Publication Year	2020
Acceptance in OA@INAF	2021-12-27T16:34:57Z
Title	þÿ No Evidence for [C II] Halos or High-velocity Outflows Galaxies
Authors	Novak, Mladen; Venemans, Bram P.; Walter, Fabian; Neeleman, Marcel; Kaasinen, Melanie; et al.
DOI	10.3847/1538-4357/abc33f
Handle	http://hdl.handle.net/20.500.12386/31262
Journal	THE ASTROPHYSICAL JOURNAL
Number	904

DRAFT VERSION OCTOBER 29, 2020
Typeset using LaTeX **twocolumn** style in AASTeX62

arXiv:2010.14875v1 [astro-ph.GA] 28 Oct 2020

No evidence for [C II] halos or high-velocity outflows in $z \gtrsim 6$ quasar host galaxies

MLADEN NOVAK,[1] BRAM P. VENEMANS,[1] FABIAN WALTER,[1,2] MARCEL NEELEMAN,[1] MELANIE KAASINEN,[1,3] LICHEN LIANG,[4] ROBERT FELDMANN,[4] EDUARDO BAÑADOS,[1] CHRIS CARILLI,[2] ROBERTO DECARLI,[5] ALYSSA B. DRAKE,[1] XIAOHUI FAN,[6] EMANUELE P. FARINA,[7] CHIARA MAZZUCCHELLI,[8] HANS–WALTER RIX,[1] AND RAN WANG[9]

[1]*Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany*
[2]*National Radio Astronomy Observatory, Pete V. Domenici Array Science Center, P.O. Box O, Socorro, NM 87801, USA*
[3]*Universität Heidelberg, Zentrum für Astronomie, Institut für Theoretische Astrophysik, Albert-Ueberle-Straße 2, D-69120 Heidelberg, Germany*
[4]*Institute for Computational Science, University of Zurich, Zurich CH-8057, Switzerland*
[5]*INAF - Osservatorio di Astrofisica e Scienza dello Spazio, via Gobetti 93/3, I-40129, Bologna, Italy*
[6]*Steward Observatory, The University of Arizona, 933 N. Cherry Ave., Tucson, AZ 85721, USA*
[7]*Max Planck Institut für Astrophysik, Karl–Schwarzschild–Straße 1, D-85748, Garching bei München, Germany*
[8]*European Southern Observatory, Alonso de Córdova 3107, Vitacura, Región Metropolitana, Chile*
[9]*Kavli Institute for Astronomy and Astrophysics, Peking University, Beijing 100871, People's Republic of China*

ABSTRACT

We study the interstellar medium in a sample of 27 high-redshift quasar host galaxies at $z \gtrsim 6$, using the [C II] 158 μm emission line and the underlying dust continuum observed at $\sim$ 1 kpc resolution with ALMA. By performing uv-plane spectral stacking of both the high and low spatial resolution data, we investigate the spatial and velocity extent of gas, and the size of the dust-emitting regions. We find that the average surface brightness profile of both the [C II] and the dust continuum emission can be described by a steep component within a radius of 2 kpc, and a shallower component with a scale length of 2 kpc, detected up to $\sim$10 kpc. The surface brightness of the extended emission drops below $\sim$1% of the peak at radius of $\sim$ 5 kpc, beyond which it constitutes 10 – 20% of the total measured flux density. Although the central component of the dust continuum emission is more compact than that of the [C II] emission, the extended components have equivalent profiles. The observed extended components are consistent with those predicted by hydrodynamical simulations of galaxies with similar infrared luminosities, where the dust emission is powered by star formation. The [C II] spectrum measured in the mean uv-plane stacked data can be described by a single Gaussian, with no observable [C II] broad-line emission (velocities in excess of $\gtrsim 500\,\mathrm{km\,s^{-1}}$), that would be indicative of outflows. Our findings suggest that we are probing the interstellar medium and associated star formation in the quasar host galaxies up to radii of 10 kpc, whereas we find no evidence for halos or outflows.

Keywords: AGN host galaxies — High-redshift galaxies — Dust continuum emission — Interstellar line emission

1. INTRODUCTION

Quasars and their hosts are ideal targets to probe the properties of massive high-redshift galaxies. Powered by the rapid accretion of material, near the Eddington limit, onto a supermassive black hole (SMBH, see e.g., De Rosa et al. 2011, 2014; Willott et al. 2010), quasars within the first Gyr of the Universe ($z > 6$) are easily detected by current facilities. Several hundred quasars have now been identified at $z > 5.5$, owing to both large surveys (see York et al. 2000; Arnaboldi et al. 2007; Lawrence et al. 2007; Chambers et al. 2016), and improved selection methods with follow-up observations (see e.g., Fan et al. 2006; Venemans et al. 2007; Willott et al. 2010; Mortlock et al. 2011; Morganson et al. 2012; Bañados et al. 2016; Jiang et al. 2016; Mazzucchelli et al. 2017; Matsuoka et al. 2018; Bañados et al. 2018; Wang et al. 2018; Yang et al. 2020).

