cosmic microwave background (e.g. Bouwens et al. 2015b; Robertson et al. 2015) using results obtained by Planck Collaboration et al. (2016). Comparison of the $\text{Ly}\alpha$ forest in a number of quasar spectra at high redshift has provided evidence for variance in the HI fraction in the IGM, supporting the notion of a patchy progression of reionization (e.g. Becker et al. 2004). Further evidence for patchy reionization obtained from the study of the spectra of high redshift galaxies is outlined in Becker et al. (2015), in which a significant trough in the $\text{Ly}\alpha$ forest at $z \approx 5.6 - 5.8$ signifies a fraction of neutral hydrogen significantly greater than other lines of sight at the same redshift, which is significant enough that it can not be explained simply by density fluctuations alone. This requires fluctuations in neutral hydrogen fraction that can be explained by fluctuations in the ionizing UV radiation field. This is substantiated by Bosman et al. (2018), who find that density fluctuations are insufficient to fully explain the scatter in IGM opacity.

High redshift galaxy surveys in recent years, have enabled the UV galaxy luminosity function to be determined to redshifts as high as $z \sim 10$ (e.g. Bouwens et al. 2015a; Finkelstein 2016; Atek et al. 2018), which provide an insight into the galactic population into the epoch of reionization, and its evolution with time. The relative contribution to ionizing flux from the faint or bright galactic population, however, remains unclear. Faint, low mass galaxies exhibit high specific star formation rates, and the escape of ionizing flux is likely being facilitated by redistribution of neutral hydrogen driven by feedback processes (e.g. Erb 2015; Livermore et al. 2017; Sobral et al. 2017a). Other recent studies (i.e. Sharma et al. 2017), however, suggest that the faint end of the galactic population plays a more minor role compared to the brightest galaxies at high redshift, arguing that escape fraction of ionizing photons is higher for galaxies with higher $\text{Ly}\alpha$ equivalent widths. (e.g. Steidel et al. 2018; Fletcher et al. 2018; Sobral & Matthee 2018). Furthermore, recent analysis of simulations presented in Katz et al. (2018) suggests that as the epoch of reionization progresses, contributions to ionizing flux from faint galaxies dominated in the earliest epochs, but by $z \sim 6$, massive galaxies with evolved stellar populations may dominate, suggesting a complex interplay between both faint and bright galaxies contributing to cosmic reionization.

As well as studying the relative contribution of sources to cosmic reionization, another topic of interest, and one that is addressed in this paper concerns the topology of reionization. In particular, is an underlying overdensity of galaxies conducive to the rapid formation of an ionized bubble? This question was addressed observationally for the first time in Castellano et al. (2016), with follow up in Castellano et al. (2018), in which a search is conducted for companions to a pair of $z \sim 7$ LAEs which are alone

insufficiently luminous to ionize the surrounding region. The latest analysis confirms the existence of a close third LAE counterpart, strengthening the case for a galactic overdensity. In an ‘overdense’ region of space with an excess of sources when compared with the wider field, there is significantly more ionizing flux that contributes to the formation and expansion of the ionized ‘bubble’ surrounding the galaxy. This is not the complete picture, however, as overdense regions likely contain a greater quantity of neutral intergalactic matter to ionize in the first place, and hence recombination rates would be higher. Conversely, individual galaxies residing in an inherently less dense region, despite producing less ionizing flux than a group of galaxies, there is likely less intergalactic matter to ionize and thus the ionized bubble could more readily form (i.e. Choudhury et al. 2009). The first direct evidence of a galaxy directly contributing to ionizing the surrounding IGM and forming an ionized bubble is presented in Matthee et al. (2018), in which double-peaked $\text{Ly}\alpha$ emission at $z=6.593$ (Hu et al. 2016) is detected. Further searches for double peaks in the $\text{Ly}\alpha$ line at high redshift show promise in unveiling the prevalence of ionized regions surrounding galaxies in the EoR (e.g. Songaila et al. 2018).

The detection of $\text{Ly}\alpha$ photons requires a sufficiently low HI fraction in the line of sight, requiring sufficient ionizing flux to have been produced from the sources and their surrounding environment. Hence, strong emission in $\text{Ly}\alpha$ detected beyond $z \sim 6$ implies the presence of an early ionized bubble (e.g. Stark et al. 2017). Thus, LAE’s within the epoch of reionization may be candidates of over-dense regions. Furthermore, the most luminous $\text{Ly}\alpha$ emitters at high-z are typically comprised of multiple distinct components in the rest frame UV, indicative of recent or ongoing mergers (e.g. Matthee et al. 2017a). Merger events suggest a likely overdensity of sources alone and signifies the production of large quantities of ionizing photons.

In the far infra-red (FIR), there are a series of bright emission lines that can be observed that do not suffer from attenuation due to a neutral IGM or dust. This can enable the detection of a population of galaxies that are not detected when using typical bright (e.g. UV) emission. Modern ground-based facilities such as the Atacama Large Millimetre/Submillimetre Array (ALMA) enable the study of FIR emission originating in the early Universe with high resolution, enabling sensitive detection of faint sources at high redshift (e.g. Matthee et al. 2017b; Aravena et al. 2016; Pentericci et al. 2016; Maiolino et al. 2015; Watson et al. 2015). This enables facilities such as ALMA to peer into the epoch of reionization, and detect sources which are either too faint to detect with optical/UV observations, or have a significant fraction of intervening intergalactic neutral hydrogen such that they are not detectable in $\text{Ly}\alpha$. Particularly bright FIR emission lines such as the [CII] $158\mu\text{m}$ emission, a coolant of the interstellar medium (ISM), also provide a tracer of star formation (e.g. Vallini et al. 2015). Another prospect is the luminous [OIII] $88\mu\text{m}$ emission line, with its detection extending beyond $z>7$ (e.g. Inoue et al. 2016; Moriwaki et al. 2018), where it shifts into favourable ALMA bands.

Presented in this paper is a search for [CII] emitting sources in the region surrounding CR7 (Cosmos Redshift 7, Sobral et al. 2015). CR7 is the most luminous $\text{Ly}\alpha$ emitter at $z \sim 6.6$, with direct detection of the $\text{Ly}\alpha$ line suggesting it exists in a reionized bubble. CR7 has been studied extensively since its discovery, with much debate over its nature. We identify the underlying

Table 1. Information on the data cubes involved in this study, including the frequency range and [CII] redshift range contained within each cube.

Name	Frequency range (GHz)	[CII] redshift range
spw3	249.3 - 250.8	6.57-6.62
spw2	247.1 - 248.8	6.63-6.69
spw1	234.1 - 235.7	7.06-7.12
spw0	233.9 - 232.0	7.13-7.19

overdensity in which CR7 likely resides, and form a catalogue of our most robust candidates for [CII] emission. We also obtain observational constraints on the high- z [CII] luminosity function at the faint end for the very first time. We can then compare the results of the luminosity function with current observational constraints provided by Aravena et al. (2016) with a survey of the HUDF, which probes the bright end of the luminosity function, as well as with predictions provided by Popping et al. (2016) and Lagache et al. (2017). The predictions provided in Popping et al. (2016) also include predictions of the CO line population at lower redshift, providing the opportunity to consider the fraction of CO line contaminants which potentially contaminate our sample, as well as utilizing the wealth of data available within the COSMOS field to search for our most robust candidates of emission.

This paper is structured as follows; We describe the ALMA data cubes and reduction in section 2.1. We then present our method to identify blind [CII] emitters within the ALMA data in section 3, and outline the subsequent analysis in section 4. A discussion of the results and the wider implications is presented in section 6 and a conclusion follows in section 7. Throughout the paper, we assume a Λ CDM cosmology with $\Omega_M = 0.30$, $\Omega_\Lambda = 0.70$, $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$, and a scale of $5.409 \text{ kpc}''$.

2 DATA

2.1 CR7 data overview

ALMA surveyed CR7 and the surrounding region, observed for a total on source integration time of 6.0h on 22, 23, 24 May and 4, 8 November 2016, with the scientific goal of detecting the [CII] line and searching for FIR continuum emission from CR7 (Matthee et al. 2017b). ALMA band 6 was used during cycle 3 for the observations (program ID #2015.1.00122.S; PI: Sobral), with four spectral windows centered at 249.94592, 247.94636, 234.94917 and 232.9496 GHz and with a band width of 1875 MHz. The names of the cubes, with the corresponding frequency channels and the [CII] redshift ranges are outlined in Table 1. The frequency range of the cubes, with the redshifts of expected emission lines including [CII] and CO emission are displayed in Figure 1. The cube spw3 is centered upon the [CII] emission from CR7. The other three cubes placed to obtain constraints on the IR continuum emission, which we utilize as comparison volumes at different [CII] redshifts.

