An ALMA [CII] survey of the environments around bright Lyman-α emitters in the epoch of re-ionisation

Ryan Cooper[1], Bentley Newton[1], Matthew Fletcher[1], Catherine Muller[1], Kyle Turner[1] & David Sobral[1]*

[1] *Department of Physics, Lancaster University, Lancaster, LA1 4YB, UK*

ABSTRACT

The bright early Lyman-α emitters (LAEs) studied here (CR7, VR7 and MASOSA) are some of the brightest known LAEs at high redshift. These galaxies likely reside in early ionised bubbles, producing a significant amount of ionising photons to sustain and reionise the surrounding intergalactic medium during the epoch of reionisation (EoR). It is not clear how those early ionised bubbles formed, but studying luminous LAEs in the EoR could bring us closer to understanding the processes involved. Here we use deep ALMA observations ($\sigma = 5$mJy beam^{-1} km s^{-1}) obtained around each bright distant LAE to survey for [CII] emitters at $z \approx 6.5 - 7.1$. This was done by looking for line emitters in the ALMA data using SExtractor and other statistical analysis methods. We have identified 9, 11 & 5 robust candidate [CII] line emitters within the full data-sets for CR7,VR7 & MASOSA respectively. By dividing the number density of sources in the vicinity of each LAE by the average number density of [CII] emitters at $z \approx 7$, we find that the immediate environment of bright LAEs is over-dense by factors of about 2 to 4. Our results are consistent with luminous LAEs within the epoch of re-ionisation being in moderately over-dense regions.

Key words: Galaxies: High Redshift, Galaxies: Statistics, Galaxies: Luminosity Function, Radio lines: Galaxies

1 INTRODUCTION

Approximately 400,000 years after the Big Bang, the Universe cooled enough for neutral hydrogen to form (Zaroubi 2012). The existence of this neutral hydrogen caused the Universe to become opaque to ionising photons. (Greig et al. 2019). The epoch of reionisation is the point at which large amounts of neutral hydrogen were ionised and the Universe slowly became transparent to ionising radiation.

There are large amounts of uncertainty regarding when reionisation occurred. Suggestions have been made that reionisation started on small scales due to the formation of the first stars at very high redshifts $z = 20 - 30$ (Natarajan et al. 2013). However, the process likely accelerated rapidly following the formation of the first galaxies at redshifts $6 < z < 12$ (Planck Collaboration et al. 2016). Massive stars within galaxies are largely considered the primary source of ionising photons, with other sources such as Active Galactic Nuclei (AGN's) contributing less (Meyer et al. 2019; Robert-son et al. 2015). The reionisation process first accelerated around these early galaxies, creating bubbles of ionised gas, leading to the idea of patchy reionisation (Natarajan et al. 2013; Sobacchi & Mesinger 2015). Ultimately these bubbles overlap, causing the Universe to become fully ionised. Reionisation is largely believed to have completed at around z = 6, whilst attempts to constrain a starting point are ongoing.

Since galaxies are likely the prime contributor to the reionisation process, it is worthwhile examining those present at high redshifts. Notably, Lyman-α Emitters (LAEs) are potentially the best candidates to identifying the earliest point at which the reionisation epoch occurred. Lyα lines are strong and recognisable in young, star forming galaxies (Behrens et al. 2019).This is because Lyman-α is sensitive to the presence of neutral hydrogen due to it being a highly resonant emission line. This makes them useful for studying high redshift galaxies, including at redshifts within the epoch of reionisation. It is worth noting however, that a rapidly declining luminosity function at redshifts beyond z=6 (e.g. Konno et al. 2017; Ota et al. 2017) may limit the effectiveness of these observations at earlier points. However,

* PHYS369 supervisor.

since highly luminous LAEs and the Ly-α photons they emit can be detected within the epoch of reionisation, the areas surrounding these galaxies must lack neutral hydrogen. This suggests there is an ionised bubble around the source that may have been created with the help of nearby galaxies contributing to the reionisation process.

It is possible that dimmer galaxies are providing more to the ionising photon budget than expected. Specifically, over-dense regions with a highly luminous LAE present may be the cause of the rapid reionisation of their surroundings and the visibility of the LAE's emissions. This is supported by previous findings that faint sources had contributed to patchy reionisation (Atek et al. 2015). Prior results have also suggested a possible over-density of galaxies surrounding the environment of the bright, Lyα emitter CR7 will have contributed to the reionisation rate of its local area (e.g. Day et al. 2019).

The experiment detailed in this paper looks at the environments surrounding three bright, high redshift LAEs, CR7, VR7 and MASOSA. With redshifts of 6.604, 6.532, and 6.541 respectively (e.g Sobral et al. 2015; Matthee et al. 2019), these reside well into to the epoch of reionisation. Each galaxy was discovered and analysed using data from some of the largest telescopes available, including the VLT and Keck (Sobral et al. 2015). The importance of these galaxies comes from their luminosity. Due to the nature of the intergalactic medium at such redshifts, many dimmer LAEs are undetectable. The unique opportunity to observe the surroundings of these galaxies allows us to potentially discover fainter companion galaxies.

We aim to identify possible sources surrounding our subject galaxies, and determine whether an over-density is present. Our primary method for detecting sources will be to identify [CII] emitters with a rest-frame wavelength $\lambda = 158\mu$m (Lagache, G. et al. 2018). [CII] emission is mostly from neutral gas (Lebouteiller et al. 2019) and can be used as a good indication for star formation. Therefore [CII] is often the brightest emission line in star forming galaxies (Lagache, G. et al. 2018).

The structure of this paper is as follows. In section 2 we present the data used. This includes data from ALMA and Hubble across various frequencies. Our method is presented in section 3, outlining our methodology and analysis processes. Results, including extracted spectra are presented in section 5 and discussions are made in section 6, where we estimate over-densities. Our results are summarised in section 7. Throughout this paper we will use $\Omega_M = 0.30$, $\Omega_\Lambda = 0.70$ and $H_0 = 70$km s^{-1} Mpc^{-1} with a ΛCDM cosmology.

2 DATA

2.1 Sample of Galaxies

2.1.1 CR7

The first of the three LAEs we will be studying is CR7 (COSMOS Redshift 7, (Sobral et al. 2015).) CR7 is located in the constellation of Sextans. With a redshift of $z \approx 6.6$, CR7 is the most luminous Lyα emitter found at a redshift of z>6 and likely resides within a bubble of reionisation. CR7 consists of three UV clumps, and is predicted to be populated with low metallicity stars. (Sobral et al. 2015)CR7

Table 1. Frequency and redshift ranges of the CR7,VR7 and MASOSA data cubes. CR7 is present in spw3, VR7 in spw0 and MASOSA in spw0 of the respective data cubes.

Cube Name	Frequency (GHz)	[CII] Redshift
CR7spw0	232.1-233.9	7.18-7.11
CR7spw1	234.1-235.9	7.11-7.04
CR7spw2	247.1-248.9	6.68-6.62
CR7spw3 (CR7)	249.1-250.9	6.61-6.56
VR7spw0 (VR7)	251.5-253.2	6.56-6.51
VR7spw1	249.5-251.2	6.62-6.57
VR7spw2	236.3-237.8	7.05-6.99
VR7spw3	234.1-235.8	7.12-7.06
MASOSAspw0 (MASOSA)	251.2-252.9	6.57-6.52
MASOSAspw1	249.5-250.9	6.63-6.58
MASOSAspw2	236.3-238.1	7.04-6.99
MASOSAspw3	234.1-236.1	7.11-7.05

may reside in an over-dense region and be the progenitor of a bright cluster galaxy, as (Day et al. 2019) found evidence of an over-density of candidate [CII] emitters around CR7.

2.1.2 VR7

VR7 is a large star forming galaxy with an RA of 22:18:56.36 and a Dec of +00:08:07.32. It has a redshift of z=6.532. While it is not as luminous as CR7, VR7 is still a relatively strong Lyα emitter and is even brighter in [CII] and UV emissions than CR7. (Matthee & Sobral 2019). The area around this galaxy has not yet been looked at, so we aim to catalogue sources in the environment around VR7 and work out whether it is in an overdense region or not.

2.1.3 MASOSA

MASOSA is the third high redshift galaxy that will be looked at in this study and has a RA of 10:01:24.80 and a Declination of +02:31:45.34. With a redshift value of z = 6.541, it too is a strong emitter of Lyα photons. In contrast to CR7 and VR7, MASOSA remains undected with current deep ALMA data, implying a very poor metallicity. It should therefore be noted that MASOSA itself will not appear in the catalogues made during the process detailed in the following report. Much like VR7, the environment in which MASOSA resides has not yet been studied. Therefore, we will use data collected from Sections 2.2, 2.3 and ?? to discover what is surrounding the galaxy and whether or not it is in an overdensity.

2.2 ALMA data

The Atacama Large Millimeter/submillmeter Array (ALMA) is an array of 66 radio telescopes located in the Atacama desert in northern Chile, operating at wavelengths between 0.32mm and 3.6mm. The telescopes have distances ranging from 150m to 16km between them and the main array is formed of 50 antenna acting together as an interferometer. It was this set up that was used to survey the areas surrounding CR7, VR7 and MASOSA.

CR7 was observed by ALMA on 22, 23, 24 May 2016

and 4, 8 November 2016 for a total on source time of 6 hours (program ID #2015.1.00122.S; PI: Sobral). During cycle 3, ALMA band 6 was used (band width 1875 MHz) to detect [CII] emission. Four spectral windows were used, centred at 249.94592, 247.94636, 234.94917 and 232.9496 GHz. (e.g. Day et al. 2019). These windows yielded four different data cubes, named spw0, spw1, spw2 and spw3. Each cube has different frequency and redshift ranges and these values can be found in Table 1. We will use the cube which samples [CII] at the redshift of CR7, spw3 to probe the immediate environment of CR7, and the other data-cubes as average volumes which are unrelated to CR7.