At $z \sim 6$, the singly ionized carbon emission line at 158 μm, arising from the $^2P_{3/2} \rightarrow {}^2P_{1/2}$ transition (hereafter referred to as the [C II] line) falls conveniently within the atmospheric transition window of ground based interferometers such as the Atacama Large Millimeter Array (ALMA) and the NOrthern Extended Millimeter Array (NOEMA). This emission line is one of the brightest far-infrared (FIR) emission lines and can be used to trace the cold molecular gas of the interstellar medium (ISM, see Carilli & Walter 2013 for a review of high-redshift galaxies). The line predominantly arises from within the photodissociation regions (PDRs) found around newly formed stars,

Corresponding author: Mladen Novak
novak@mpia.de

although it can also stem from ionized regions (e.g., Hollenbach & Tielens 1999; Vallini et al. 2017; Ferrara et al. 2019). The [C II] line has been used to probe the gas distribution and kinematics on kpc scales for dozens of quasar host galaxies, simultaneously providing precise redshift measurements (e.g., Walter et al. 2009; Maiolino et al. 2012; Wang et al. 2013; Willott et al. 2015; Venemans et al. 2017a,b; Decarli et al. 2018; Neeleman et al. 2019).

Detections of large-scale [C II] emission and outflows in a high-redshift quasar host galaxy were first reported for the $z = 6.4$ system SDSS J1148+5251 (Maiolino et al. 2012; Cicone et al. 2015). However, further statistical studies of [C II] outflows in various, partly overlapping, samples of $z > 4.5$ quasar host galaxies remain without consensus. Some studies report tentative or strong evidence for outflows (e.g., Bischetti et al. 2019; Stanley et al. 2019), whereas other studies reported no outflows (e.g., Decarli et al. 2018). However, different image plane stacking techniques were used among the different studies.

We revisit this topic by using the deepest available ALMA data and novel analysis techniques. In this study, we conduct a multi-resolution analysis of 27 quasar host galaxies at $z \gtrsim 6$ observed with ALMA in order to search for signatures of extended [C II] emission and outflows. We require ~ 1 kpc imaging capability in order to accurately derive surface brightness profiles, but complement our data with lower spatial resolution observations of the same sources to minimize potential issues of missing flux and outresolving sources. For the spectral analysis, we use a novel technique of spectral uv-plane stacking. By directly averaging observed visibilities of the sample of galaxies, in velocity bins of interest, we circumvent various problems present in interferometric image-based stacking (e.g. dirty beam residuals and beam matching), simultaneously ensuring that emission on multiple spatial scales can be recovered in the imaging step.

This is the third paper in a series of studies, in which different aspects of the quasar host galaxies observed in [C II] at ~ 1 kpc resolution are discussed. The first paper, Venemans et al. (2020, accepted), describes the sample in detail and provides the analysis of both the [C II] and the dust continuum emission of individual galaxies. The second paper, Neeleman et al. (2020, in prep.), capitalizes on the high resolution of these observations by modeling the gas kinematics of the host galaxies, yielding rotation and/or dispersion velocities, and estimates of the dynamical masses of the hosts. These two studies investigate properties of individual galaxies, making use of the brightest emitting regions, where there is sufficient signal. In this paper, we perform stacking and uv-plane analysis to constrain the amount of faint, but extended (both spatially and spectrally), emission, that is below the detection threshold of individual objects.

The paper is structured as follows. In Section 2, we present the sample of $z \gtrsim 6$ quasar host galaxies along with the new and archival ALMA data. We also describe the data reduction steps, along with the details of the uv-stacking procedure. In Section 3, we elaborate on the methods for accurately measuring fluxes, both from the image, and the uv-plane. In Section 4, we present the results in three subsections, focusing on the 1) spatial extent of the [C II] emission, 2) spatial extent of the dust continuum, and 3) spectral analysis of the [C II] line. We interpret and discuss our findings in the context of published studies in Section 5, and summarize our main conclusions in Section 6.

Throughout the paper we assume the concordance Lambda cold dark matter (ΛCDM) cosmology, defined by a Hubble constant $H_0 = 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, dark energy density $\Omega_\Lambda = 0.7$, and matter density $\Omega_\mathrm{m} = 0.3$. The mean redshift of our sample is $z = 6.4$, at which an angular size of $1''$ corresponds to a projected physical distance of 5.5 kpc.

2. DATA AND SAMPLES

The main sample for our study consists of 27 quasar host galaxies[1] at redshifts $z \gtrsim 6$ observed with ALMA in the [C II] emission at ~ 1 kpc resolution ($\sim 0\farcs25$). With the high-resolution criteria satisfied, we also supplemented the data with low-resolution archival observations of these sources, where available. Our sample spans the redshift range of $z = 5.8 - 7.5$, and contains quasar host galaxies with [C II] luminosities in the range of $(0.8 - 9) \times 10^9\,\mathrm{L_\odot}$, and FIR luminosities in the range of $(0.5 - 12) \times 10^{12}\,\mathrm{L_\odot}$ (see Venemans et al. 2020, accepted, for further details on the sample selection).