CASA version 4.7.0 (McMullin et al. 2007) has been utilized to reduce the data, with natural weighting and channel averaging resulting in a velocity resolution of 38.8 km s^{-1} . The optimal S/N is obtained by using uv tapering of the visibilities with a Gaussian width FWHM = 0.7''. Resulting from the observational set up and the subsequent reduction of the data, we have a beam size of $0.35 \times 0.33''$ (gaussian σ ; $0.82 \times 0.77''$ FWHM) at a position angle

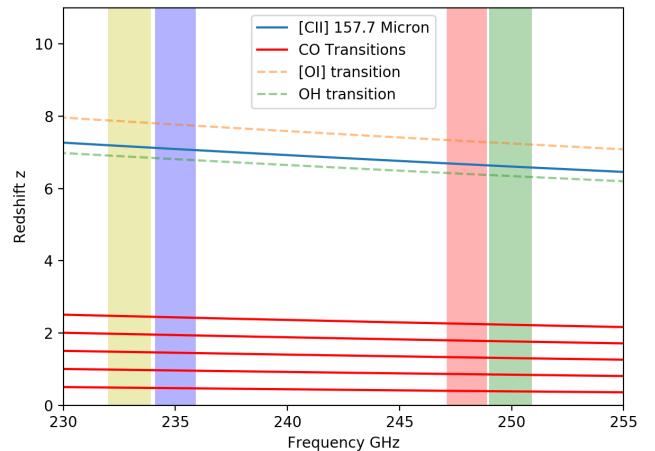


Figure 1. The redshift coverage within the four spectral windows in ALMA band 6, with the associated emission lines expected to be observed. [CII] is the most luminous in this range at high redshift, with potential for contamination from CO lines at lower redshift. Contamination due to [OI] and OH emission is possible however less likely, as the lines are inherently less luminous.

of -49.5° . Reduction was performed with uv tapering with different smoothing kernels in a range of $0.4\text{-}0.8''$, finding that variation in luminosity, line-width and size measurements remain consistent within the error-bars. An estimate of the noise level of the pixels within a $20''$ radius of the center of observation, and in channels of 38.8 km s^{-1} obtains $\sigma = 0.06 \text{ mJy beam}^{-1}$. We mapped the data on a scale of $0.05'' \text{ pix}^{-1}$.

2.2 Hubble Space Telescope Data

In our analysis, we utilize also three different data sets covering CR7 provided by the Hubble Space Telescope (*HST*) utilizing different filters. The F110W filter forms part of the Wide-Field Camera 3 (WFC3) infra-red filters, detecting wavelengths between $0.90 - 1.40 \mu\text{m}$ and includes the redshifted Ly α emission of CR7. The F160W filter operates between $1.45 - 1.75 \mu\text{m}$. Both the F110W and F160W data obtained with exposure time of 2.6ks (43 minutes) each, and are presented in the analysis of CR7 in Sobral et al. (2015). We also utilize F814W data from the COSMOS field *HST* Advanced Camera for Surveys (ACS) imaging, with wavelength range $0.73 - 0.95 \mu\text{m}$. We use these data sets to search for correspondence with the most likely candidates for genuine emission that we present in Section 4.1.1, to infer the likely nature of the candidate galaxies that we find.

2.3 COSMOS field catalogue

The Cosmic Evolution Survey (COSMOS; Koekemoer et al. 2007; Laigle et al. 2016) is a survey of a 2 square degree equatorial field, providing multi-wavelength imaging provided by major ground and space based telescopes. The COSMOS field catalogue presented in Laigle et al. (2016) contains over 2 million galaxies, and spans over 75% of cosmic history. In this study we use the COSMOS field catalogue to search for contamination within the catalogue of galaxies that we obtain in this study, from the blind survey of the

ALMA data. The COSMOS catalogue contains all of the known sources detected in NIR data within the scope of the COSMOS field, enabling the study of our most robust candidates.

3 METHOD

3.1 Visual examination of the CR7 data cube

3.1.1 Cube collapse and noise estimation

We begin by visually inspecting the data cube containing CR7 (spw3, 249.94592 GHz, Band width 1875 MHz). The first stage is to collapse the cube into adjacent slices of velocity width 74 km s⁻¹ (twice the channel resolution of the cube after reduction had taken place). We expect any real emission to exhibit widths >50km s⁻¹, with a narrower width obtaining a greater fraction of fake detections, and so we select this width so that detection of candidate sources is optimized. The cube is collapsed into 30 adjacent slices suitable for individual study. Following the collapse of the cube, the next stage is to estimate the noise level within each collapsed slice, and is obtained from the standard deviation of the distribution of fluxes through 50,000 randomly placed PSF (0.35") apertures, for each slice. Once the value of the standard deviation σ is obtained (typically ~ 0.05 mJy beam⁻¹), each slice is individually studied to detect candidates for emission. Following this procedure, a manual search for candidates takes place in each slice, and a catalogue of detections is obtained for each slice comprising the data cube.

3.1.2 Studying the negative cube

The entire procedure is then repeated using precisely the same technique for the flux reversed negative cube. The principle underpinning this technique is that any negative flux in the original cube corresponds to noise in the data, and hence enables us to obtain a measure of the number of false positives in our positive catalogue. It is the comparison between the positive and negative tables that enables determination of the fraction of most robust [CII] emitting candidates (see section 4). It is expected that the level of noise measured from the negative cube is matched in the positive cube, hence any offset in positives over negatives is potentially a signature of genuine emission existing within the cube.

Following this procedure, we obtain 1022 candidate sources in the positive mapping of the data cube, and 966 within the negative. This primitive result provides an initial insight into the population of sources that may reside surrounding CR7. This visual analysis is the foundation upon which we calibrate the automatic detection of galaxies with the automatic galaxy detection software in Section 3.2.

3.2 Developing the blind survey

Once the results are obtained from the visual inspection, we develop a blind survey technique that is calibrated with the produced catalogues. The software SExtractor (Bertin & Arnouts 1996) is utilized to replace the manual examination element of the procedure outlined in section 3.1. The capability of the software, including background determination and detection are utilized to handle the detection of candidate galaxies within each slice, upon which the photometry is performed. It is important to note that the

photometry features of SExtractor are not applied to the candidates, where we use alternative scripts instead.

When handling the detection of candidates within the slices, SExtractor operates on a series of parameters which affect the background determination and the detection thresholds. To optimize SExtractor, the parameters which the algorithm uses to handle the detection threshold of candidates, as well as determining the background within each slice, are altered over a range of values (See Appendix A for more detail on this procedure). For each parameter configuration, SExtractor is operated on a slice of which we have already visually studied, and have obtained a number of positive and negative candidates. The SExtractor obtained results are then compared with the candidates from the visual inspection. The parameter configuration which most closely matches the results of the visual inspection, we use for our blind survey.

Once SExtractor is optimized, a full blind cube analysis procedure is formulated which is reapplied to the CR7 data cube, and then performed upon the three other data cubes of CR7 to detect candidates for galaxies. In principle, the procedure can be readily applied to any other comparable data available on the ALMA archive. The script begins with the collapse of the cube into regions of 74km s⁻¹. This time, the cube not only collapses the adjacent slices, but also the intermediate slices (available due to the collapse width being twice the channel resolution). This maximises the detection potential of candidates with particularly narrow line widths which may remain undetected otherwise, as well as providing a means of detecting the extension of sources in velocity space. SExtractor with optimised operational parameters is then applied to each of the slices, producing a catalogue of coordinates of detections. These coordinates are then input into the photometry script outlined in the previous section, measuring the flux through a PSF (0.35") aperture and producing a table of the key quantities of interest (i.e S/N, frequency, [CII]-SFR, [CII] luminosity etc.). Once this has taken place for the entirety of the positive cube, the procedure is repeated in precisely the same fashion for the negative cube. The end result is photometric catalogues for candidates the positive cube, with an identically produced catalogue of the negatives.

3.3 Analysis of the catalogues

Photometry catalogues are obtained for positive and negative candidates within the four data cubes involved in this study (see section 2.1 for further details), the statistics of which provide the basis of the analysis which follows in this section. Firstly, duplicates are removed from the catalogues. This is achieved using an internal matching algorithm, replacing any sources within a 5 pixel radius in any adjacent slices with only the most luminous source. This reduces the number of overall candidates. The resulting number of detections are presented in Table 2.

There are two main features when studying the statistics of the tables, that are signatures of potential real emission. One signature is an excess in positive detections over the negative detections. This is because the degree of noise contamination is expected to be traced symmetrically in the negative and positive catalogues. Another sign of real emission, is an excess in positive detections over negative detections at the highest S/N end of the catalogues. It is at the highest S/N values where real emission is expected to reside (this can be supported with the study of line spectra outlined in section 4.1). Overall, we find an excess of 20 sources in the

Table 2. Number of positive and negative detections forming the photometry catalogues for each data cube, after duplicated detections have been removed.

Cube	Positive detections	Negative detections	Excess
spw3	1619	1599	20
spw2	1428	1442	-14
spw1	1469	1440	29
spw0	1529	1595	-66

Table 3. The number of the most robust appearing spectra appearing in each cube, resulting from a manual examination of all candidates. Cube SPW3 contains CR7, and the number of spectra excludes the spectra of CR7.

cube	number of candidate spectra
spw3	5
spw2	7
spw1	3
spw0	1

number of positives over the negatives in the CR7 cube. This is a potential signature of some significant emission beyond just CR7 in the cube, even after removing CR7 from the data. In the cube spw2, we find an deficit of 14 positive to negative candidates, however at the highest signal to noise we do find there is an excess in positives over negatives, suggestive of some likely real emission, which can be studied with further analysis outlined in this section. The cubes spw1 and spw0 at higher [CII] redshifts towards ~ 7.2 , on the other hand, we find an excess of 29 and deficit of 59 respectively, but with no excess in positives to negatives at the highest signal to noise values throughout the cube. This suggests that these cubes are likely tracing the spurious fraction of detections which are a characteristic of the noise inherent in ALMA data.