VR7 was observed by ALMA on 23, 24 March 2018 and MASOSA on 6, 7 September 2018, both using 43 of ALMA's antenna in configuration C43-4. Each galaxy was observed 3 times, each for 49 minutes. (program ID #2017.1.01451.S). The phase calibrator for MASOSA was quasar J0948+022 and the atmospheric bandpass and flux calibrator was quasar J1508+0133. The atmospheric bandpass and flux calibrators for VR7 were quasars J2148+0657 and J2226+0052 respectively (Matthee & Sobral 2019). Similarly to CR7, four spectral windows were used to observe the galaxies. Two of these windows were centred around a frequency of 252Hz, and two around 235Hz. The full range of frequencies for each spectral window for each galaxy can be seen in the table 1.

Each cube is then in turn split into 60 slices. It is this data that formed the basis of this study; it will be used to identify candidate galaxies and extract their spectra for analysis.

For each of these direct data cubes, a corresponding cube containing primary beam corrected data (PBCorrected) is used for analysis of detections found in the ALMA data. PBCorrected data compensates for the increased signal and reduced noise towards the centre of the field of view using a multiplication factor that increases towards the edge of the images to compensate for the reduced flux.

2.3   Data from the Hubble Space Telescope

Each of the galaxies were also covered by the Hubble Space Telescope (HST), and the data will be used throughout the analysis. Two different infrared filters were used, the first being the F160W filter. This operates between 1.45-1.75μm. The second filter, F110W, operates in a lower wavelength band, between 0.90-1.40μm and is part of the Wide-Field Camera 3 (WFC3) on the HST. The exposure time used to collect the data was 2.6ks for both of these filters. This data will primarily be used to help confirm the existence and nature of the potential galaxies found in Section 3.

3   METHODOLOGY

3.1   Cube collapse

3.1.1   Data Cube Collapse

For much of the analysis performed on the data it was most convenient to split each of the data cubes into a number of slices. The slicing of the data cubes was performed primarily to facilitate the use of the software SExtractor (Bertin & Arnouts 1996) .

Prior to the employment of both a manual examination process and the use of SExtractor for production of a candidate catalogue, the data cubes (further explained in 3.3), are sliced into slices of double the initial velocity width. From each initial data cube, containing 60 slices, 28 slices are obtained following the re-slicing. It is useful to double the velocity width of each slice as any genuine source is likely to be detectable in more than one element.

When collapsing the PBCorrected data cubes the original velocity width is retained, this collapse simply removes the empty data slices, resulting in 57 slices of each original cube being retained. This data is used primarily in obtaining the spectra for candidates (further explored in 4.1).

3.1.2   Inverted Data

In order to interpret any detections and their significance, we have performed the main analysis on inverted data-sets. Inverted forms of the collapsed data cubes (as in 3.1.1) were generated specifically for this. This simple inversion was performed by multiplying each data value by negative one. Within the original data any negative flux must correspond to noise, therefore with use of this inverted data in the catalogue generation processes, an estimation for the number of false positive detections is obtained through the assumption that positive and negative noise is comparable. Any excess of positive detections to false positive detections suggests the presence of genuine emission lines.

3.2   Manual Cataloguing

3.2.1   Noise Estimation For Each Data Cube

Across each of the data cubes an estimation for the noise was calculated using the standard deviation of the flux through 3000 regions of the beam size (0.7"x0.7") with randomly generated centre coordinates per data slice (as generated in 3.1.1).

Following the determination of noise throughout each data cube a manual inspection of the data and the inverted data began to find candidates for an initial catalogue of sources.

3.2.2   Cataloguing

The process of manual cataloguing involved counting each 'bright spot' visible in the collapsed data cubes. These bright spots represented the brightest sources in the positive data and in the negative. In order to do this, the cubes were opened in DS9 and the scale parameters adjusted to improve the contrast and eliminate as much noise as possible (Scale parameter values of 0.0002 and 0.0004 were used). Once these parameters were set, any white spot on the black background was counted on each slice. This process enabled us to locate CR7 and VR7. CR7 was found in the data cube CR7spw3 (at a redshift of approximately 6.6), and VR7 can be seen in VR7spw0 (at a redshift of approximately 6.5). MASOSA is found in MASOSAspw0. However, it is not detectable at the wavelengths we are using to observe the galaxy, so is not included in the numbers in Table 3. After this had been done, the difference between the total number of positive and negative sources for each cube and galaxy

Table 2. The method steps taken to reduce the number of sources from the original catalogues to the final amount of 25 suitable sources as discussed later.

Step	Description	Conditions	Number of remaining sources
0	ALMA selected for cataloging	Sourced from Sobral	12613
1a	Noise reductions -MASOSA	Signal $> 3\sigma$	3180
1b	Noise reductions -CR7	Signal $> 3\sigma$	3722
1c	Noise reductions -VR7	Signal $> 3\sigma$	3867
2a	Deep Cleaning -MASOSA	Distance from center <380 pixels & Distance between sources > 5 pixels	1683
2b	Deep Cleaning -CR7	Distance from center <380 pixels & Distance between sources > 5 pixels	1583
2c	Deep Cleaning -VR7	Distance from center <380 pixels & Distance between sources > 5 pixels	1608
3a	Signal/Noise cutoff purity -MASOSA	S/N > 4.0	81
3b	Signal/Noise cutoff purity -CR7	S/N > 3.9	172
3c	Signal/Noise cutoff purity -VR7	S/N > 3.9	162
4a	Signal/Noise cutoff -MASOSA	S/N > 4.6	10
4b	Signal/Noise cutoff -CR7	S/N > 4.65	17
4c	Signal/Noise cutoff -VR7	S/N > 4.4	35
5	Signal/Noise overdensity	known good source	25
6a	Luminosity Bin -Best - MASOSA	known good source	5
6b	Luminosity Bin -Best - CR7	known good source	9
6c	Luminosity Bin -Best - VR7	known good source	11

Cube	Positive	Negative	Excess
CR7-spw0	1135	1192	-57
CR7-spw1	1307	1216	91
CR7-spw2	664	677	-13
CR7-spw3 (CR7)	795	781	14
CR7-All	3901	3866	35
VR7-spw0 (VR7)	1059	952	107
VR7-spw1	919	967	-48
VR7-spw2	921	828	93
VR7-spw3	1248	1345	-97
VR7-All	4147	4097	55
MASOSA-spw0 (MASOSA)	1460	1433	27
MASOSA-spw1	992	751	241
MASOSA-spw2	760	882	-22
MASOSA-spw3	1353	1348	5
MASOSA-All	4565	4414	151

Table 3. Number of positive and negative detection for CR7, VR7 and MASOSA from manual analysis of the datacube; not accounting for duplicates.

in total were calculated. The totals for CR7, VR7 and MASOSA can be seen in table 3. Any duplicates were not accounted for in this method; as the purpose of manual calibration was to provide an initial suggestion of the quantity of galaxies that reside around each source. This data was then used to check the automated galaxy detection (3.3) results as well as to calibrate its settings, as we needed to decide upon the minimum number of pixels which had to be at the threshold value in order for SExtractor to detect the source.

3.3   Automated Catalogue Generation

3.3.1   Utilising a SExtractor Wrapper

To automate the process of creating a catalogue of detection points a process was created that was able to repeatedly run the software SExtractor (Bertin & Arnouts 1996) over each of the data slices (3.1.1) and the inverted data slices (3.1.2). Python code was used for this purpose and the wrapper module Source Extractor Wrapper for Python (generally known as 'sewpy') was employed (Bertin & Arnouts sew). 'sewpy' was utilised over other wrapper modules available due to its simple interaction with the 'astropy' (Robitaille et al. 2013) modules table functions. The use of the these tables facilitated the later analysis of detections. It should be noted that, although SExtractor has the ability to output fluxes for detections, SExtractor was only used to locate candidate points to be added to the catalogues.

3.3.2   SExtractor Parameters

While conducting the search for sources using the automated SExtractor method the survey employed parameters as defined by a configuration file that was maintained throughout all of the data. Following the recommendations of Holwerda (2005) much of the default config file was retained, primarily the activation of the software's 'CLEANING' function which attempts to reduce the bias on a source due to the effect of its neighbours.

A background area, BACK-SIZE in SExtractor, determines the area over which the average background reading was found around each detection. A value was chosen as 64 for the BACK-SIZE based upon information and recommendations from (Holwerda 2005) This is chosen to allow the software to best take the influence of gradients in the background noise across a detector while maintaining a large enough area that each source should not contribute strongly.

A minimum area, DETECT-MINAREA in SExtractor,

Table 4. The signal to noise cutoff points for CR7, VR7 and MA-SOSA.Where CR7-spw3, VR7-spw0 and MASOSA-spw0 contain the subject galaxy.

Subject	S/N Cutoff
CR7-spw0	N/A
CR7-spw1	4.6
CR7-spw2	4.6
CR7-spw3	4.6
VR7-spw0	4.5
VR7-spw1	4.85
VR7-spw2	4.3
VR7-spw3	4.6
MASOSA-spw0	N/A
MASOSA-spw1	N/A
MASOSA-spw2	N/A
MASOSA-spw3	N/A

of 20 pixels for a detection was chosen to eliminate the detection of individual pixel noise. This was tested by counting the minimum number of pixels required for us to manually identify a source and by comparing manually observed sources to the list generated by SExtractor. Higher values were shown to eliminate detections that were found in manual detections and lower values began to identify a large number of less-significant points that were easily identifiable as noise.