2.1. *New and archival ALMA data*

For our analysis, we consider all available 12 m and 7 m ALMA observations performed in cycles 1 through 6, which were obtained at a resolution of $0\farcs1$ or coarser. The bulk of the high spatial resolution data comes from our recent programs with IDs 2017.1.01301.S and 2018.1.00908.S. We have excluded cycle 0 data from our selection to avoid possibly lower quality data and known difficulties[2] of combining these observations with subsequent cycles. We have also excluded higher resolution (sub-kpc) observations available for two of the quasar host galaxies, because there is insufficient overlap of available baseline lengths compared to our main data sample. Finally, we have excluded quasar host galaxies without ~ 1 kpc observations from the analysis, in order to have a consistent sample across the paper series. These selection criteria yield a main sample of 27 quasar host galaxies, where all objects, by construction, have ~ 1 kpc data available.

Although the high-resolution data resolve the gas and dust structure and the kinematics of the host galaxy, they may not be sensitive to large-scale emission, which may be either below the detection limit, or out-resolved due to the lack of

[1] Although we happen to have the same number of galaxies, our sample is not the same as the one described in Decarli et al. (2018), as only 15 objects are shared between the two samples. For our study, we require data obtained at high resolution ($\sim 0\farcs25$), in contrast to the $\sim 1''$ used in Decarli et al. (2018).

[2] E.g., fewer antennas available, sometimes missing calibration data, no pipeline support.

shorter baselines. In order to quantify this effect and mitigate the issue as much as possible, we make use of all available archival data targeting the [C II] emission line in the galaxies of our sample. Datasets with lower spatial resolution observations, obtained across multiple ALMA observational cycles, are available for 20 objects in our sample.

In total, this work uses 54 different observations targeting 27 objects, as listed in Table 1, where all datasets correspond to 12 m array observations, unless noted otherwise. The redshift, [C II] integrated line flux density, and the full width at half maximum (FWHM), reported in Table 1, are all measured from the aperture-extracted [C II] spectra, whereas coordinates are measured from the [C II] intensity maps, as reported in Venemans et al. (2020, accepted). For each object we list used observation runs in various ALMA cycles, including their time on source (TOS), the synthesized beam size, and the maximum recoverable scale (MRS). The last two values are computed based on the baseline statistics following the ALMA Technical Handbook (Eqs. 7.4 and 7.7), i.e. $\theta_{\rm beam}[{\rm rad}] = 0.574\lambda/L_{80}$ and $\theta_{\rm MRS}[{\rm rad}] = 0.983\lambda/L_5$, where λ is the observing wavelength, and L_5 and L_{80} are the 5$^{\rm th}$ and the 80$^{\rm th}$ percentile of the available uv-distances, respectively. The MRS is defined as the largest angular size at which at least 10% of the total flux density of a uniform disk is recovered and should only be used to describe the data limitations to first order.

2.2. *Data reduction*

To obtain the calibrated uv-visibilities for imaging, we reduced the raw data using the default pipeline restoration scripts executed in the appropriate version of the Common Astronomy Software Applications (CASA) package (McMullin et al. 2007), i.e. the same version applied during the original calibration (CASA 4.1 to 5.4.0-68). Additional data quality checks and data manipulation were performed as described in the following section.

2.2.1. *Flux calibration accuracy*

We combine data from multiple objects and ALMA observation cycles. Because the addition of visibilities at incorrect flux scales could translate into erroneous spatial structure in the image plane, we performed several quality assurance checks on the flux calibration. As noted by Stanley et al. (2019), several datasets required recalibration in a newer CASA version due to incorrect flux calibrator models present for Pallas and Ceres in CASA version 4.6 or lower. There are a total of nine datasets affected by this issue, as indicated in Table 1. Flux scales after recalibration were $60 - 80\%$ of the original values, with the mean flux level reduction of 15% for these nine datasets.

As a precaution, we reran the full pipeline calibration process on each of the 54 datasets, using the latest CASA version at the time of writing (version 5.6.1-8), as described in the ALMA Science Pipeline User's Guide for CASA 5.6.1 (Section 6.4). Additionally, we manually enabled the query to the online ALMA flux calibrator catalog. The goal of this approach was to check for the existence of additional significant flux calibration changes introduced in the CASA development and debugging process, as well as to get the most up-to-date flux calibrator model values. For most of the datasets, the process finished successfully[3] and yielded no flux level changes beyond 10%, which is the usually reported calibration accuracy value. Because we found no further anomalies, we continued to use the calibrated visibilities provided by the default restoration scripts (except in the case of Pallas and Ceres calibrators as mentioned above).