4 DATA ANALYSIS

4.1 Extraction of line spectra

Alongside the study of the large scale statistics of [CII] emitting candidates within the data cubes to estimate the number of previously undetected sources, the extraction of line spectra through the data cube enables characterisation of each individual source. By studying and comparing the line spectra of the confirmed [CII] emission within the clumps of CR7, and studying the typical form of the most significant negative spectra which correspond to noise within the data cube, we identify likely sites of genuine emission within the cube from their line spectra. Robust spectra are expected to exhibit a Gaussian form, which is both high in flux and velocity width, and exhibiting a high S/N. Spectra on the other hand whose magnitude may be high, but line width low, or vice versa are likely to be noise.

After manually examining the line spectra obtained from every candidate detection, we find the number of robust spectra within each cube given in Table 3. The candidates with the most convincing spectra we then compare directly with data from *HST* with the F814W, F110W (YJ) and F160W (H) data (see section 2.2). The following scenarios explain the nature of the candidate emitters; For correspondence between the coordinates of the candidate emitter with the F814W data, the source is likely a CO emitter at lower

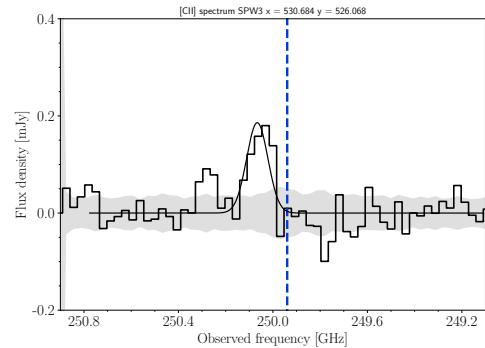


Figure 2. The line spectrum of clump B of CR7. The grey region corresponds to the 1σ noise level, and the blue line is the center of Lyman- α emission of CR7. (Matthee et al. 2017b)

Table 4. Properties of the sources corresponding to the 5 convincing line spectra obtained from the CR7 data cube spw3. The properties of the candidates are detailed in table 4.

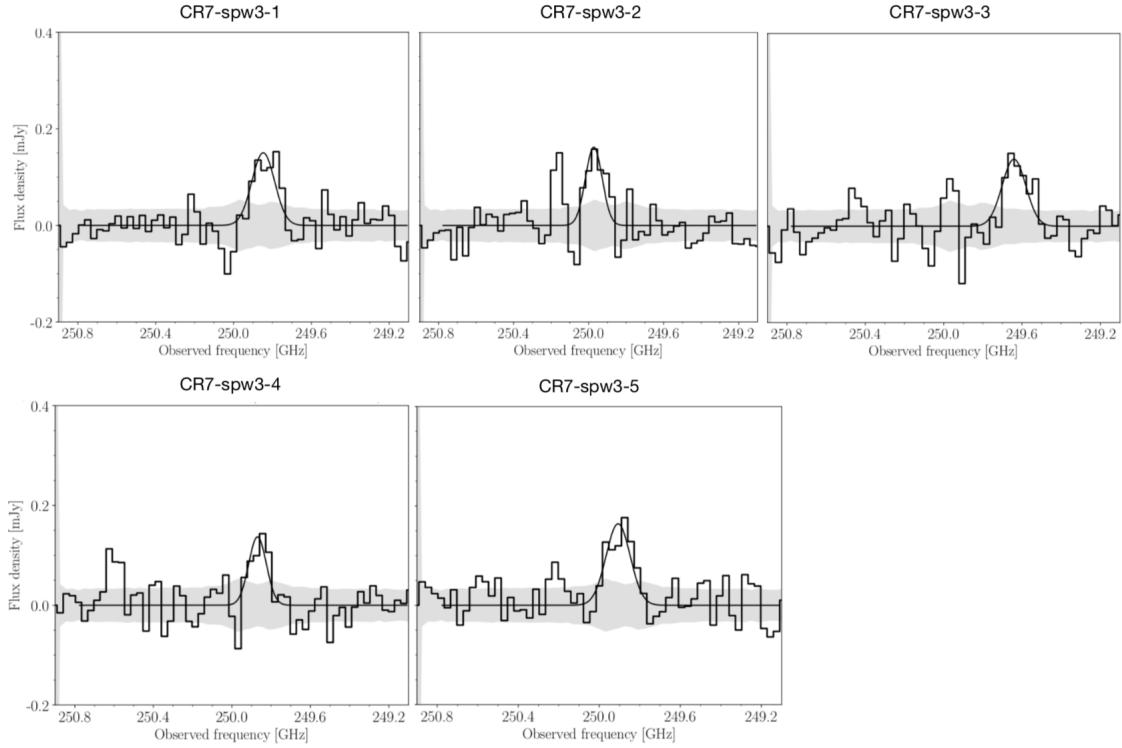
Source	R.A.	Dec.	z_{CII}	$\log_{10}(L/L_\odot)$
CR7-spw3-1	10 00 58.1	01 48 28.46	6.608	6.87
CR7-spw3-2	10 00 58.68	01 48 28.22	6.604	6.81
CR7-spw3-3	10 00 57.2	01 48 27.05	6.613	7.06
CR7-spw3-4	10 00 56.54	01 48 19.22	6.606	6.88
CR7-spw3-5	10 00 57.02	01 48 10.0	6.605	6.91

redshift. For no detection in F814W but detected in the F110W/YJ or F160W/H stack, then the source is either a high-z source or a dusty low-z detection. For no detection in any of the HST data, we can conclude that the source is either a genuine site of [CII] emission at high redshift, or a spurious detection.

4.1.1 Catalogue of robust line spectra

In Figure 3 we present the catalogue of most robust line spectra for the sources immediately surrounding CR7 in the cube spw3, following the procedure outlined in section 4.1. We present the thumbnail images to directly compare the region of the candidates with ALMA and HST data in Figure 4. In total for spw3, we find 5 spectra that appear to exhibit characteristics that are expected for real sources, including a high S/N, significant line width, and approximately Gaussian profile. The line spectra of the vast majority of the candidates in the catalogue did not satisfy these criteria, and in fact matched closer with the form of the spectra that formed the negative catalogues which correspond purely to noise. This does not mean that all of these sources must be simply noise, however, and presence in this catalogue of most convincing spectra does not guarantee their existence, and only future observation/follow up will unveil which of these sources are in fact real. The candidates presented here, stood out from the rest of the catalogue, and hence are the most likely sites of real emission. The candidates exhibit varying degrees of convincingness, with some of the sources appearing stronger candidates than others. Their properties are outlined in Table 4.

From Figure 4, a foreground star present in the COSMOS field catalogues (Laigle et al. 2016) is clearly visible in the HST data for source CR7-spw3-1. At present we are uncertain whether the star could be bright in ALMA at these frequencies, however,

Figure 3. The best line spectra of candidates surrounding CR7, obtained resulting from a manual examination of the spectra of all candidates in the cube SPW3.

as well as noting the slight offset in position. It remains uncertain if the foreground star will affect our measurement, and if this is a contaminant or genuine high-z emission. For this reason, we disregard it as a strong candidate for now. For the remaining sources, we find strong detections in ALMA through apertures and line spectra, and no *HST* counterparts, and hence following the procedure outlined in section 4.1 to characterise the nature of the sources, they likely correspond to either genuine high-z [CII] emission, or are extremely strong spurious detections.

In the Appendix B, we present the results in the same form for spw2, spw1 and spw0. For spw2, we find 6 convincing spectra with varying degrees of confidence. We find for source 5, an object visible in all 3 Hubble bands in Figure B, a match in the COSMOS field at $z = 1.55$, rendering it likely a CO(5-4) contaminant (see Figure 1).

Within spw1, in Figure B we find 3 convincing line spectra. Source 1 is detected in the three HST bands, however is not detected in the COSMOS field catalogues. The likely scenario is that this is low luminosity CO emitter somewhere between $0.5 < z < 2$ as can be inferred from Figure 1. Source 2 appears strong in the spectrum, although is towards the end of the frequency channel and hence is more prone to noise contamination. The spectrum of source 3 exhibits characteristics not entirely dissimilar to that of source 1.

From the study of the spectra in spw0, we find one source with a strong spectrum which lacks detection in HST data. The nearest source in the cosmos field is offset by $0.8''$ and is unlikely to affect the detection. The spectrum appears relatively strong when compared with the rest of the cube.