A threshold for the detections of 3 standard deviations was used to leave only the sources with the greatest significance in the SExtractor generated catalogues. This threshold was chosen to remove the least significant detections while maintaining a large data set for later analysis.

3.4 The Refinement of Detection Catalogues

The catalogue produced through the automated SExtractor method is expected to contain many false positives. Further analysis on the candidates is performed using operations on fluxes taken from the PB-corrected data cubes. The PB-corrected data cubes have a smaller region that contains data than the raw data sets, it is therefore necessary to remove candidates that do not lie within the shared region. Conveniently the region on a slice from the PB-corrected data is circular about the central coordinates, it was therefore possible to simply calculate the distance a detection was from the centre and compare this to the radius of the PB-corrected data region. Some sources within the initial catalogue are extremely close to one another and upon inspection many of these 'duplications' in the catalogue were seen to be a single region of brightness, with pixels of negative noise splitting the area. To remove these duplications a manual inspection of the coordinates was performed and sources within five pixels of one another on the same data slice were removed with the priority in retaining a point given to those with a greater detected signal.

3.4.1 MASOSA Bright-spot

Within the MASOSA data cubes we found a continuum-bright sub-millimetre galaxy which we will refer to as the Masosa Bright-Spot, MBS. As discussed in Section 3.4 we
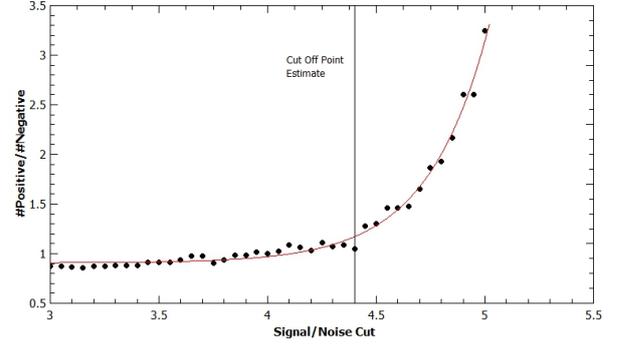
Figure 1. The ratio of positive to negative sources above specific signal to noise %cut off points for the entire VR7 data set.
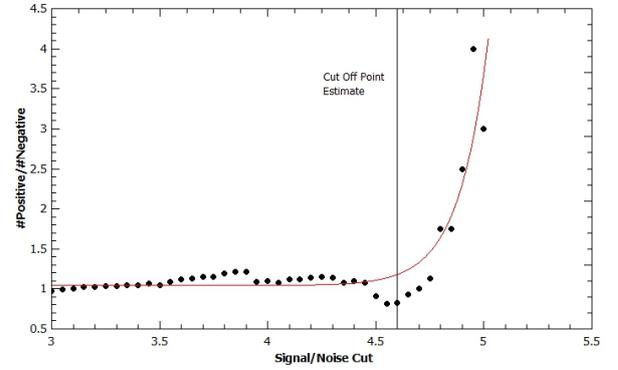
Figure 2. The ratio of positive to negative sources above specific signal to noise %cut off points for the entire MASOSA data set. Here the MASOSA bright spot is %excluded.

removed duplicates from the data, for the MBS however we created a set of data for which any detection of the MBS within it was removed and a set of data with those detections left in. The MBS was visually inspected and no emission lines were found: only continuum.

3.5 Signal to noise calculation

Similarly to the method for calculation of noise conducted prior to the manual cataloguing process (see section 3.2.1) the calculation of signal to noise ratio utilises the measurement of flux through apertures on a data slice. On the direct data slice corresponding to the slide where the initial detection was made an aperture is placed with a size equal to the beam size (0.7"x0.7") of the telescope. This aperture is chosen as it is the minimum size for a point source detectable with that telescope. The flux is taken through this aperture and used to calculate the signal to noise. The value for noise used in the signal to noise ratio for a source is found as the standard deviation of the flux through 200 apertures spread randomly across the data slice.

3.6 Determining cutoff values for S-N

Using a method similar to that in (Aravena et al. 2016), we were able to establish cutoff points for selecting our candi-

Cube	Positive	Negative	Excess	Purity
CR7-spw0	4	4	0	0%
CR7-spw1	4	3	1	25%
CR7-spw2	4	4	0	0%
CR7-spw3	4	3	1	25%
CR7-All	16	14	2	13%
VR7-spw0	15	9	6	40%
VR7-spw1	8	11	-3	N/A
VR7-spw2	7	4	3	43%
VR7-spw3	6	8	-2	N/A
VR7-All	36	32	4	11%
MASOSA-spw0	32	5	25	84%
MASOSA-spw1	31	9	22	71%
MASOSA-spw2	23	4	19	83%
MASOSA-spw3	55	6	49	89%
MASOSA-All	141	24	115	83%

Table 5. The positive, negative number of sources as well as purities for each of the data cube for CR7, VR7 and MASOSA

dates. This involved using the ratios calculated in 3.5 and identifying how many positive and negative sources were detected above each ratio. It is expected for most signal to noise cuts that the ratio of positive to negative sources remains close to 1, therefore the point at which this ratio begins to increase can be considered the point above which sources may be significant. For CR7, the plot produced showed a less obvious trend compared to VR7 and MASOSA. We show examples in Figures 1 and 2.

3.7 Purity

Again taking inspiration from (Aravena et al. 2016), another method in quantifying the contamination of our catalogues is to calculate "Purity" or "Fidelity". This uses the ratio of negative to positive sources, above a S/N cutoff point; in this case 4.65σ for CR7; 4.4σ for VR7 and 4.6σ for MASOSA, as shown in Equation 1.

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

From Table 5 we can see that for MASOSA, using the cutoff we obtain a purity of 84% on SPW3 suggesting significant genuine emission, in agreement with Figure 6. Further for MASOSA, with cubes spw1,spw2 and spw3 getting 71%,83% and 89% respectively, this again would further suggest likely emissions at the high S/N ratios; however, this purity is most likely due to the MBS "Masosa bright spot" adding a significant amount of high S/N sources in the data cubes, creating a positive bias, even though the source causing the MBS is not a galaxy nearby to the early LAE, MASOSA.

4 DATA ANALYSIS

4.1 Spectral Graphing

For the candidates that remained above the signal-to-noise cut off point (see section 3.6) it was possible to distinguish between CII emitting galaxies and strongly positive areas of noise through comparison of the spectra detected. It was expected that real emission lines should have a defined peak in its spectrum which would fit a Gaussian curve whilst noise would not exhibit this quality. By producing graphs of emission spectra, we are narrowing down the potential emission lines we have found and can begin to identify potential [CII] emitters.

To generate the spectra for each data point in the refined catalogue a process similar to that detailed in section 3.5 is used. For all the data slices in a data cube in which a source was found, an aperture is applied at the coordinates of the source and 200 randomly placed apertures are spread over the slice, the flux of the source aperture is then plotted against the noise found from the standard deviation of the flux for the apertures on each slice.

4.1.1 Visual Inspection and Fitting of Gaussian Curves

Following the generation of the spectra plots a process of manual examination began. The manual examination identified candidates that had a spectrum which showed approximately Gaussian characteristics. For these candidates a Gaussian curve was then fitted for the points that best exhibited the behaviour with QtiPlot's fitwizard tools.

4.2 Analysis of ALMA/Hubble images

Any true [CII] emitters will be at redshifts $z > 6$, meaning any optical light would be redshifted to wavelengths undetectable by the F814W filter. Therefore any sources visible with this filter should be reconsidered as potential CO emitters at a lower redshift. The majority of our candidates show little signal in these filters and are therefore unlikely to be CO emitters.

Our most likely candidates will show detections in Hubble's infrared filters (F110W and F160W), this is because the optical light from these galaxies is shifted into infrared wavelengths. From our final candidates presented in appendix figures 2, 4 and 6 there are several clear sources that are distinct at these wavelengths. CR7-spw0-02, VR7-spw2-02 and MASOSA-spw1-01 show clear signals which suggests they are our most robust candidates, whilst several other candidates show weaker, less central emissions. These candidates are still worth analysing further given the size of the cutouts and the precision of DS9. Those worth further investigation are: CR7-spw0-01, CR7-spw2-01, CR7-spw3-01, VR7-spw2-02 and MASOSA-spw1-01. This leaves us with 3 potential [CII] candidates, 3 CO candidates and over 15 candidates that would require more detailed analysis.