2.2.2. *Data products per galaxy*

We have created several data products (measurement sets, maps and cubes) in CASA for every observational cycle of each of the studied quasar (see Table 1), as described in Venemans et al. (2020, accepted). We briefly summarize these here. We imaged the calibrated visibilities, using the CASA task TCLEAN, with the Högbom deconvolver, and a channel spacing of 30 MHz (corresponding to 35 km s^{-1} on average). We weighted the data with the Briggs[4] algorithm (ROBUST = 0.5) and cleaned (deconvolved) the maps down to 2σ level, within a cleaning mask radius of $2''$. We use these clean cubes to extract the [C II] spectrum including the continuum emission. With the knowledge of the line position, we performed the continuum subtraction using the UVCONTSUB task on the two spectral windows around the [C II] line, excluding channels in total width of $2.5 \times$ FWHM around the [C II] line (i.e. the total excluded bandwidth is on average 900 km s^{-1}), while allowing for a continuum slope (FITORDER = 1). Using these datasets we imaged the continuum-free cube (same imaging parameters as above), as well as the [C II] intensity map by averaging channels across $1.2 \times$ FWHM of the line, chosen to optimize the signal-to-noise ratio (S/N). We justify this integration width in Appendix A.

2.2.3. *Stacking the data in the uv-plane*

Stacking can be performed in the image plane by either averaging the pixel/voxel values of maps or cubes, or by averaging extracted properties such as the spectra. Both methods are susceptible to several issues pertaining to the nature of interferometric data. I.e., the synthesized beam is not the same between datasets, and the faint stacked emission is usually uncleaned, which will yield ill-defined hybrid units in the stacked map (i.e., every pixel value will be a non-trivial combination of various cleaned and dirty beams). Another side-effect of image based stacking is that different spatial scales might be probed in different sources. In order to circumvent these issues, we aim to add all visibilities in the uv-plane and thus obtain the best estimate of the mean emission of the population. To this effect we have adapted the method

[3] Due to data format changes, CASA task definition changes, and backward incompatibility, this process failed for some datasets.

[4] We note that we recover the same aperture flux densities in both natural and Briggs weighted maps using the residual scaling method explained in a later section.

Table 1. ALMA observations of [C II] emission in $z \sim 6$ quasar host galaxies used in our study (27 objects, 54 datasets).