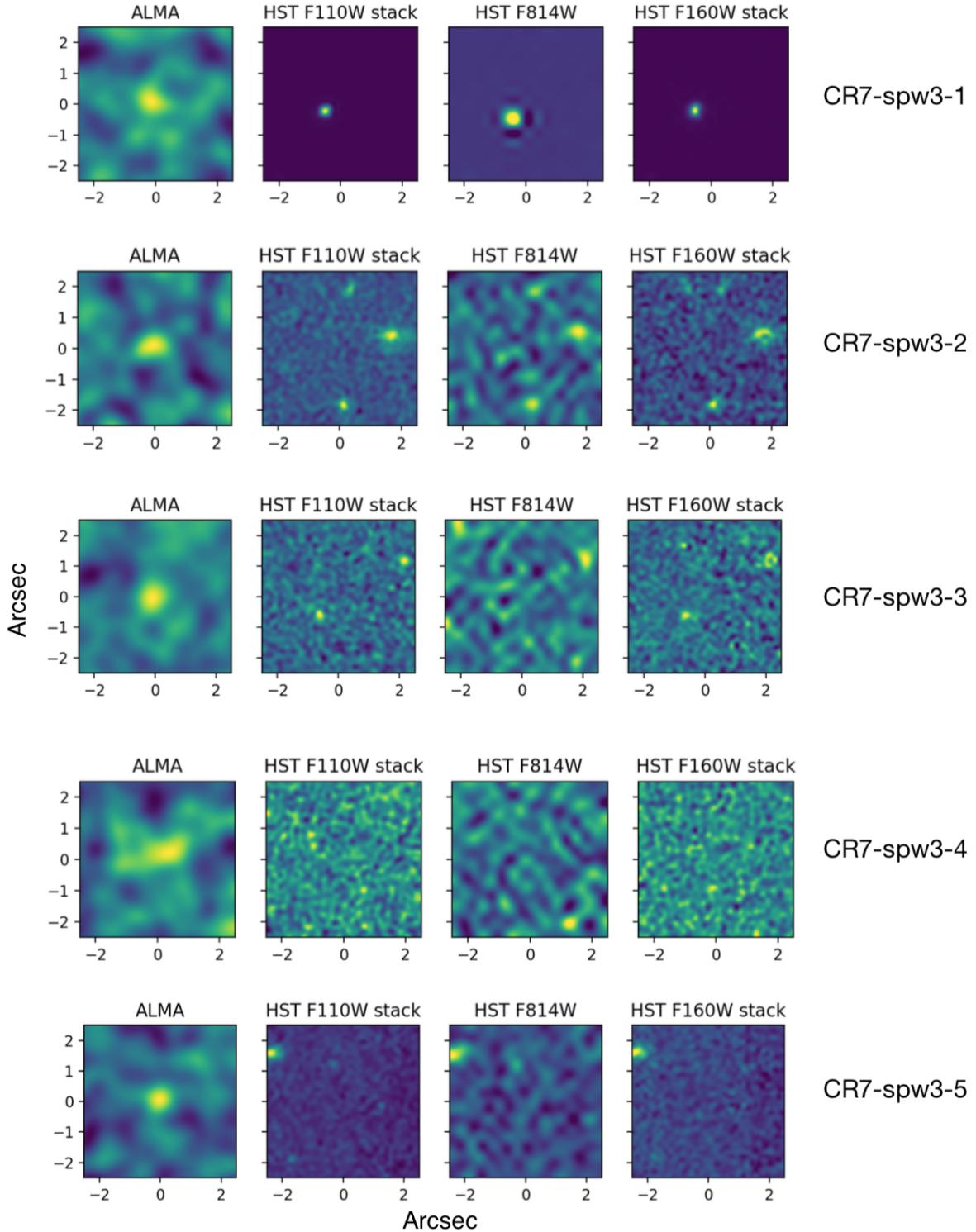
Table 5. We compare the location of matches in our positive and negative catalogues produced from our analysis, with the COSMOS field photometry catalogue (Laigle et al. 2016). There is expected to be an excess of matches between the positive and cosmos catalogues, compared with the negative and cosmos catalogues, if significant contamination due to lower redshift CO interlopers is present. There is no significant excess found between the positives and negatives, and in fact a deficit is found in most cases, suggestive of little/no influence on our results due to CO line emission.

Cube	Positive matches	Negative matches
spw3	171	180
spw2	165	171
spw1	171	167
spw0	176	180

4.2 Quantifying the fraction of contaminants

To search through our tables for any significant contamination due to CO emission at lower redshift, we crossmatch our produced catalogues from both the positive and negative cubes with the COSMOS field photometry catalogues (Laigle et al. 2016, see section 2.3). We expect for any significant contamination to take place, there will be an excess of matches of sources in the positive catalogues over the negative catalogues. Also, when studying the redshift distribution of the matches in the cubes, it is expected that there will be a peak of matches at the expected CO line redshifts outlined in Figure 1.

The results of the crossmatch between the cube catalogues and the COSMOS field outlined in Table 5 suggest that there is no significant contamination in the data due to CO lines. Particularly within spw3, spw2 and spw0 which exhibited a deficit in positive

Figure 4. The ALMA and HST data corresponding to the locations of the sources presented in figure 3, as 5"×5" thumbnails images.

catalogue matches with the COSMOS data when compared with the negatives. The cube spw1 which exhibited an excess in matches, is not a signature of anything significant, particularly as there are no clear signs of genuine emission within the cube in the first place.

The redshift distribution of the matches includes a significant amount of variation, and hence we must take care to check for

correspondence with our most robust candidates which we find using the procedure outlined in section 4.1, with results given in section 4.1.1.

4.2.1 Establishing the purity

A method to analyse the photometric catalogues is as described in Aravena et al. (2016), in which the highest signal to noise candidates in the positive and negative photometry catalogues are directly compared. The ‘purity’ or ‘fidelity’ of the sample is defined as:

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

i.e for a S/N cut off value, the number of positive and negative sources residing above the threshold are directly compared. This establishes a catalogue of sources for further consideration, with each of the positive candidates above the S/N threshold having an estimated probability P of corresponding to a real as opposed to spurious detection. This method has the benefit of providing a more robust catalogue for consideration, meaning the analysis can be targeted to those within the catalogue. By directly comparing the highest signal to noise candidates in the positive and negative catalogues, and measuring the excess in positives, for each cube we measure the likelihood of genuine emission within the cube. If no offset is detected, then following from this method it is unlikely that the cubes contain any real sources. Once the catalogues have been analysed in this fashion, points plotted on the luminosity function can be directly compared with luminosity function predictions.

4.2.2 Purity results

We apply the method as described in section For the CR7 cube spw3, we obtain for the 70% purity limit at 4.77σ , a total of 10 sources in the positive catalogue and 3 in the negative. In the cube spw2 slightly beyond CR7, for the same cut at 4.77σ , we obtain a purity of 38% suggesting likely significant genuine emission, albeit less significant than for the CR7 cube. In the cube spw1 we find at 4.77σ a purity of 43%, however this drops off rapidly and so may not be fully representative. For spw0 at $z_{CII} \sim 7.2$, we can not establish a purity limit because there is no significant offset of positive to negative candidates at the highest S/N. Whilst there may be a small number of real sources in the cube, explored in the study of the line spectra, there is no catalogue which we may obtain using the purity technique.

4.2.3 Searching for excess in bins of Luminosity

Here we outline a method which can be used to analyse the positive and negative photometry catalogues, which determines a value for the number of sources within the data cube, thus enabling production of a luminosity function, and enabling the estimation of the relative overdensity of the data cubes. The technique we use counts the difference between the number of positive over negative sources within bins of luminosity. It is the difference in the number of sources that defines the number of sources we expect to be real, within the luminosity range and within the full cube volume being surveyed. By dividing the number of sources within the bin with the volume of the cube (Table 1), we obtain a measurement of the number density of sources for the data cube. The three cubes beyond CR7 (enclosing three times the CR7 cube volume, $\sim 300\text{Mpc}^3$) are computed in a single bin, with the result normalised. For the higher luminosity bins, this method is expected to provide an accurate representation of the quantity of real sources within the cube, with lower luminosity bins becoming significantly noise contaminated. The resulting number density of each data cube enables the plot-

ting of the complete luminosity function in section 5.1, as well as estimating the relative overdensity of CR7 presented in section 5.2.

5 RESULTS

5.1 Constraining the high-z [CII] luminosity function

In Figure 5, we present the results of the method of analysis outlined in section 4.2.3, in which in bins of luminosity, we subtract the number of negative candidates from the positive candidates in the photometric catalogues. We consider the flux rather than the [CII] luminosity to not presume the flux necessarily corresponds to purely [CII] emission. At present, there is a lack of observational constraints on the high-z [CII] luminosity function, with only the aforementioned constraints provided by the [CII] survey of the HUDF from Aravena et al. (2016). We compare our results with predictions, namely those produced by Popping et al. (2016) who provides predictions of the [CII] line at $z = 6.0$, and also of the expected CO transitions at lower redshift. The predictions are generated from the coupling of a semi-analytical model of galaxy formation combined with radiative transfer modelling. We also compare with the predictions of Lagache et al. (2017), who utilize photoionization code CLOUDY modelling with G.A.S (semi-analytical model of galaxy evolution), comparing with the predicted [CII] luminosity function at redshift $z = 6.7$. We convert all of the predicted lines from luminosity to flux, to enable direct comparison. The pink line combines the predictions of the CO line luminosity function expected to trace the level of contamination, with the [CII] line prediction at $z = 6.0$ provided in Popping et al. (2016). This line is incomplete for likely contaminants, as the CO(3-2) predictions provided in Popping et al. (2016) do not go faint enough to be comparable with our sample, however the contribution from this line would likely be marginal. The bin encompassing the reference volume resides only slightly above this predicted line, providing further evidence to suggest it traces the expected number density of CO and [CII] emitting sources for a wider field, and supporting the overdensity estimate in section 5.2. We can also readily compare the CR7 spw3 cube (black) and the same cube with the source CR7 removed (blue) points with the reference volumes, once more suggesting a significant overdensity for CR7 even neglecting the contribution by the source itself.