4.3 Luminosity bins

An alternative method we have used to analyse the positive to negative excess in our catalogues of data, is to use luminosity bins. This secondary method allows us to calculate
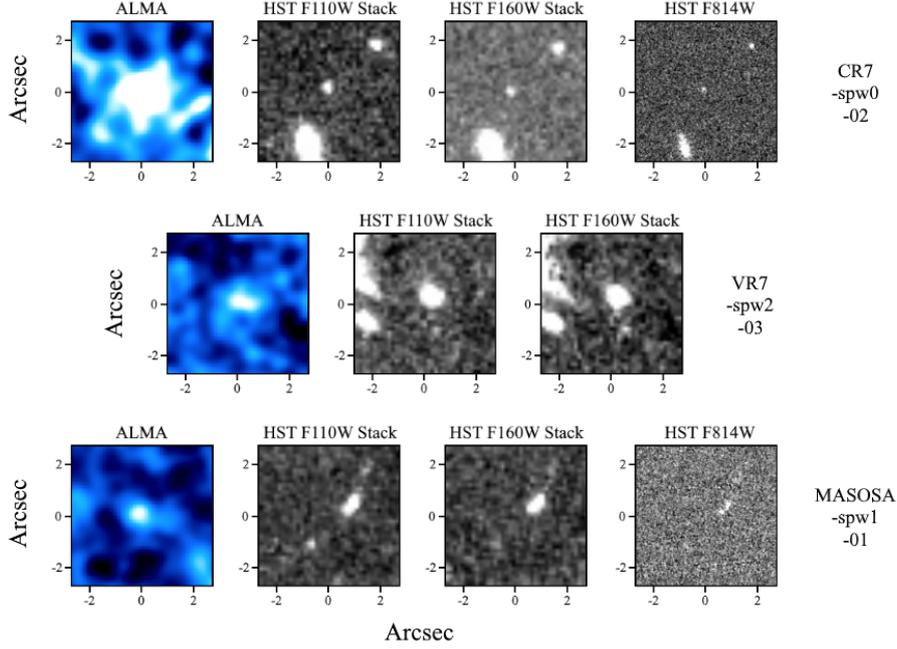
Figure 3. The ALMA/Hubble data for our most robust candidates, presented as $5 \times 5''$ cutouts.
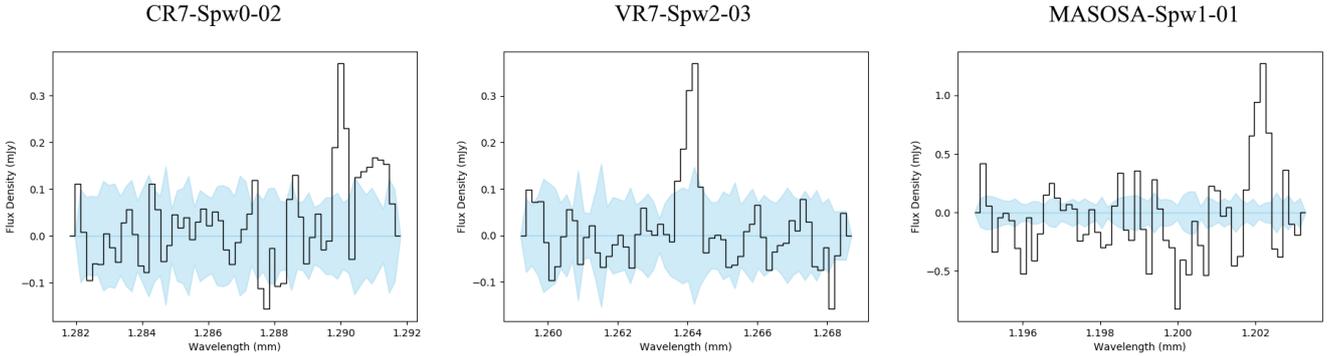
Figure 4. The emission spectra of our most robust candidates, showing clear peaks above the noise background (blue).

a luminosity function for each galaxies environment, as well as the reference volumes to calculate relative over-densities.

We have used luminosity here rather than flux due to the more absolute nature of using luminosity. Further the volume of the datacubes has been calculated by taking the area of the sky being measured in each data slice (approximately $500 arcsec^2$) and the difference between the maximum and minimum redshift values for that particular data cube

Firstly we calculated the luminosities for our catalogues by using the equation (Where luminosity distances in this equation were calculated using an online tool (Wright 2006)):

$$L = f \times 4\pi \times DL^2 \qquad (2)$$

This method has been used in two variations:

• A Statistical Method, similarly to the use of cut off points in signal to noise (see section 3.5) with purity, we calculate the excess between the positive and negative sources

(sources found in the inverted data cubes) within a range of luminosity. We have used bins of $10^7 L_o$ width, removed outliers that have negative luminosity and above the luminosity of the subject galaxy.

• A Direct method compares the number of acceptable sources that have made it through all of the data reduction to the normalised number of sources in other data cube. Using this number of sources (which do not contain any number of false sources) and again dividing by the $Log_{[}10]$(bin width) ($7.5L$) and the volume of the data cube ($100mpc^3$) produces a second luminosity function.

5 RESULTS

5.1 Notable Detections

Beyond the candidates for likely [CII] emitters there were some notable detections made and excluded from the final
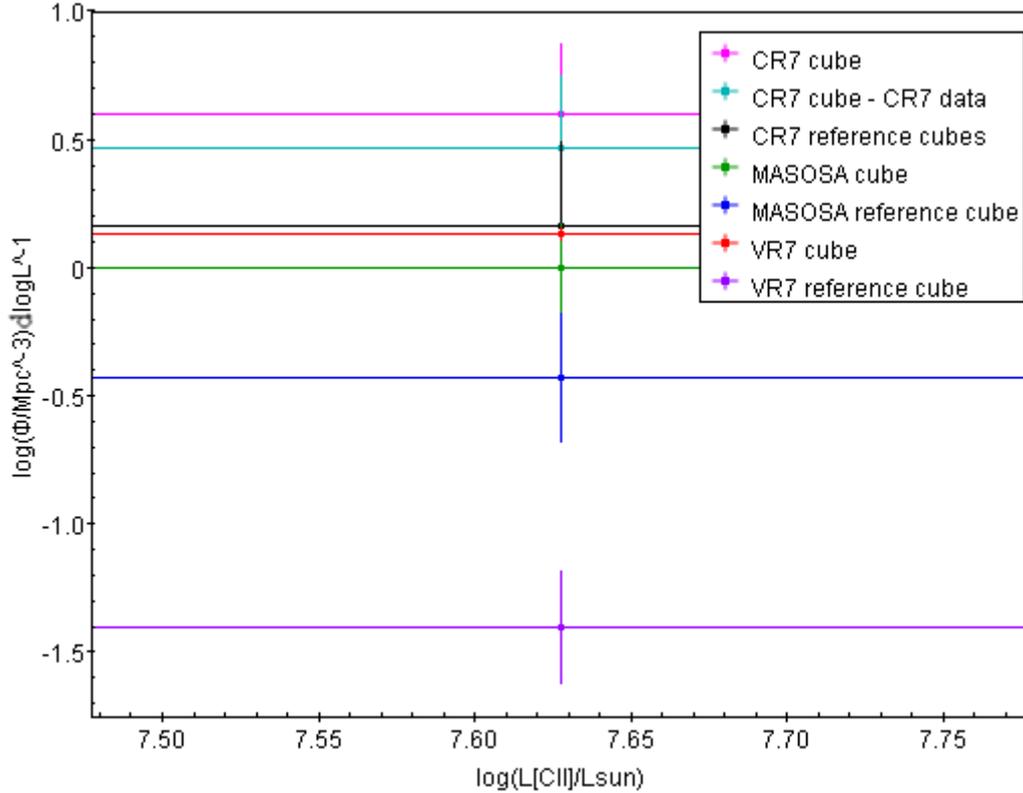
Figure 5. The luminosity functions for CR7, VR7 and MASOSA compared to the reference volumes surrounding the subject galaxies

catalogue. The MASOSA bright spot (see section 3.4.1) has shown a significant luminosity through all the data cubes of MASOSA. Located at approximately R.A. 10:01:24.708 and Dec. +2:31:32.717 this source was not thoroughly investigated but an early interpretation of the data led to an unproven identification of this source as a dusty sub-millimeter galaxy. A candidate believed to be a nearby object emitting a [CO] spectra was found within the VR7 data cubes at R.A. 22:18:56.103 and Dec. 0:08:14.846.

5.2 The high-z luminosity function for [CII] emitters

The luminosity function quantifies the number density of sources as a function of luminosity. There are currently few results for luminosity functions at high redshift, except for in the [CII] survey performed by (Aravena et al. 2016) where a limiting luminosity of $L_{CII} = (1.6 - 2.5) \times 10^8 L_o$ was found. The [CII] luminosities for our candidate galaxies are all below this limit which suggests this estimate is reasonable. Various other groups have previously predicted the [CII] luminosity function at specific redshifts (e.g. Popping et al. 2016). These results were produced by comparing two theoretical models for galaxy formation and energy transfer. The results present in the referenced paper show a declining luminosity function that is sufficiently negative when compared to our results, even when potential [CO] contaminants are accounted for.

It is also possible to compare luminosity functions with and without the target galaxies present. The results in fig-

ure 5 show that CR7 has a substantially higher luminosity function even when CR7 itself is excluded (light blue). The data for MASOSA (green) and VR7 (red) already have the target galaxies removed and are still substantially brighter than the results from (e.g. Popping et al. 2016). This initially suggests overdensities for all three sources despite the central galaxies being removed.

5.3 The density of CR7, VR7 and MASOSA

5.3.1 Results using luminosity bins

The method for use of luminosity bins is discussed in section 4.3. Figure 5 suggests that CR7 and VR7 lie in extremely overdense regions (due to the large disparity between the reference values and the luminosity functions of the cubes containing the observed galaxies). In comparison the overdensity of the region around the galaxies (as seen in Figure6) suggests that MASOSA is in an overdense region far more populated that expected, whilst CR7 resides in an area of lower overdensity than expected. These over densities were found as $2^+_-0.28$ and $4.5^+_-0.18$ respectively. VR7 is found to be in a area of suggested regular density $1.125^+_-0.20$.