Quasar host galaxy	$z_{[CII]}$	RA (ICRS) °	Dec (ICRS) °	$F_{[C\,II]}$ Jy km s^{-1}	FWHM km s^{-1}	Cycle	TOS min	Beam[a] ″	MRS ″	Project code	Member ObsUnitSet ID
P007+04	6.0015	7.02736	4.95706	1.7 ± 0.1	370 ± 22	3[c]	7.4	0.44	5.0	2015.1.01115.S	uid://A001/X2fb/X3b8
						5	29.1	**0.23**	3.2	2017.1.01301.S	uid://A001/X1273/X360
P009–10	6.0040	9.73553	-10.43168	9.6 ± 0.7	437 ± 33	3[c, b]	8.4	0.4	4.5	2015.1.01115.S	uid://A001/X2fb/X3bc
						5	25.7	**0.23**	3.3	2017.1.01301.S	uid://A001/X1273/X364
J0100+2802	6.3268	15.05426	28.04051	3.7 ± 0.2	405 ± 20	3[b]	72.4	**0.18**	3.5	2015.1.00692.S	uid://A001/X2d6/X1a8
J0109–3047	6.7904	17.47135	-30.79065	1.7 ± 0.1	354 ± 34	1	15.4	0.27	4.0	2012.1.00882.S	uid://A002/X5a9a13/X53b
						2[c]	35.8	0.2	2.1	2013.1.00273.S	uid://A001/X148/X6b
						3	32.4	**0.16**	2.9	2015.1.00399.S	uid://A001/X5a3/X52
J0129–0035[d]	5.7788	22.49381	-0.59440	2.1 ± 0.1	206 ± 9	1[c, b]	76.5	**0.18**	2.0	2012.1.00240.S	uid://A002/X7fb989/X1c
						3	60.7	0.34	4.4	2015.1.00997.S	uid://A001/X2fb/X5de
						3 - 7m	209.4	4.0	25.0	2015.1.00997.S	uid://A001/X2fb/X5e0
J025–33	6.3373	25.68218	-33.46264	5.5 ± 0.2	370 ± 16	3	7.9	0.61	5.8	2015.1.01115.S	uid://A001/X2fb/X3c4
						5	24.2	**0.23**	3.2	2017.1.01301.S	uid://A001/X1273/X368
P036+03	6.5405	36.50782	3.04979	3.2 ± 0.1	237 ± 7	3	75.5	**0.13**	1.6	2015.1.00399.S	uid://A001/X5a3/X4a
J0305–3150[e, f]	6.6139	46.32052	-31.84888	5.4 ± 0.3	225 ± 15	1	15.9	0.26	3.7	2012.1.00882.S	uid://A002/X5a9a13/X543
						2[c]	15.5	0.2	2.1	2013.1.00273.S	uid://A001/X148/X6f
						3	37.7	**0.17**	3.0	2015.1.00399.S	uid://A001/X5a3/X4e
P065–26	6.1871	65.40851	-26.95432	1.7 ± 0.2	289 ± 31	3	15.8	0.74	7.3	2015.1.01115.S	uid://A001/X2fb/X3e4
						5	24.7	**0.23**	3.3	2017.1.01301.S	uid://A001/X1273/X36c
J0842+1218	6.0754	130.62266	12.31402	0.8 ± 0.1	378 ± 52	3	7.4	0.98	9.2	2015.1.01115.S	uid://A001/X2fb/X3ec
						4	53.6	**0.25**	2.7	2016.1.00544.S	uid://A001/X885/X36b
J1044–0125[d]	5.7846	161.13767	-1.41724	1.8 ± 0.2	454 ± 60	1[b]	76.5	**0.17**	1.9	2012.1.00240.S	uid://A002/X7fb989/X20
						3	60.7	0.57	5.0	2015.1.00997.S	uid://A001/X2fb/X5e4
						3 - 7m	179.4	3.9	24.0	2015.1.00997.S	uid://A001/X2fb/X5e6
J1048–0109	6.6759	162.07949	-1.16123	1.9 ± 0.1	299 ± 24	3	11.9	0.95	9.1	2015.1.01115.S	uid://A001/X2fb/X3f4
						5	25.7	**0.23**	2.8	2017.1.01301.S	uid://A001/X1273/X370
P167–13[e]	6.5144	167.64160	-13.49607	5.3 ± 0.3	519 ± 25	3 - CW[b, g]	42.8	0.61	5.4	2015.1.00606.S	uid://A001/X2d6/X7d
						3 - FW[g]	7.9	0.88	8.5	2015.1.01115.S	uid://A001/X2fb/X3f8
						4	42.7	**0.25**	3.7	2016.1.00544.S	uid://A001/X885/X367
J1120+0641	7.0848	170.00611	6.68996	1.0 ± 0.1	416 ± 39	1	160.7	**0.22**	2.6	2012.1.00882.S	uid://A002/X5a9a13/X537
P183+05	6.4386	183.11240	5.09266	6.8 ± 0.3	397 ± 19	3	8.4	0.92	8.9	2015.1.01115.S	uid://A001/X2fb/X408
						4	47.1	**0.25**	2.7	2016.1.00544.S	uid://A001/X885/X363
J1306+0356[e]	6.0330	196.53441	3.94061	1.2 ± 0.1	246 ± 26	3	8.4	0.85	8.2	2015.1.01115.S	uid://A001/X2fb/X40c
						5	27.7	**0.23**	3.3	2017.1.01301.S	uid://A001/X1273/X374
J1319+0950[d, e]	6.1347	199.79701	9.84763	4.1 ± 0.4	532 ± 57	1[b]	50.4	**0.22**	2.0	2012.1.00240.S	uid://A002/X7fb989/X18
						3	30.3	0.93	9.0	2015.1.00997.S	uid://A001/X2fb/X5d2
						3 - 7m	119.6	3.9	26.0	2015.1.00997.S	uid://A001/X2fb/X5d4
J1342+0928[e]	7.5400	205.53375	9.47736	1.0 ± 0.1	353 ± 27	5	114.1	**0.16**	3.0	2017.1.00396.S	uid://A001/X1296/X976
P231–20[e]	6.5869	231.65765	-20.83354	3.3 ± 0.3	393 ± 35	3	7.4	0.87	8.5	2015.1.01115.S	uid://A001/X2fb/X440
						4	43.1	**0.17**	2.1	2016.1.00544.S	uid://A001/X885/X35f
P308–21[e]	6.2355	308.04167	-21.23399	3.4 ± 0.2	541 ± 32	3	12.4	0.63	5.9	2015.1.01115.S	uid://A001/X2fb/X418
						4	55.6	**0.24**	3.6	2016.A.00018.S	uid://A001/X11a4/Xf
J2054–0005	6.0389	313.52708	-0.08735	3.2 ± 0.1	236 ± 12	6	84.9	**0.11**	2.1	2018.1.00908.S	uid://A001/X133d/X261d
J2100–1715	6.0807	315.22792	-17.25610	1.4 ± 0.1	361 ± 41	3[c]	7.9	0.57	5.2	2015.1.01115.S	uid://A001/X2fb/X41c
						5	25.2	**0.22**	3.2	2017.1.01301.S	uid://A001/X1273/X378
P323+12	6.5872	323.13826	12.29865	1.3 ± 0.2	271 ± 38	6	42.9	**0.1**	1.7	2018.1.00908.S	uid://A001/X133d/X2621
J2318–3113	6.4429	349.57651	-31.22955	1.5 ± 0.1	344 ± 34	3[c]	7.9	0.66	5.7	2015.1.01115.S	uid://A001/X2fb/X428
						5	23.7	**0.24**	3.5	2017.1.01301.S	uid://A001/X1273/X37c
J2318–3029	6.1456	349.63792	-30.49266	2.3 ± 0.1	293 ± 17	6	39.9	**0.1**	1.7	2018.1.00908.S	uid://A001/X133d/X2619
J2348–3054[f]	6.9007	357.13895	-30.90285	1.5 ± 0.2	457 ± 49	1	16.7	0.41	5.3	2012.1.00882.S	uid://A002/X5a9a13/X53f
						2[c]	36.8	0.19	2.0	2013.1.00273.S	uid://A001/X148/X73
						3	51.8	**0.17**	3.1	2015.1.00399.S	uid://A001/X5a3/X56
P359–06	6.1719	359.13517	-6.38313	2.7 ± 0.1	341 ± 18	3[c]	7.9	0.66	6.3	2015.1.01115.S	uid://A001/X2fb/X430
						5	25.7	**0.22**	3.1	2017.1.01301.S	uid://A001/X1273/X380