5.2 The relative overdensity of CR7

The plot as shown in Figure 6 measures the relative overdensity of the CR7 data cube when compared with a reference volume comprised of cubes spw0, spw1 and spw2 ($z \sim 6.6\text{--}7.2$, 3x CR7 cube volume). By taking the ratio of the number density of candidates in the CR7 cube with the full reference volume, the relative overdensity of CR7 is calculated to be $3.95^{+0.91}_{-0.76}$. This result is obtained when considering a bin which encompasses the brightest end of the statistics, and assuming no/little CO line contamination, the scenario favoured by the analysis outlined in section 4.2. We also observe the three cubes which comprise the normalized reference volume reside slightly above or below the reference volume value, suggesting that they are more representative of the general population in a field of that size, providing further evidence to suggest that CR7 resides in a unique and populated region. At lower [CII] luminosities the sample being considered becomes significantly contaminated with low S/N candidates. We expect only the brightest end of the luminosities to be representative of the real sources, where any significant excess in the number of positive over negative candidate detections

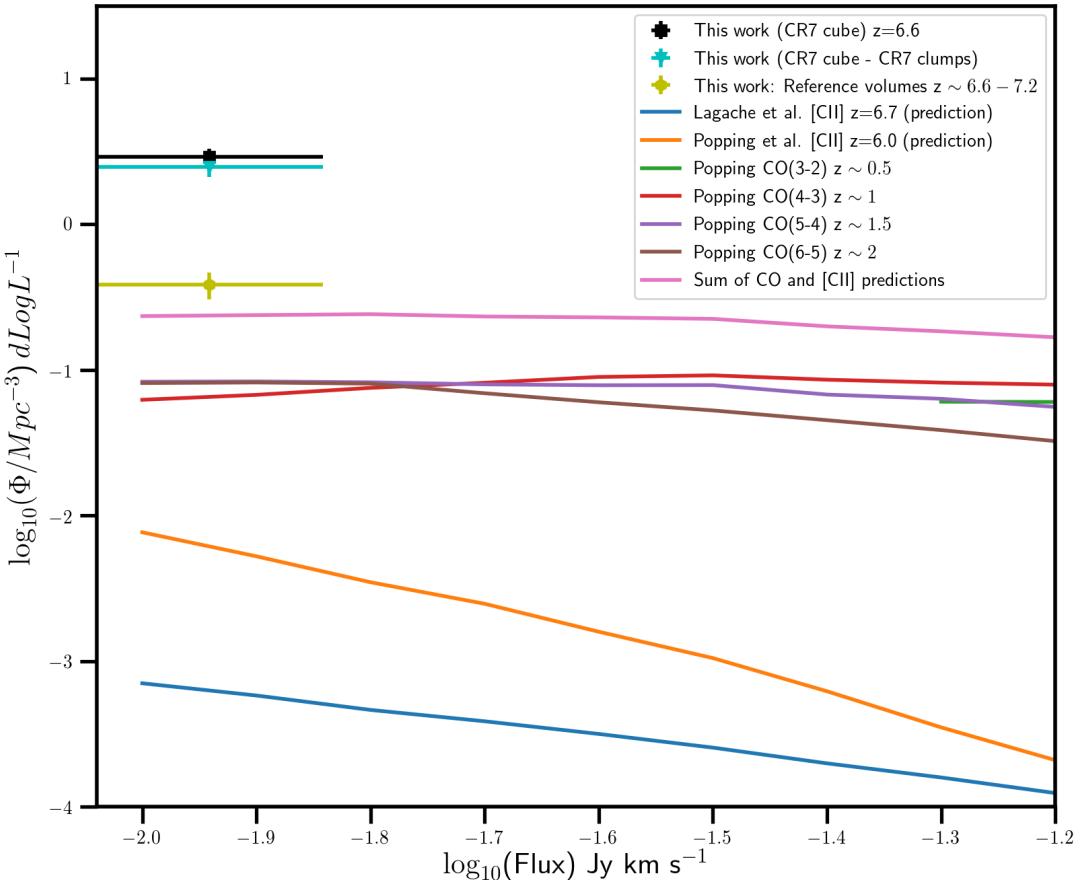


Figure 5. Luminosity function compared with LF predictions of [CII] as well as likely CO contaminants which could exist within our frequency channels. (Popping et al. 2016; Lagache et al. 2017)

is present. The errors obtained from this calculation are calculated using percentiles.

6 DISCUSSION

6.1 Summary of our results

As discussed in the introduction to this study, CR7 exists in a unique environment (e.g. Sobral et al. 2015, 2017b; Matthee et al. 2017b) when considering the epoch it resides within. The strong detection of CR7 in Ly α well into the epoch of reionization at $z \sim 6.6$ (Sobral et al. 2015), suggests that ionizing photons have cleared the region surrounding CR7 of neutral hydrogen such that the Ly α photons are directly detectable in the line of sight (e.g. Robertson et al. 2015). The question of the nature of the reionization of CR7's surrounding intergalactic hydrogen is the focus of this study. In particular, could an underlying overdensity of galaxies in the surrounding region contribute to the ionizing flux required to cause rapid reionization, leading to CR7's detectability in Ly α ? (e.g. Choudhury et al. 2009; Robertson et al. 2015; Faisst 2016; Castellano et al. 2016)

Using catalogues of candidate sources obtained from a blind survey of ALMA data covering CR7 (Matthee et al. 2017b) and the region beyond CR7 ($z \sim 6.6 - 7.2$), we present evidence in support of the scenario that CR7 is likely residing in an overdense region. Whilst each of the techniques utilized to demonstrate the overdensity has their limitations, the result remains the same in support of the relative overdensity of CR7.

Presented in section 5.2, is a direct comparison of the number density of the CR7 data cube with the number density obtained from a volume that is three times larger than the CR7 cube volume, expected to be representative of the general field (spw0, spw1 and spw2 combined). The number of sources within the volumes is defined by the number of negative blind detections subtracted from the number of positive blind detections, which at the highest end of the luminosities is expected to be representative of the actual number of sources. We utilize the negative mapping of the cube to determine the degree of contamination in the data cube, which is expected to be symmetric, and hence any significant excess in the number of positive over negative candidates is a likely sign of real emission. We set the upper limit of the luminosity bin and choose a bin that is wide enough such that our result is conservative.

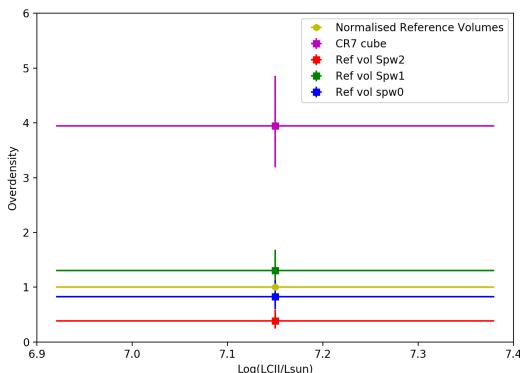


Figure 6. The relative overdensity of CR7: In bins of luminosity, the number of positive candidates is subtracted from the number of negatives. The remaining excess of sources are considered to be real. The beige point encompasses the statistics of the three cubes beyond CR7, which has been normalised for comparison with the individual cubes. The ratio of the number density of each individual cube with the number density of the reference volume we use as a measure of the relative overdensity of the cube. This data suggests that CR7 resides in an overdensity of $3.95^{+0.91}_{-0.76}$

Following this, the value of the overdensity we obtain is $3.95^{+0.91}_{-0.76}$. Defining the luminosity bin width is relatively arbitrary, with small changes in bin width having potentially a significant effect on the obtained value of the overdensity, so we remain on the conservative side.

This claim is substantiated when we consider the luminosity function we present in Figure 5, providing further evidence to suggest that CR7 resides in an overdense region. We see that the CR7 data cube catalogue with the three clumps of CR7 removed (blue), exhibits a significantly higher number density when compared with the reference volume (beige) for the most luminous, high S/N end of the catalogue. Within the flux bin the CR7 data cube spw3 with CR7 itself removed, contains 46 sources in 93 Mpc^3 , whereas the reference volume contains only 26 sources in 297 Mpc^3 . This is suggestive of a significant overdensity for the CR7 volume.

The purity technique presented in section 4.2.2 and as outlined by Aravena et al. (2016), provides a further source of evidence to suggest that CR7 resides in a significant overdensity, by measuring the excess of positive to negative candidates above a S/N limit for each of the cubes. We find for CR7 a relatively significant offset of positive to negative candidates, an indication that there is likely genuine emission occurring in the cube. Likewise, we find a smaller yet distinct excess in the spw2 data cube just beyond CR7, another signature of genuine emission. The cube spw1 returns some signatures of genuine emission using this technique, however this is subject to where the S/N is cut is placed. The cube spw0 returns no result utilizing this technique, indicating that the cube is relatively empty, and any potential genuine emission will be uncovered through the study of line spectra. This method when compared with the luminosity binning technique as outlined above, agree on the scenario that CR7 resides in an overdense region. The disagreement, however, concerns the scenario presented by the other three data cubes. Whilst purity considerations suggest that spw2 at $z \sim 6.6$ has genuine signs of emission, when considering

the luminosity binning technique spw2 appears to exhibit a number density lower than that obtained from the cubes spw1 and spw0 at $z \sim 7.2$. This is likely the result of the wide luminosity bin which we utilize in figure 6 in an effort to remain conservative in our selection, coupled with the inherently different range of luminosities between the different data cubes.