6   DISCUSSION OF RESULTS

6.1 Summary

Our results have been generated from observing the environments around 3 high redshift LAEs within the epoch of reionisation: CR7, VR7 and MASOSA. This study aims to
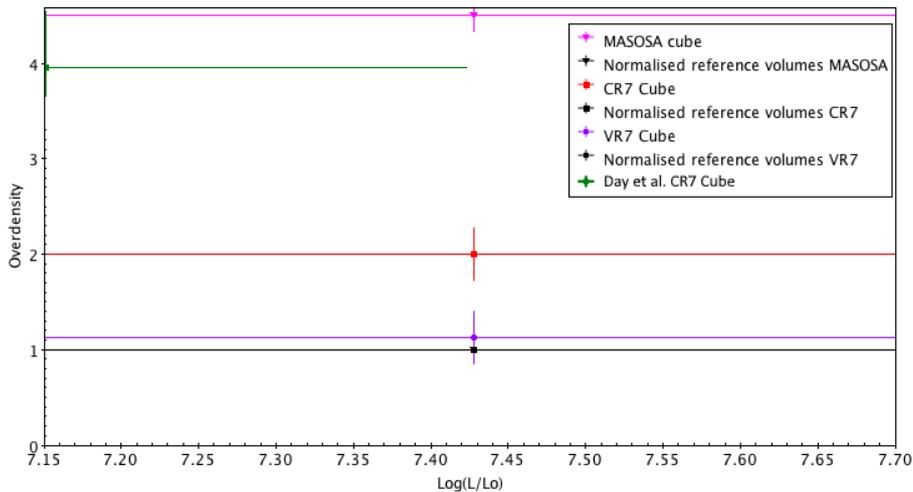
Figure 6. The relative overdensities of CR7, VR7 and MASOSA : in the higher luminosity bins the excess between positive and negative sources are considered to be real sources.

quantify whether a significant overdensity of galaxies in the areas surrounding CR7, VR7 and MASOSA aids the reionisation process (e.g. Faisst 2016).

The data used has been collected by ALMA and Hubble, with radio data provided by ALMA and optical/infrared provided by the HST. This data has been used to determine potential candidates for [CII] emitting galaxies and to attempt to quantify the overdensity of CR7, VR7 and MASOSA. Previous results (e.g. Day et al. 2019) show a significant overdensity for CR7, a result we attempt to replicate.

By calculating emission spectra (Figure 4) and comparing data from ALMA/Hubble (Figure 3, we have identified 3 robust candidates worth considering as [CII] emitters. They are CR7-spw0-02 ("Twinings"), VR7-spw2-03 ("A little grey") and MASOSA-SPW1-01 ("MCF1"). There are many other candidates stated in table 2.1 that may also warrant further investigation. In order to determine the nature of these sources, further observations are required.

Our initial overdensity estimations involved comparing the number of positive sources in the data cube containing each galaxy to the combined data cubes surrounding each galaxy. The number of sources was calculated by finding the difference between positive and negative sources first identified in our manual detections. This method produced largely inconsistent results. The reasons for this discrepancy are that a large amount of faint but positive sources were screened out of the data early in the data reduction process. This often left an excess of negative sources in our data which caused problems with using this method. We have also calculated purity for each spw file using the method outlined in Aravena et al. (2016), but the results were largely unhelpful as many of the data cubes showed negative excesses, meaning purity could not be calculated. We thereby switched to using a more reliable method to calculate our overdensities by finding luminosities and luminosity functions.

By selecting a large bin width, our results are likely to be smaller than expected. This holds true when comparing our results to Day et al. (2019) which uses a slightly smaller bin than our results. For CR7 they produce an overdensity

of $3.95^{+0.91}_{-0.76}$, whilst this paper produces a result of 2.0 ± 0.28, a percentage error of 49.4%. We also produced results for MASOSA and VR7 at 4.5 ± 0.18 and 1.125 ± 0.20 respectively. Our extreme overdensity for MASOSA can be accounted for as our positive candidates were present in two cubes only, one of which is the cube MASOSA is present in. This skews the normalised values and leads to a large overdensity.

We also qualitatively compared the luminosity functions produced in this study by those theorised in Popping et al. (2017). This displayed that the luminosity functions for CR7, VR7 and MASOSA are substantially higher than expected, providing significant evidence for overdensities around each galaxies environment.

6.2   Impacts on Reionisation

The data we have collected suggests that the galaxies CR7, VR7 and MASOSA all are all found within regions of varying overdensity. (Castellano et al. 2016) suggests that the overlapping of ionised bubbles coincides with the overdensity of galaxies and implies a build up of galaxy clusters at high redshifts ($z \approx 7$). Our results, including the existence of the three robust sources CR7-spw0-02 ("Twinings"), VR7-spw2-03 ("A little grey") and MASOSA-SPW1-01 ("MCF1") support this picture. This therefore could explain early reionisation of the IGM, especially when considering MASOSA which has a particularly high overdensity.

6.3   Comparison to other papers

In comparison to other known papers, such as Day et al. (2019), our values for the luminosity functions are comparatively similar with values ranging from 0.6 to 0.4 for the CR7 luminosity function with a comparable reference value. If comparing to CR7 (overdensity of 2 ± 0.28), MASOSA (overdensity of 4.5 ± 0.18) appears to lie in a significantly overdense region whilst VR7 (overdensity of 1.125 ± 0.20)lies in a slighty more underdense region. It is unknown whether these values are accurate as we have little data to compare

to, however the suggestion that each galaxy is within an overdense region is supported by theory.

Our discrepancies in CR7 are mostly due to the configuration files used by SExtractor early in the catalogue process. Our configuration file had stricter conditions than the ones used in Day et al. (2019), meaning we lost many potential candidates early on. This coupled with the fact our luminosity bins were wider would ultimately produce a lower value for the overdensity as is seen in this paper.

7 CONCLUSIONS

In this project, we have analysed ALMA data for the three distant LAEs CR7, VR7 and MASOSA. We first reduced the data by merging adjacent slices to help with a blind manual survey performed on all 3 galaxies and the environments around them. Initial counts of sources allowed for statistical analysis which suggested how many useful candidates we were likely to find.

By introducing noise reduction, we have used SExtractor to identify a large amount of potential candidates, which have been reduced further to only include those above certain signal to noise ratios. Emission spectra were found for these candidates with the aim of identifying potential [CII] peaks that were expected from any true galaxies.

Our final candidates were observed using data from both ALMA and Hubble. Our best galaxy candidates were expected to be detected in the infrared Hubble filters (F110W, F160W) whilst any potential CO emitters would show up in the optical (F814W).

• The visualisation of the most probable [CII] emitting galaxy candidates for CR7, VR7 and MASOSA. This includes 8 candidates for CR7, 11 for VR7 and 5 for MASOSA. From these candidates, 4 from CR7, 2 from VR7 and 1 from MASOSA show significant signals in Hubble data. The large number of candidates for CR7 and VR7 suggests a possible formation of galactic clusters around these galaxies.

• Calculations of the densities of each galaxies environment using purity and statistical methods. These were found to be 1.07 for CR7, 0.79 for VR7 and 3.00 for MASOSA. The value for CR7 is in disagreement with the known value of $3.95^{+0.91}_{-0.76}$ which suggests the results found for VR7 and MASOSA may also be inaccurate.

• Calculations of the densities by binning luminosities and creating luminosity functions. This created overdensities of $2.0^+_-0.28$, $4.5^+_-0.18$ and $1.125^+_-0.20$ for CR7, MASOSA and VR7 respectively. These results are more reliable and suggest overdensities for each galaxy.

Further work could include re-analysing the environments around each galaxy to confirm or disprove the densities stated. Similar experiments could also be performed on other high redshift LAEs (e.g. Hu et al. 2017; Krug et al. 2012) to provide further data for the processes governing the epoch of reionisation. This could include how the environments around bright LAEs change the ionisation rate. It would also be enlightening to further explore the [CII] luminosity functions at high redshifts. Given the overdensitites present in our results it may suggest an overall higher [CII] luminosity function than predicted.

Considering the imminent launch of the James Webb Space Telescope (JWST), the ability to probe to higher redshifts will improve greatly. This would benefit further studies on both the subjects of this study and other, perhaps more distant LAEs. Being able to reach higher redshifts is imperative to understanding the processes taking place in the early Universe including the epoch of reionisation. This improved data may act as the evidence needed to constrain the redshift at which reionisation began.

REFERENCES

Aravena M., et al., 2016, ApJ, 833, 71
Astropy Collaboration et al., 2013, A&A, 558, A33
Atek H., et al., 2015, Are Ultra-faint Galaxies at z=6-8 Responsible for Cosmic Reionization ? Combined Constraints from the Hubble Frontier Fields Clusters and Parallels (arXiv:1509.06764)
Behrens C., Pallottini A., Ferrara A., Gallerani S., Vallini L., 2019, Monthly Notices of the Royal Astronomical Society, 486, 2197–2209
Bertin E., Arnouts S., 1996, A&AS, 117, 393
Castellano M., et al., 2016, The Astrophysical Journal, 818, L3
Day L., Sobral D., Matthee J., Dodd E., Draper L., 2019, Monthly Notices of the Royal Astronomical Society, 492, 1778–1790
Draper P. W., 2016, GAIA - Graphical Astronomy and Image Analysis Tool, http://star-www.dur.ac.uk/~pdraper/gaia/gaia.html
Faisst A. L., 2016,] 10.3847/0004-637x/829/2/99, 829, 99
Greig B., Mesinger A., Koopmans L. V. E., 2019, Monthly Notices of the Royal Astronomical Society, 491, 1398
Holwerda B. W., 2005, arXiv preprint astro-ph/0512139
Hu W., et al., 2017, The Astrophysical Journal, 845, L16
Hunter J. D., 2007, Computing In Science & Engineering, 9, 90
Jones E., Oliphant T., Peterson P., et al., 2001, SciPy: Open source scientific tools for Python, http://www.scipy.org/
Konno A., et al., 2017, Publications of the Astronomical Society of Japan, 70
Krug H. B., et al., 2012, The Astrophysical Journal, 745, 122
Lagache, G. Cousin, M. Chatzikos, M. 2018, A&A, 609, A130
Lebouteiller V., et al., 2019, Astronomy & Astrophysics, 632, A106
Matthee J., Sobral D., 2019, Unveiling the most luminous Lyman-alpha emitters in the epoch of reionisation (arXiv:1911.04774)