[a]Synthesized beam obtained from baseline statistics (see main text), given in arcsec. Nominal high-resolution (∼1 kpc) datasets are marked with boldface.
[b]Removed from uv stacking due to data weights outliers or partial line frequency coverage (in case of P009–10, cycle 3).
[c]Recalibrated due to wrong Ceres or Pallas flux calibrator models.
[d]Cycle 0 observations are also available, but are excluded from our study.
[e]Merger or nearby companion identified at < 10 kpc.
[f]Higher resolution (<40 mas) also available, but discarded here (see main text).
[g]The letters refer to the PI initials distinguishing between same cycle observations.

described in e.g. Fujimoto et al. (2019), and extended it to be applicable to the whole spectrum. This method requires the data to be further reduced as follows.

For every individual cycle, starting from the continuum subtracted visibilities, the required channels were split out. Depending on the science goal, we either split out the equivalent of the [C II] intensity map ($1.2 \times$ FWHM), or, a smaller chunk corresponding to a specific velocity bin (i.e. a channel map, and/or the resolution of our spectral extraction). We used the function IM.ADVISECHANSEL from the CASA toolkit to select the channels of interest in the kinematic local standard of rest frame (LSRK). Individual channels are selected based on the nearest neighbour algorithm. The data that we consider for uv-stacking were all observed at velocity resolutions of $\lesssim 10\,\mathrm{km\,s^{-1}}$, which is high enough to render any additional velocity-based data regridding or interpolation unnecessary. The data was further time-averaged inside 30 s bins to reduce the data volume[5]. The coordinate reference system in the header was changed to the International Celestial Reference System (ICRS) as it was sometimes incorrectly written as J2000. The visibilities were phase-shifted to the centroid of the [C II] emission using the task FIXVIS[6]. The header was modified further to artificially place the object at the $(0, 0)$ coordinate[7]. Finally, the data weights were recomputed with the STATWT task using the same time and channels bin for all of the datasets. The calculation of the default weights is implemented differently across various CASA versions and can depend on the system temperature or gain factors. We manually checked that the ranges of recomputed weights are similar between all datasets, allowing the data to be co-added in a way that ensures the weighted average is not erroneously dominated by any single source. Several anomalous datasets were identified as having (at least) an order of magnitude larger data weights (they also have much lower observed velocity resolution of $\gtrsim 30\,\mathrm{km\,s^{-1}}$), and were excluded from any further uv-stacking analysis. These exclusions are noted in Table 1.

We imaged the split, renormalized and re-centered data using the CASA task TCLEAN, considering the full list of all sources and cycles. We imaged the data in the continuum mode (SPECMODE = MFS) and applied the multi-scale deconvolver, with scales corresponding to 0 (delta-function), 1, 3, 5 and 10 synthesized beam sizes. Cleaning was performed within a $2''$ radius mask down to a 2σ threshold. To obtain the uv-stacked cube we imaged every velocity bin individually. The velocity bin was either of fixed size, i.e. $30\,\mathrm{km\,s^{-1}}$, or scaled with the width of the [C II] line, i.e. $0.2\sigma_{\mathrm{line}}$. We used both approaches in our analysis. These separately imaged velocity bin slices were further joined into a single cube using the CASA toolkit task IA.IMAGECONCAT.

The preparation of the data for uv-plane continuum stacking was performed in the same way as in the [C II] case, but instead of selecting the line emission channels, we flagged them. A total of width of $2.5 \times$ FWHM centered at [C II] was removed, and the remaining channels inside individual spectral windows were averaged together. All available continuum data were used (i.e. data from all four science spectral windows). We validated that our choice of excluded frequencies/velocities did not bias the dust continuum results by re-imaging the data, this time excluding a much broader bandwidth of $8 \times$ FWHM around the line ($\sim$2800 $\mathrm{km\,s^{-1}}$ on average).