From the study of the line spectra throughout the data cubes, we find some relatively convincing signs of emission throughout the ALMA data sets. The most convincing signs appear to be found in spw3 with 4 robust sources surrounding CR7, and 5 in spw2 just beyond CR7. As we progress to the higher redshift cubes spw1 and spw0, the number of sources decreases to 2 and 1 respectively. Its important to highlight that the selection criteria at present is arbitrary and is merely a relative comparison between the full sample of spectra observed. Whilst these spectra were selected based upon their appearance, due to the complex nature of the spectra, as well as the number of factors which contribute to their classification as either a potential candidate source or simply noise, the selections made in this section arise from conjecture and their nature remains speculative. Follow up observation is required to elucidate their true nature.

6.2 Implications on reionization

The evidence presented in this study suggests that CR7 does indeed reside in a significant overdensity, obtaining a factor of $3.95^{+0.91}_{-0.76}$ overdense. We utilize extremely deep ALMA observation which is sensitive to sources as faint as $10^{7.0-7.3} L_\odot$, with FIR [CII] emission which is unaffected by intervening intergalactic hydrogen (e.g. Vallini et al. 2015; Maiolino et al. 2015; Watson et al. 2015; Aravena et al. 2016; Pentericci et al. 2016; Matthee et al. 2017b). It is this overdensity which could likely be the cause the rapid and early reionization of the surrounding IGM, and explain CR7's visibility in Ly α and LyC. As well as demonstrating this with a statistical analysis, we find 4 robust [CII] line spectra within CR7's immediate surroundings, which are the most likely candidates for real galaxies in the data. The evidence we present in this study supports the picture presented in Castellano et al. (2016), in which an overdensity of galaxies coincides with the overlapping of ionized bubbles deep into the epoch of reionization at $z \sim 7$. The signatures of the existence of a population of faint galaxies surrounding the extremely luminous galaxy CR7, suggests we are likely witnessing the build up of a cluster of galaxies.

6.3 Next Steps

It is of interest to extend this study where possible to other luminous Ly α emitters deep into the epoch of reionization, to see if rapid reionization corresponds with relatively overdense regions, and begin to explore if this is a characteristic of high-z Ly α emitters in general. Using tools such as ALMA to perform deep surveys in the FIR, we can extend this study further. Similar high quality data already exists for Himiko (Ouchi et al. 2013; Carniani et al. 2018), a $z = 6.595$ Ly α emitter, which will provide the opportunity to consider this question in more depth. Furthermore, as well as providing further opportunity to explore the questions of environment and reionization, increased statistics will enable improved constraints to be placed on the [CII] luminosity function

into the epoch of reionization.

Another question of interest, is of the nature of the environments surrounding LBG's which are not detected significantly in Ly α at similar redshifts. This could unveil further evidence which will inform the question of how the environment of galaxies into the epoch of reionization may affect the rate at which reionization takes place.

7 CONCLUSION

In this study, we have presented the development of an ALMA blind survey performed upon the CR7 data set, including the CR7 data cube and the three data cubes at successively higher redshifts. We produce photometric catalogues for both the positive and negative mapping of the data cube, to account for noise contamination within the data sets. We then analyse the photometric catalogues to obtain the key scientific results of the study. The results of this study include:

(i) Demonstration that CR7 resides in an $3.95^{+0.91}_{-0.76}$ times overdensity. This provides evidence suggesting that CR7 is surrounded by a population of faint galaxies, whose ionizing flux could be the cause of CR7's rapid and early reionization of its surrounding hydrogen in the IGM.

(ii) Constraining the faint end of the high- z [CII] luminosity function between $10^{7.0-7.3} L_{\odot}$, enhancing the present understanding of the population of [CII] emitting galaxies in the epoch of reionization.

(iii) Presentation of the 4 robust line spectra of candidates for previously undetected [CII] emitting galaxies surrounding CR7, as well as a further 8 candidates at higher redshift, based upon the strength of their [CII] emission. After searching for correspondence in HST data, we find 4 strong candidates for genuine [CII] emission surrounding CR7, suggesting we may be witnessing the build up of a cluster of galaxies in the early Universe, with CR7 at its center.

Further studies conducted in a similar fashion upon other Ly α emitters will provide insight into the question of the nature of the environments conducive to rapidly reionizing the region surrounding them. As future studies take place, a greater number of statistics will provide further constraints upon the high- z [CII] luminosity function, as well as increase the size field reference volumes to improve the statistics and obtain more illustrative results for the overdensity.

REFERENCES

- Alvarez M. A., Abel T., 2012, *ApJ*, **747**, 126
 Aravena M., et al., 2016, *ApJ*, **833**, 71
 Atek H., Richard J., Kneib J.-P., Schaerer D., 2018, *Mon. Not. Roy. Astron. Soc.*, Becker R. H., et al., 2001, *AJ*, **122**, 2850
 Becker G. D., Sargent W. L. W., Rauch M., 2004, *ApJ*, **613**, 61
 Becker G. D., Bolton J. S., Madau P., Pettini M., Ryan-Weber E. V., Venemans B. P., 2015, *Mon. Not. Roy. Astron. Soc.*, **447**, 3402
 Bertin E., Arnouts S., 1996, *aaps*, **117**, 393
 Bosman S. E. I., Fan X., Jiang L., Reed S. L., Matsuoka Y., Becker G. D., Haehnelt M. G., 2018, *mnras*
 Bouwens R. J., et al., 2015a, *ApJ*, **803**, 34
 Bouwens R. J., Illingworth G. D., Oesch P. A., Caruana J., Holwerda B., Smit R., Wilkins S., 2015b, *ApJ*, **811**, 140
 Carniani S., Maiolino R., Smit R., Amorín R., 2018, *apjl*, **854**, L7
 Caruana J., Bunker A. J., Wilkins S. M., Stanway E. R., Lorenzoni S., Jarvis M. J., Ebert H., 2014, *MNRAS*, **443**, 2831
 Castellano M., et al., 2016, *ApJ*, **818**, L3
 Castellano M., et al., 2018, *Astrophys. J.*, **863**, L3
 Choudhury T. R., Haehnelt M. G., Regan J., 2009, *Mon. Not. Roy. Astron. Soc.*, **394**, 960
 Dijkstra M., Wyithe S., Haiman Z., Mesinger A., Pentericci L., 2014, *Mon. Not. Roy. Astron. Soc.*, **440**, 3309
 Erb D. K., 2015, *Nature*, **523**, 169
 Faisst A. L., 2016, *Astrophys. J.*, **829**, 99
 Fan X., et al., 2006, *AJ*, **132**, 117
 Finkelstein S. L., 2016, *Publ. Astron. Soc. Australia*, **33**, e037
 Fletcher T. J., Robertson B. E., Nakajima K., Ellis R. S., Stark D. P., Inoue A., 2018, *ApJ*
 Furlanetto S. R., Oh S. P., 2005, *Mon. Not. Roy. Astron. Soc.*, **363**, 1031
 Gunn J. E., Peterson B. A., 1965, *ApJ*, **142**, 1633
 Hu E. M., Cowie L. L., Songaila A., Barger A. J., Rosenwasser B., Wold I., 2016, *Astrophys. J.*, **825**, L7
 Inoue A. K., Shimizu I., Iwata I., Tanaka M., 2014, *MNRAS*, **442**, 1805
 Inoue A. K., et al., 2016, *Science*, **352**, 1559
 Katz H., Kimm T., Haehnelt M., Sijacki D., Rosdahl J., Blaizot J., 2018, preprint
 Koekemoer A. M., et al., 2007, *ApJS*, **172**, 196
 Lagache G., Cousin M., Chatzikos M., 2017, preprint, ([arXiv:1711.00798](https://arxiv.org/abs/1711.00798))
 Laigle C., et al., 2016, *Astrophys. J. Suppl.*, **224**, 24
 Livermore R. C., Finkelstein S. L., Lotz J. M., 2017, *ApJ*, **835**, 113
 Maiolino R., et al., 2015, *MNRAS*, **452**, 54
 Mason C. A., et al., 2018, *Astrophys. J.*, **857**, L11
 Matthee J., Sobral D., Darvish B., Santos S., Mobasher B., Paulino-Afonso A., Röttgering H., Alegre L., 2017a, *MNRAS*, **472**, 772
 Matthee J., et al., 2017b, *ApJ*, **851**, 145
 Matthee J., Sobral D., Gronke M., Paulino-Afonso A., Stefanon M., Röttgering H., 2018, preprint, ([arXiv:1805.11621](https://arxiv.org/abs/1805.11621))
 McGreer I. D., Mesinger A., D'Odorico V., 2015, *MNRAS*, **447**, 499
 McMullin J. P., Waters B., Schiebel D., Young W., Golap K., 2007, in Shaw R. A., Hill F., Bell D. J., eds, *Astronomical Society of the Pacific Conference Series Vol. 376, Astronomical Data Analysis Software and Systems XVI*. p. 127
 Moriawki K., et al., 2018, preprint, ([arXiv:1805.07062](https://arxiv.org/abs/1805.07062))
 Mortlock D. J., et al., 2011, *Nature*, **474**, 616
 Natarajan A., Battaglia N., Trac H., Pen U.-L., Loeb A., 2013, *ApJ*, **776**, 82
 Ono Y., et al., 2012, *ApJ*, **744**, 83
 Onoue M., et al., 2017, *ApJ*, **847**, L15
 Ouchi M., et al., 2013, *ApJ*, **778**, 102
 Parsa S., Dunlop J. S., McLure R. J., 2018, *Mon. Not. Roy. Astron. Soc.*, **474**, 2904
 Pentericci L., et al., 2016, *ApJ*, **829**, L11
 Planck Collaboration et al., 2016, *A&A*, **594**, A13
 Popping G., van Kampen E., Decarli R., Spaans M., Somerville R. S., Trager S. C., 2016, *MNRAS*, **461**, 93
 Robertson B. E., Ellis R. S., Furlanetto S. R., Dunlop J. S., 2015, *ApJ*, **802**, L19
 Sharma M., Theuns T., Frenk C., 2017, preprint, ([arXiv:1712.06619](https://arxiv.org/abs/1712.06619))
 Sobacchi E., Mesinger A., 2015, *Mon. Not. Roy. Astron. Soc.*, **453**, 1843
 Sobral D., Matthee J., 2018, *A&A*
 Sobral D., Matthee J., Darvish B., Schaerer D., Mobasher B., Röttgering H., Santos S., Hemmati S., 2015, *Astrophys. J.*, **808**, 139
 Sobral D., et al., 2017b, *mnras*
 Sobral D., Santos S., Matthee J., Paulino-Afonso A., Ribeiro B., Calhau J., Khostovan A. A., 2017a, *MNRAS*
 Songaila A., Hu E. M., Barger A. J., Cowie L. L., Hasinger G., Rosenwasser B., Waters C., 2018, *Astrophys. J.*, **859**, 91
 Stark D. P., et al., 2017, *MNRAS*, **464**, 469
 Steidel C. C., Bogosavljevic M., Shapley A. E., Reddy N. A., Rudie G. C., Pettini M., Trainor R. F., Strom A. L., 2018, preprint, ([arXiv:1805.06071](https://arxiv.org/abs/1805.06071))
 Tilvi V., et al., 2014, *ApJ*, **794**, 5