Matthee J., Sobral D., Gronke M., Pezzulli G., Cantalupo S., Röttgering H., Darvish B., Santos S., 2019, Monthly Notices of the Royal Astronomical Society, 492, 1778–1790

Meyer R. A., et al., 2019, The role of galaxies and AGN in reionising the IGM – III : IGM-galaxy cross-correlations at z 6 from 8 quasar fields with DEIMOS and MUSE (arXiv:1912.04314)

Natarajan A., Battaglia N., Trac H., Pen U. L., Loeb A., 2013, ApJ, 776, 82

Ota K., et al., 2017, The Astrophysical Journal, 844, 85

Planck Collaboration et al., 2016, A&A, 596, A108

Popping G., van Kampen E., Decarli R., Spaans M., Somerville R. S., Trager S. C., 2016, Monthly Notices of the Royal Astronomical Society, 461, 93

Popping G., Puglisi A., Norman C. A., 2017, MNRAS, 472, 2315

Robertson B. E., Ellis R. S., Furlanetto S. R., Dunlop J. S., 2015, The Astrophysical Journal, 802, L19

Robitaille T. P., et al., 2013, Astronomy & Astrophysics, 558, A33 DS9

Sobacchi E., Mesinger A., 2015, Monthly Notices of the Royal Astronomical Society, 453, 1843

Sobral D., Matthee J., Darvish B., Schaerer D., Mobasher B., Röttgering H. J. A., Santos S., Hemmati S., 2015, ApJ, 808, 139

Taylor M., 2013, Starlink User Note, 253

Van Der Walt S., Colbert S. C., Varoquaux G., 2011, Computing in Science & Engineering, 13, 22

Wright E. L., 2006, Publications of the Astronomical Society of the Pacific, 118, 1711

Zaroubi S., 2012, Astrophysics and Space Science Library, p. 45–101

A Source Extractor Wrapper for Python

8   APPENDIX: ROBUST CANDIDATES FOR CR7, VR7 AND MASOSA

This paper has been typeset from a TEX/LATEX file prepared by the author.

ID	R.A.	Dec.	CII Flux (cgs)	$log_{10}(L/L_O)$	Redshift
CR7-SPW0-01 "PG Tips"	10:00:58.415	1:48:22.95	2.80e-19	7.64	7.154
CR7-SPW0-02 "Twinings"	10:00:58.202	1:48:23.65	1.85e-19	7.46	7.165
CR7-SPW0-03 "Tetley"	10:00:58.165	1:48:19.25	2.25e-19	7.54	7.117
CR7-SPW1-01 "Typhoo"	10:00:58.282	1:48:21.15	1.71e-19	7.42	7.100
CR7-SPW1-02 "Yorkshire Gold"	10:00:58.022	1:48:08.25	1.92e-19	7.46	7.046
CR7-SPW2-01 "Lipton"	10:00:58.629	1:48:13.50	4.37e-19	7.76	6.663
CR7-SPW3-01 "English Breakfast"	10:00:58.909	1:48:12.40	8.56e-19	8.04	6.598
CR7-SPW3-02 "Earl Grey"	10:00:57.735	1:48:06.30	3.94e-19	7.70	6.573
CR7-SPW3-03 "Chamomile"	10:00:57.702	1:48:01.70	5.76e-19	7.87	6.593
CR7-SPW3-CR7SOURCE1 "CR7 Clump B"	10:00:57.942	1:48:15.75			
CR7-SPW3-CR7SOURCE2 "CR7 Clump B"	10:00:57.955	1:48:15.95			6.589
VR7-SPW0-01 "Some Grey"	22:18:56.784	0:08:02.352	3.59e-19	7.65	6.521
VR7-SPW0-02 "Another Grey"	22:18:56.088	0:08:02.707	3.81e-19	7.681	6.517
VR7-SPW0-03 "Very Much Grey"	22:18:55.803	0:08:22.363	3.23e-19	7.61	6.519
VR7-SPW2-01 "Super Grey"	22:18:57.233	0:08:16.453	5.98e-19	7.95	7.004
VR7-SPW2-02 "Quite Grey"	22:18:57.150	0:08:05.463	2.82e-19	7.63	6.985
VR7-SPW2-03 "A little Grey"	22:18:56.812	0:08:10.750	3.19e-19	7.68	7.001
VR7-SPW2-04 "Extra Grey"	22:18:56.694	0:08:14.545	2.96e-19	7.64	6.974
VR7-SPW2-05 "Elephant Grey"	22:18:55.505	0:08:10.439	1.86e-19	7.44	6.999
VR7-SPW3-01 "Grey Grey"	22:18:56.438	0:08:00.590	1.63e-19	7.40	7.088
VR7-SPW3-02 "Dull Grey"	22:18:56.103	0:08:14.846	1.75e-19	7.43	7.044
VR7-SPW3-03 "Decent Grey"	22:18:55.623	0:07:50.680	8.17e-19	8.09	7.050
VR7-SPW0-VR7SOURCE1 "VR7"	22:18:56.373	0:08:07.070			6.517
VR7-SPW0-VR7SOURCE2 "VR7"	22:18:56.347	0:08:07.220			
MASOSA-SPW0-01 "BENtley"	10:01:25.004	2:31:38.590	5.01e-19	7.80	6.542
MASOSA-SPW0-02 "WALDI7"	10:01:24.983	2:31:38.382	4.03e-19	7.71	6.543
MASOSA-SPW0-03 "HCII"	10:01:24.174	2:31:52.483	4.60e-19	7.76	6.503
MASOSA-SPW1-01 "MCF1"	10:01:25.302	2:31:30.814	9.05e-19	8.07	6.609
MASOSA-SPW1-02 "KAT7"	10:01:24.7734	2:31:49.476	4.33e-19	7.75	6.590

Table 1. A table of the candidates most likely to be CII emitting galaxies from the CR7, VR7 and MASOSA data cubes. This includes data from the photometry performed on each candidate and shows the detections that were attributed to the subject galaxies.
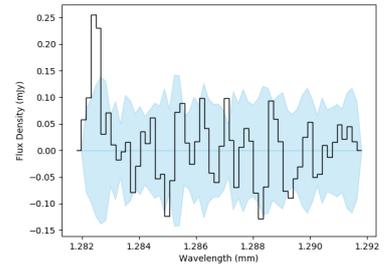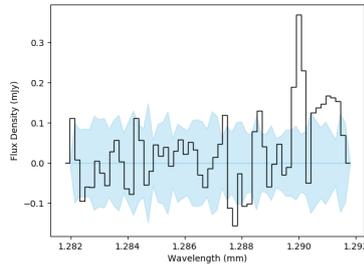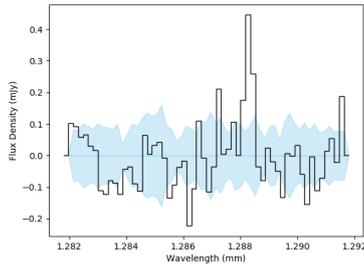
(a) CR7-spw0-01 ; CR7-spw0-02 ; CR7-spw0-03
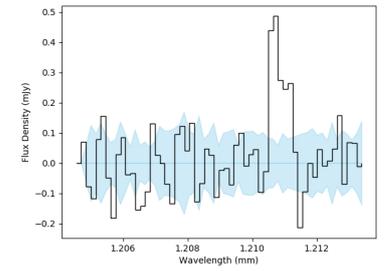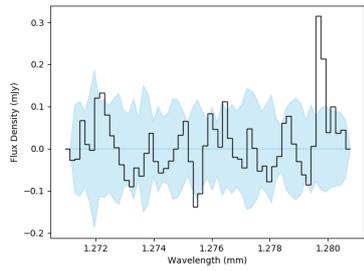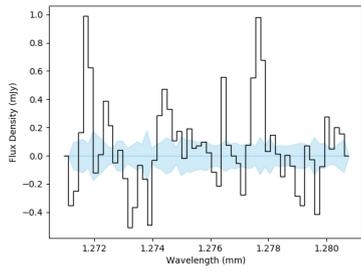
(b) CR7-spw1-01 ; CR7-spw1-02 ; CR7-spw2-01
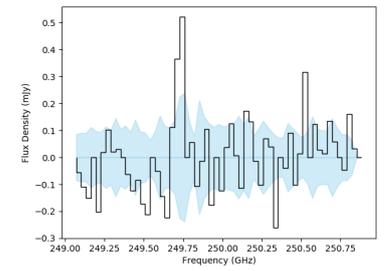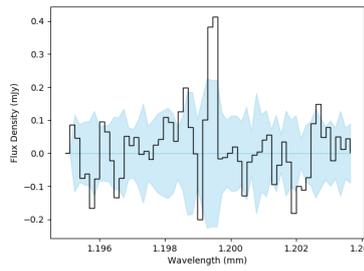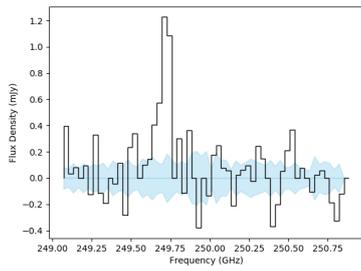
(c) CR7-spw3-01 ; CR7-spw3-02 ; CR7-spw3-03