Spectral stacking will smooth any spatial and velocity structure of individual sources. The stack itself is only meaningful if the mean property of interest (e.g. extent of the faint emission, or the line shape) is well-defined for the sample, which we cannot know a priori, and deeper observations are necessary to confirm any potential structure in individual objects.

2.3. *Mergers and companions*

Our aim is to measure extended structure and high-velocity outflows in quasar host galaxies. Thus, we take special care of objects exhibiting ongoing mergers, and of those that have a nearby companion galaxy. In both cases the accompanying source might bias our results, by either providing additional flux at larger distances from the quasar host, or at larger velocity separations. Where possible, we visually identified the presence of accompanying sources, distinct in terms of either their spatial, or velocity position ([C II] intensity maps are available in Appendix G; see also accompanying papers by Venemans et al. 2020, accepted and Neeleman et al. 2020, in prep.). Morphological details on sub-kpc scales cannot be recovered with these observations. What we consider as a single component source in this work might break into multiple sources upon a higher resolution follow-up. Thus, definitions such as a merger or a galaxy pair are, by construction, arbitrary, especially in the low S/N regime. With these limitations in mind, we classify these special cases based on visual inspection into four groups as follows (see also Decarli et al. 2017, Decarli et al. 2019, Neeleman et al. 2019 and Bañados et al. 2019a):

I) Merging system blended inside the aperture: These systems exhibit at least two clearly separated (spatially or in velocity) peaks at distances of less than ~ 5 kpc. In our data it is not possible to draw a clear line between the two sources or easily measure their individual fluxes due to a connecting bridge of [C II] emission or blending. We classify five of such systems: J1319+0950, J1342+0928, P167–13, P308–21 and J0305–3150.

II) Nearby companion: These are defined as sources that have a spatial offset of $\sim 5 - 10$ kpc between the quasar host and its galaxy pair, such that an aperture flux can be obtained for both objects individually. However, the system will show increased flux on the smallest baselines that cover

[5] The value is large enough to accomodate all datasets, but small enough not to introduce any time smearing effects.

[6] The task uses small angle approximation, which is valid in our case as shifts are at most few arcseconds large, and less than an arcsecond in most of the cases.

[7] This was achieved by putting zeros into PHASE_DIR, DELAY_DIR, REFERENCE_DIR columns of the measurement set.

both sources simultaneously. We classify two such systems: J1306+0356 and P231–20.

III) Distant companions: These are cases where there is a galaxy pair in the field, but at separations larger than 10 kpc. No significant amplitude increase can be seen on the smallest baselines due to this large separation (the primary beam response at the companion source position is also significantly lower). We classify two such systems: J0842+1218 and J2100–1715.

IV) Complex morphologies: These systems show a surface brightness morphology that is more complex than a single component source, but its origin is unclear. It is possible that these objects have companions detectable only at higher resolution, or that they contain extended and more patchy gas emission. In the absence of conclusive companion or merger evidence, we do not exclude these objects from the stacking analysis. We classify three such systems: J0100+2802, J025–33 and P009–10.

Groups I and II are removed from some of our samples, as explained in the next section. Groups III and IV are always included in our stacking samples, with the latter containing potential candidates for follow-up observations.

2.4. *Data samples*

Throughout the paper we refer to four different data samples, all drawn from Table 1, as follows.

1) High-resolution data sample: It contains all 27 quasar host galaxies, where each one is represented by a single high-resolution ($\sim$ 1 kpc) dataset. This corresponds to the data used in the accompanying papers (Venemans et al. 2020, accepted; Neeleman et al. 2020, in prep.).

2) High-resolution clean data sample: It is a subset of the previous sample, obtained by removing five objects classified as merging systems blended inside the aperture (group I from Section 2.3). This sample is used to stack aperture spectra extracted from individually imaged maps. Nearby companions (group II) will not contribute significantly to the aperture fluxes, and are therefore included in this sample.

3) Full uv-stacking data sample: It contains all of the datasets (including low resolution data), except five observing runs that have outlying data weights, and one without a full line coverage (see Table 1 footnote and Section 2.2.3). This selection results in 26 quasar host galaxies (out of 27) and a total of 48 datasets (out of 54). The cumulative on source time is approximately 34 h (out of which 25.5 h were taken with the 12 m array). This sample is used to obtain the highest S/N in the uv-stack, while attempting to recover broad-line emission, which would be indicative of outflows.

4) Clean uv-stacking data sample: It is a subset of the previous sample, obtained by further removing systems classified as mergers inside the aperture and near companions (groups I and II from Section 2.3), as well as any 7 m dataset (for details see Appendix E). This selection results in 19 quasar host galaxies and a total of 32 datasets. The cumulative time on source is approximately 18 h. We consider this sample to be the least biased, and use it to derive our main stacking results.