Parameter	Optimized Value
DETECT_THRESH	3.25σ
DETECT_MINAREA	39 pixels
BACK_SIZE	36 pixels
BACK_FILTERSIZE	4

Table A1. Table of optimized SExtractor parameter values, which we use within our blind survey. With these parameters handling the software’s detection and background determination capabilities, SExtractor blindly recovers 78% of the candidate sources we find in the test slice, with 121% of the total number of candidates we manually find during the visual inspection. The deviation between the catalogues occurs at the lowest signal to noise end of the sample, meaning the deviation will likely not affect our strongest candidates.

Vallini L., Gallerani S., Ferrara A., Pallottini A., Yue B., 2015, *ApJ*, **813**, 36
 Watson D., Christensen L., Knudsen K. K., Richard J., Gallazzi A., Michałowski M. J., 2015, *Nature*, **519**, 327
 Zahn O., Lidz A., McQuinn M., Dutta S., Hernquist L., Zaldarriaga M., Furlanetto S. R., 2007, *ApJ*, **654**, 12

APPENDIX A: OPTIMIZING SOURCE EXTRACTOR

To optimize the SExtractor parameters to survey the full cube, the software parameters are calibrated with the catalogues produced from the manual examination of the cube. To achieve this, the parameters concerning detection (DETECT_THRESH, DETECT_MINAREA) and background determination (BACK_SIZE, BACK_FILTERSIZE) are systematically varied over a wide range of parameter values and operated on a test slice of the cube, producing catalogues of detections. The produced catalogues are then systematically compared with the catalogue resulting from the manual analysis of the same data. The quantities which affect the quality of the match between SExtractor and eye detection, we denote the ‘covering fraction’ (the fraction of the same sources that are blindly recovered by SExtractor with given parameter values), and the ‘excess fraction’ (the fraction of sources that appear in the blind detections but were not detected in the manual examination of the slice). The quality of the match can be effectively parametrised by the ratio of the covering to excess fraction, finding that the closer to unity the result, the more optimal the match. In Figure A1 we present the results of this analysis. The optimal parameter values obtained are presented in Table A1 with a covering fraction of 78%, an excess fraction 121%, and with the ratio 1.55. The bulk of the deviation of the manual detections with the SExtractor produced detections, lies in the lowest S/N candidates, which are not considered in the final analysis. A good match between the highest S/N candidates remains, and hence this parameter set provides an ideal foundation for a blind analysis.

APPENDIX B: ROBUST CANDIDATES WITHIN CUBES **SPW0, SPW1 & SPW2**

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.

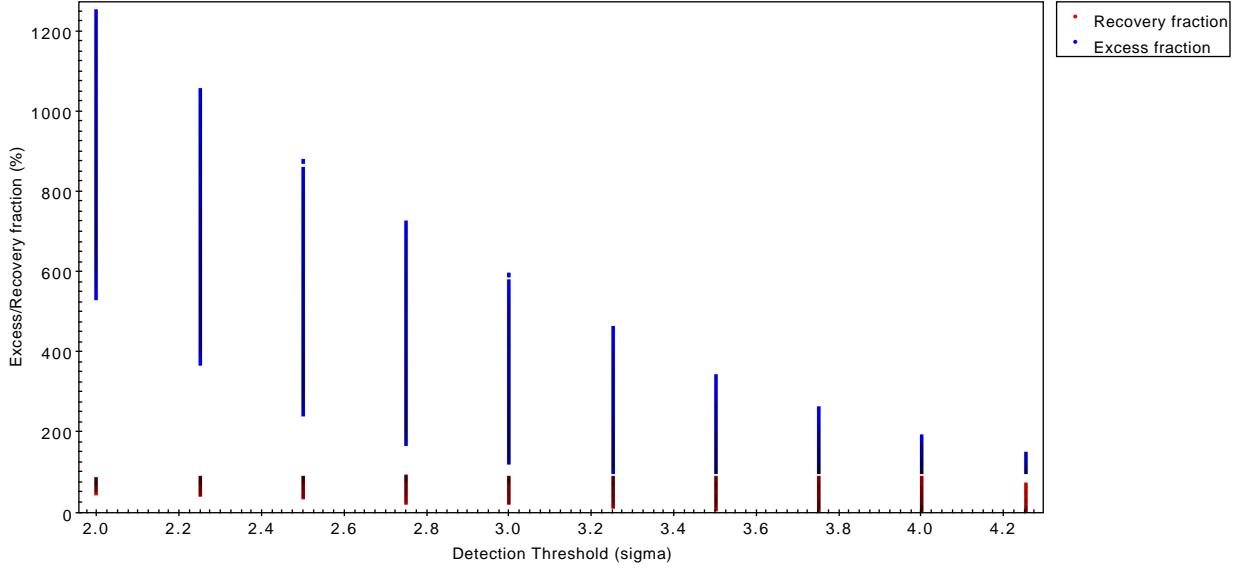
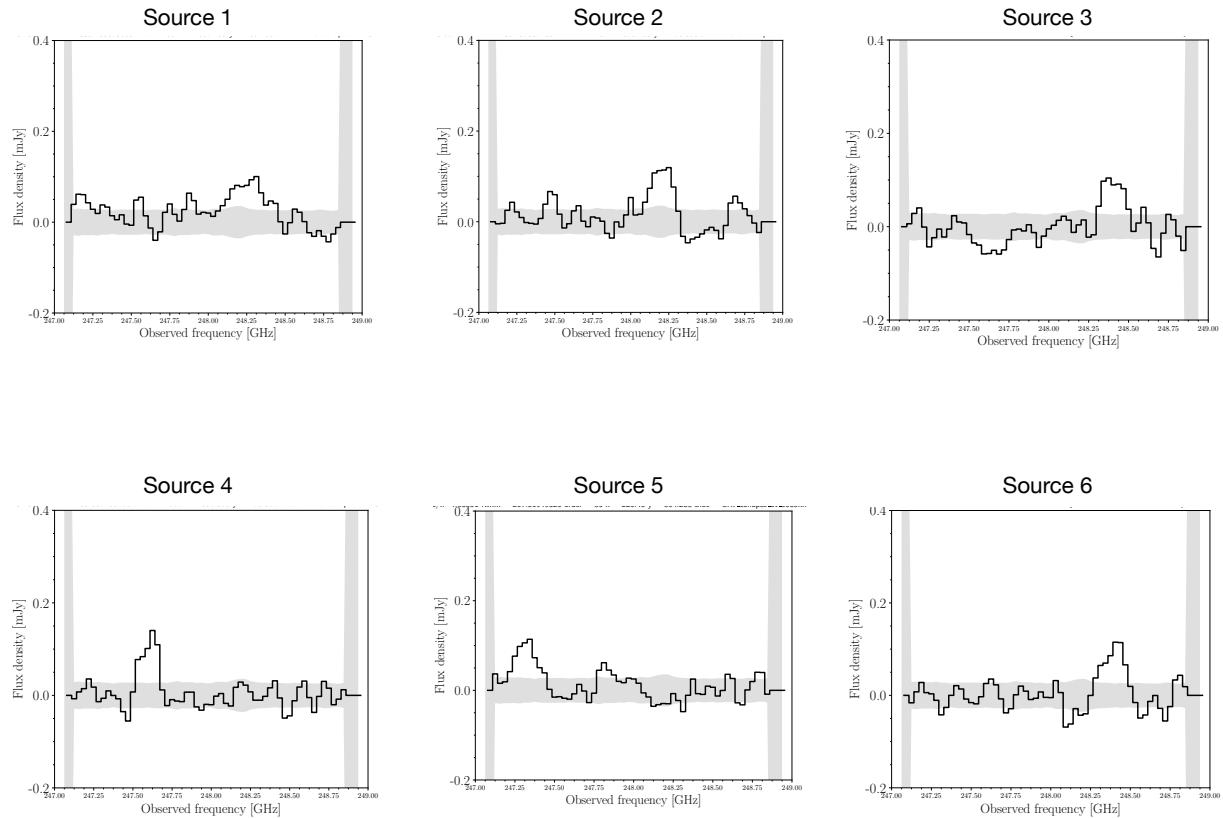