Figure 1. The best spectra of candidates surrounding CR7, from manual examination of all data cubes.
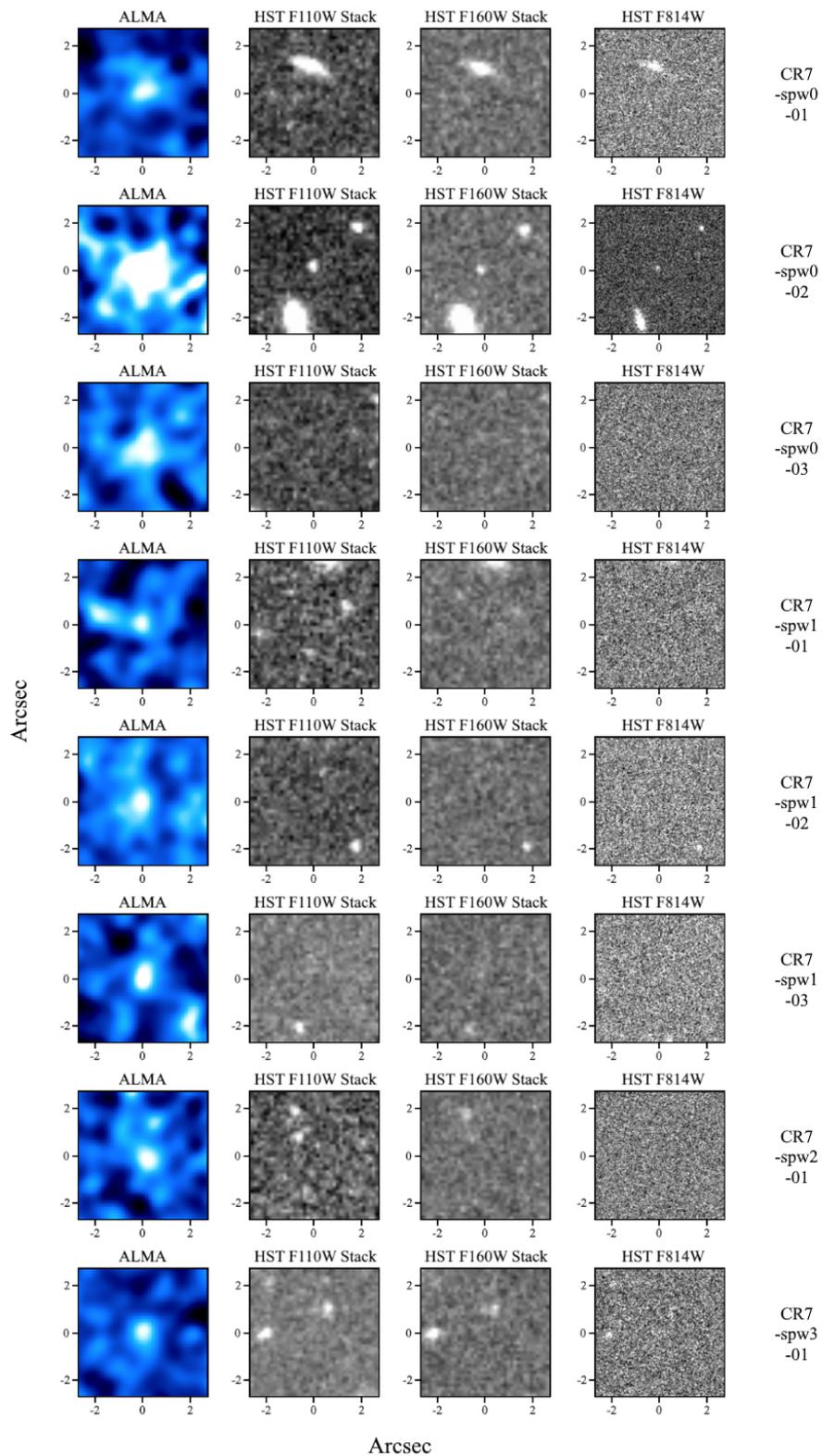
Figure 2. 5x5 arc second cutouts of the most likely galaxy candidates surrounding CR7. These include radio data from ALMA and the Hubble Space Telescope at infrared and optical wavelengths.
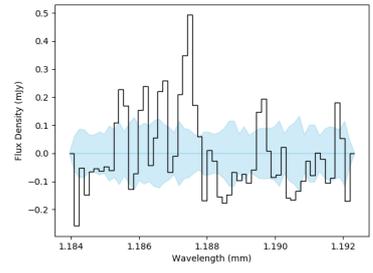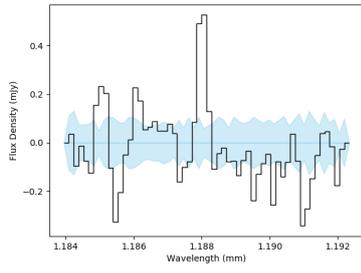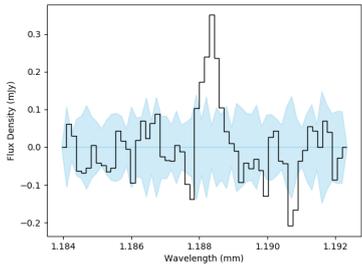
(a) VR7-spw0-01 ; VR7-spw0-02 ; VR7-spw0-03
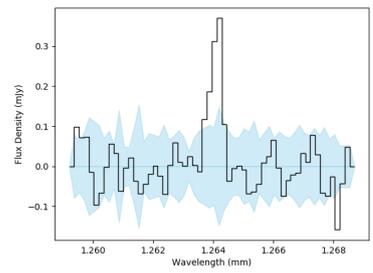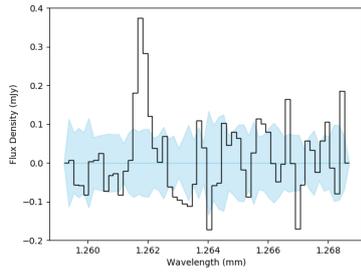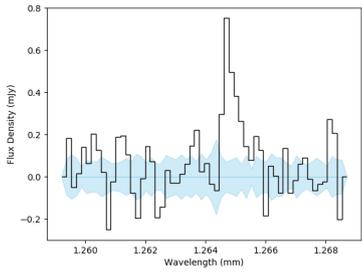
(b) VR7-spw2-01 ; VR7-spw2-02 ; VR7-spw2-03
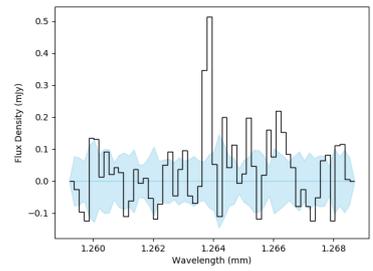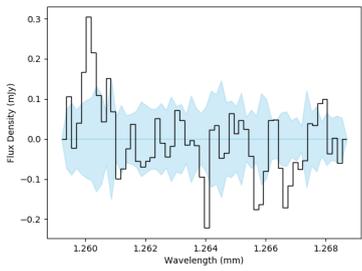
(c) VR7-spw2-04 ; ; VR7-spw2-05
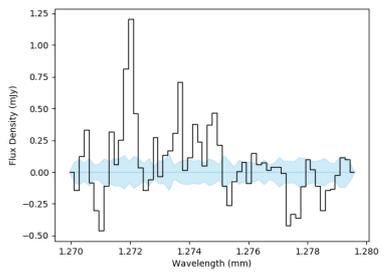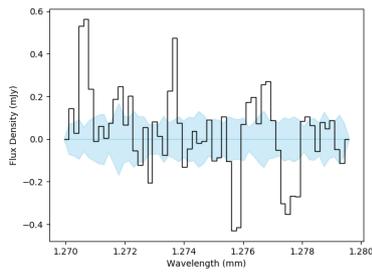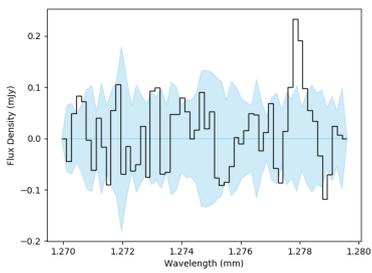
(d) VR7-spw3-01 ; VR7-spw3-02 ; VR7-spw3-03

Figure 3. The best spectra of candidates surrounding VR7, from manual examination of all data cubes.
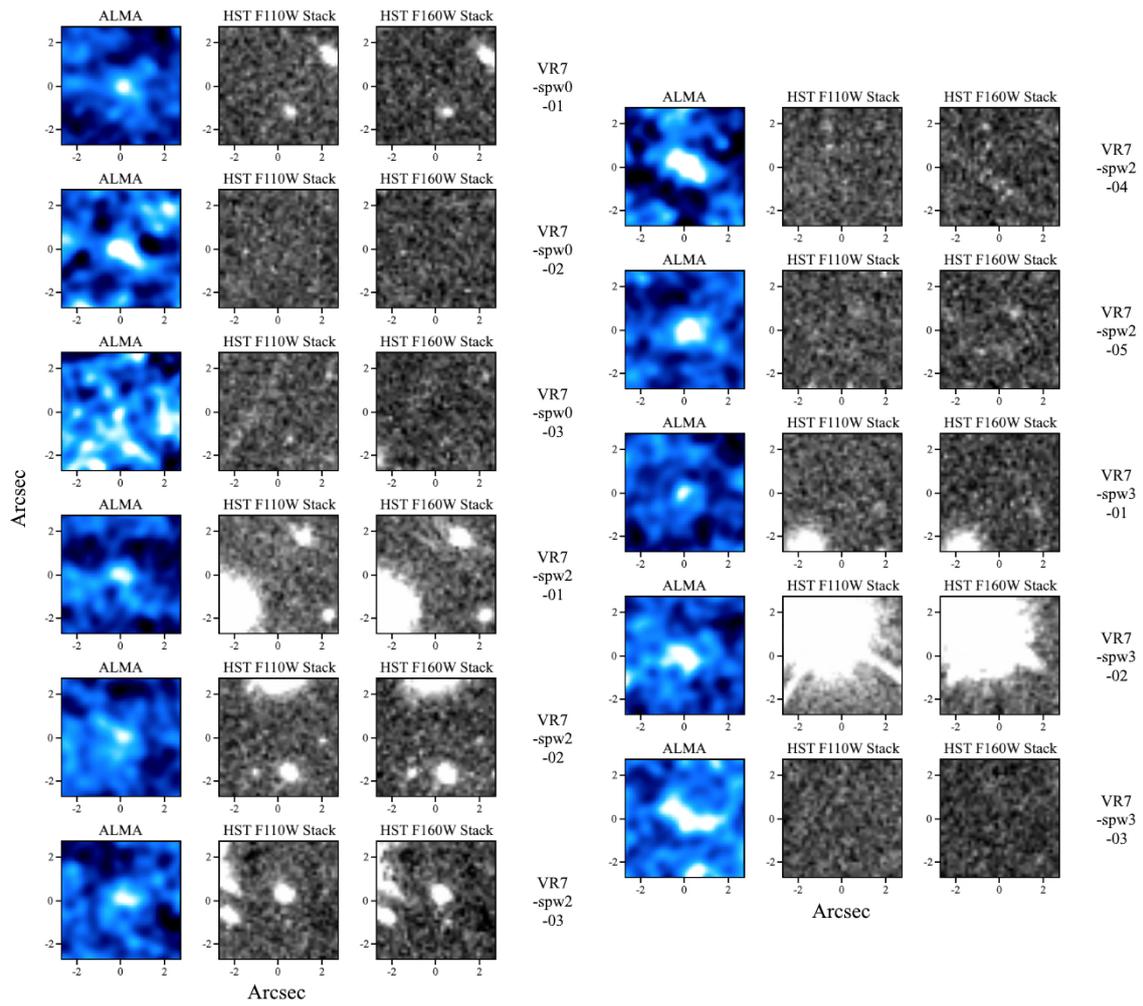
Figure 4. 5x5 arc second cutouts of the most likely galaxy candidates surrounding VR7. These include radio data from ALMA and the Hubble Space Telescope at infrared wavelengths.
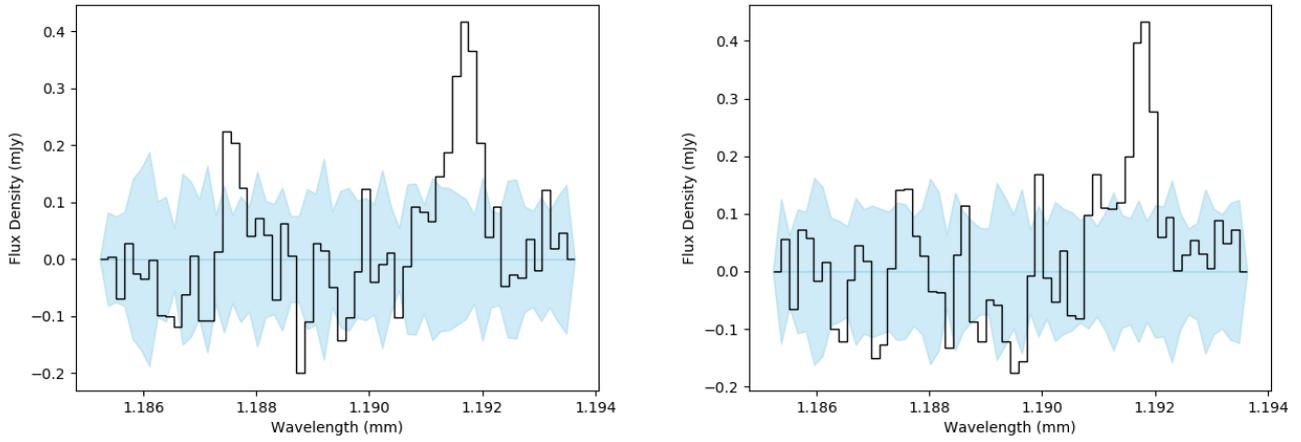
(a) MASOSA-spw0-01 ; MASOSA-spw0-02
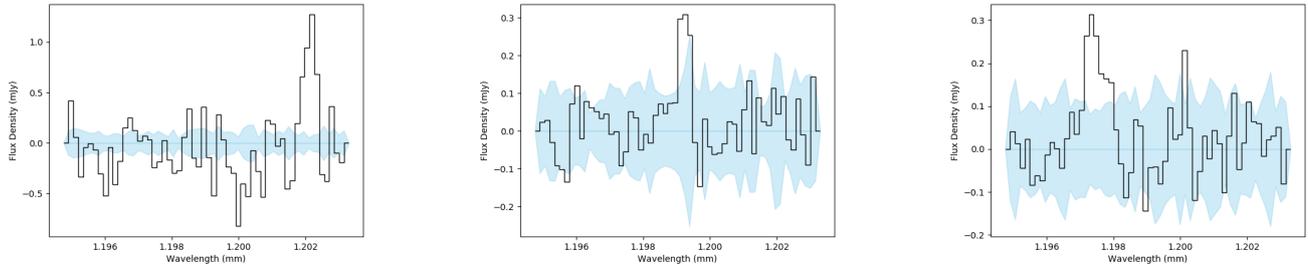
(b) MASOSA-spw0-03 ; MASOSA-spw1-01 ; MASOSA-spw1-02

Figure 5. The best spectra of candidates surrounding MASOSA, from manual examination of all data cubes.
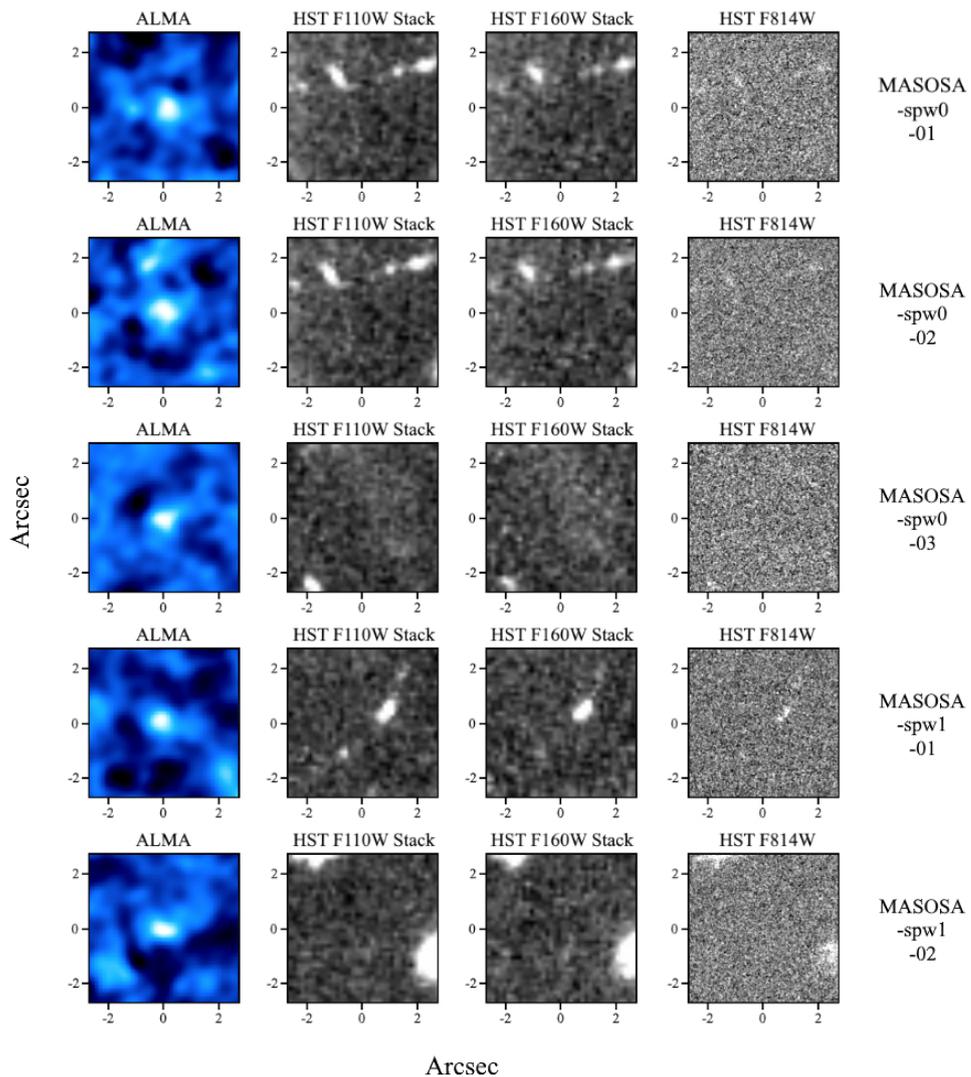
Figure 6. 5x5 arc second cutouts of the most likely galaxy candidates surrounding MASOSA. These include radio data from ALMA and the Hubble Space Telescope at infrared and optical wavelengths.