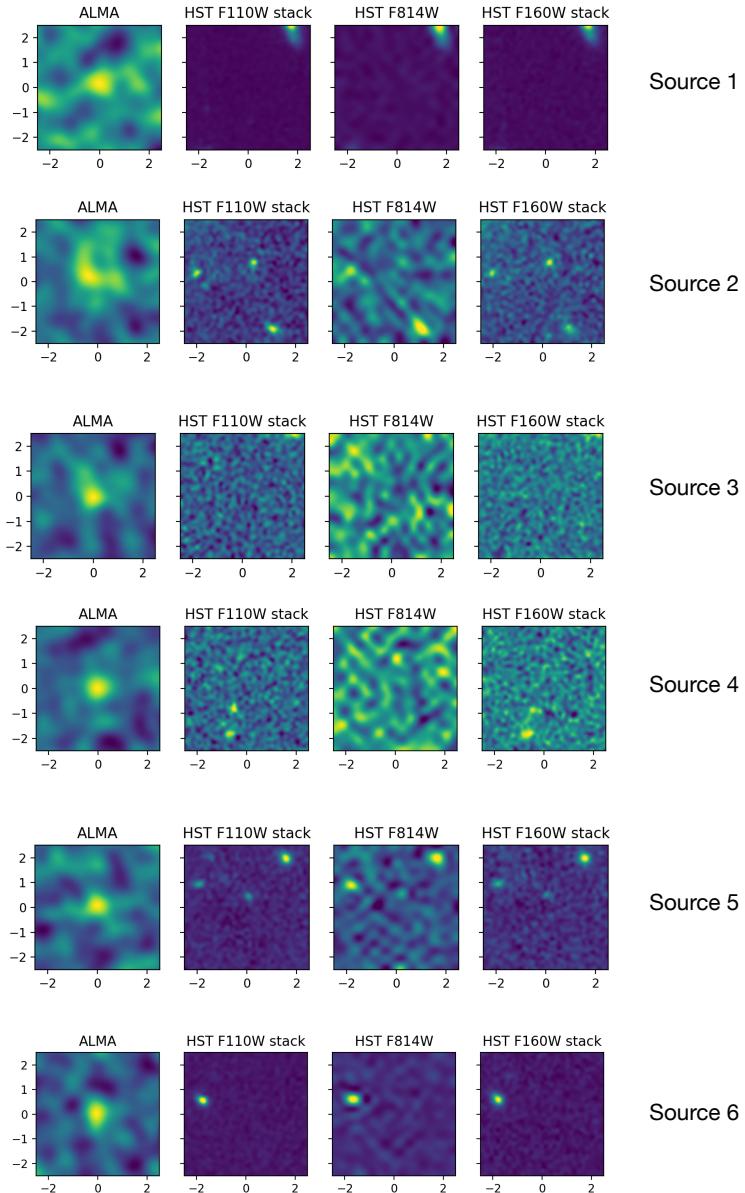
Figure A1. Variation of excess fraction (blue, which quantifies contamination of false detections from SExtractor) and recovery fraction (red, fraction of candidates successfully recovered by SExtractor) as the detection threshold parameter is varied, operated with a variety of SExtractor configurations. We find significant contamination from a high excess fraction, as the detection threshold was low. As detection threshold increases, the contamination from false detections becomes less significant. There is less variation experienced with the recovery fraction, finding it relatively steady throughout. The optimal parameter set obtained, defined by the lowest value of the ratio of excess to recovery fraction occurs at a detection threshold of 3.25σ .

[h]

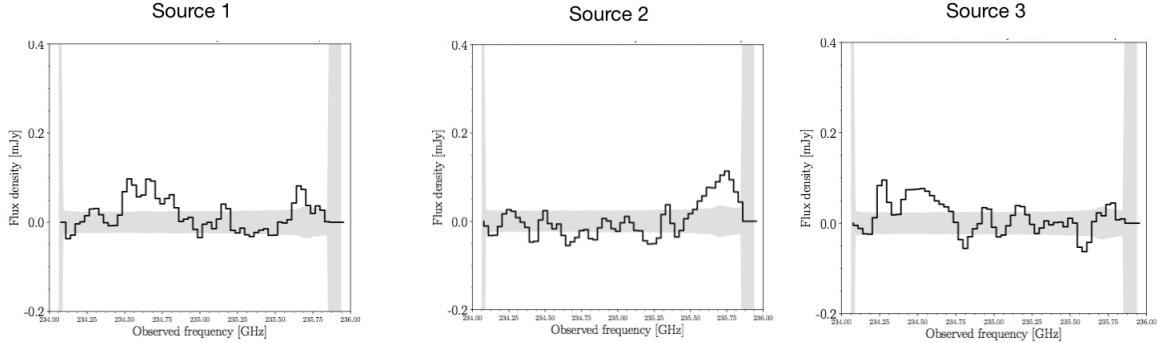
Figure B1. The most convincing line spectra of candidates in the cube spw2.



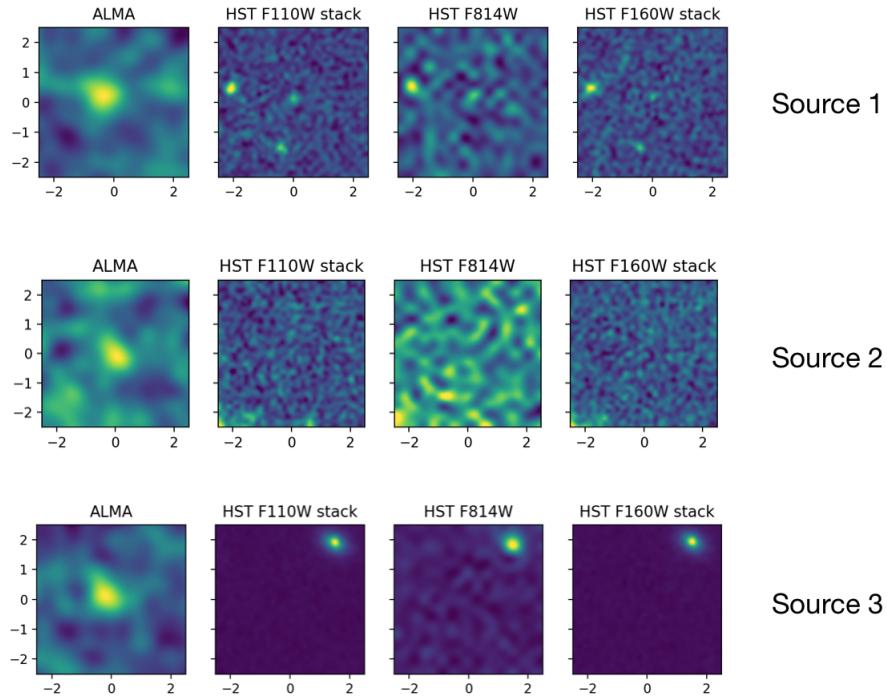
[h]

Figure B2. The thumbnails corresponding to the strongest candidates in spw2.

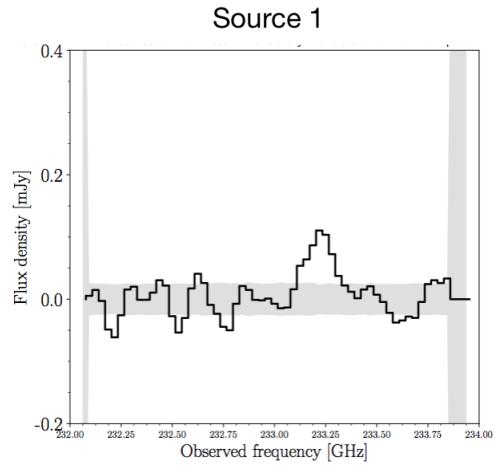
[h]

Figure B3. The most convincing line spectra of candidates in the cube spw1.

[h]

Figure B4. The thumbnails corresponding to the strongest candidates in spw1.

[h]

Figure B5. The most convincing line spectrum obtained from a visual inspection of spw0.

[h]

Figure B6. The thumbnails corresponding to the line detection in spw0