3. METHODS

In the following sections we present different methods employed in our search for extended [C II] emission and outflows. We discuss flux measurements taken from the image and the uv-plane, and the applicability and shortcomings of each method. We also discuss the details of the stacking procedures. By applying various techniques we aim to reach a consensus on the measured galaxy properties (e.g. total flux and spatial extent). This is particularly relevant for detections of low S/N. Throughout this whole section we demonstrate various diagnostics on the source J1306+0356, which was chosen as a good example, due to it being a complex system (presence of a companion, see also Neeleman et al. 2019), with data available from multiple cycles at different resolutions. Results for all individual quasar host galaxies are shown in Appendix G.

3.1. *Measuring the spatial extent of emission*

3.1.1. *Analysis in the image plane*

A common method of measuring the flux and source size from a map is by fitting a 2D Gaussian to the galaxy in the image plane. We do so by employing the IMFIT task included in CASA. The second approach that requires no priors on the source morphology is the aperture integration, for which we choose to apply circular apertures. In the remainder of this section we explore only the aperture integration, but the discussed systematics are applicable to any image based analysis (i.e. fitting a 2D Gaussian).

One problem present in image plane analysis is the non-trivial unit definition (Jy beam^{-1}). Due to the nature of the deconvolution (cleaning) process, the final cleaned map will contain hybrid units. Several maps from the cleaning process, shown in Figure 1, demonstrate this issue. The final cleaned map is obtained by summing the residual map (in units of Jy / dirty beam) and the clean components map (in units of Jy / clean beam), whose units differ due to the beam definition. The clean beam is obtained[8] by fitting a 2D Gaussian to the peak of the point spread function (PSF; the dirty beam). It has a well defined volume (integral) equal to $\Omega_{\rm beam} = \pi/(4\ln 2) \times \theta_{\rm maj}\theta_{\rm min}$, where $\theta_{\rm maj}$ and $\theta_{\rm maj}$ are major and minor axis FWHMs of the elliptical 2D Gaussian. On the other hand, the dirty beam is a sum of many sine/cosine waves, whose integral oscillates around zero. Therefore, the volume of the dirty beam is ill-defined and will depend on the integration limits (i.e. the aperture size).

The clean components map is defined as the cleaned sky model (i.e. point sources / delta functions, and, in case of the multi-scale algorithm, additional Gaussians of manually chosen sizes) convolved with the Gaussian clean beam. During the sky reconstruction, the clean algorithm does not distinguish between positive and negative peaks. The true emis-

[8] This can be manually overridden by the RESTORINGBEAM parameter in TCLEAN, but we made no such attempt.

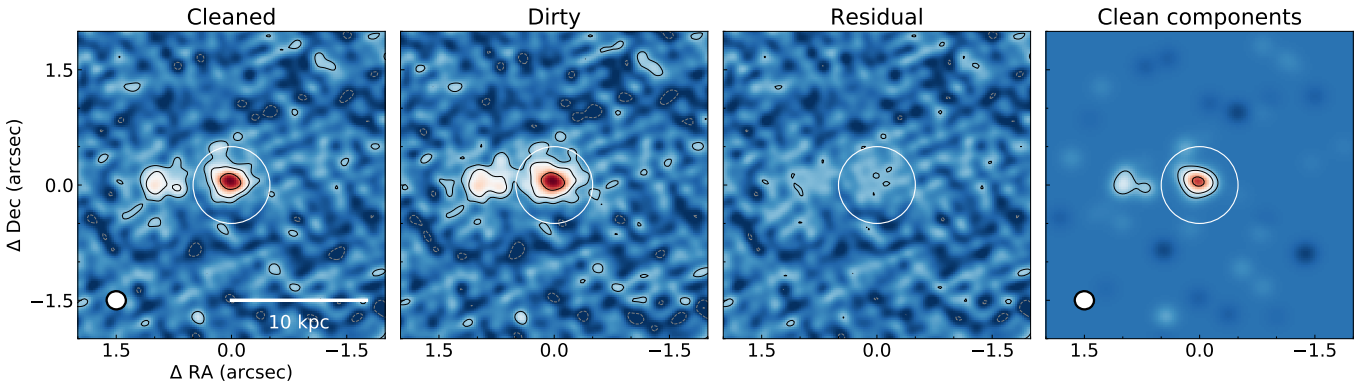

Figure 1. An example of the [C II] intensity map, imaged from the high-resolution dataset using $1.2 \times$ FWHM of the line width (the system shown is J1306+0356). Contours are logarithmic starting from 2σ, continuing in powers of 2, with dashed ones indicating negative values. The four panels show maps from various deconvolution stages. The clean beam is shown in the bottom left corner. The white circle is the manually chosen aperture (radius of $0\farcs5$), that captures the emission of the quasar host galaxy only, which intensionally excludes the companion galaxy (encompassing some fraction of emission in the connecting bridge is unavoidable). The cleaned map is the sum of the residual map (units: Jy / dirty beam) and the cleaned components map (units: Jy / clean beam), and is therefore affected by ill-defined units. Excess flux in the residual map can lead to a non-negligible bias in the total flux measurement.

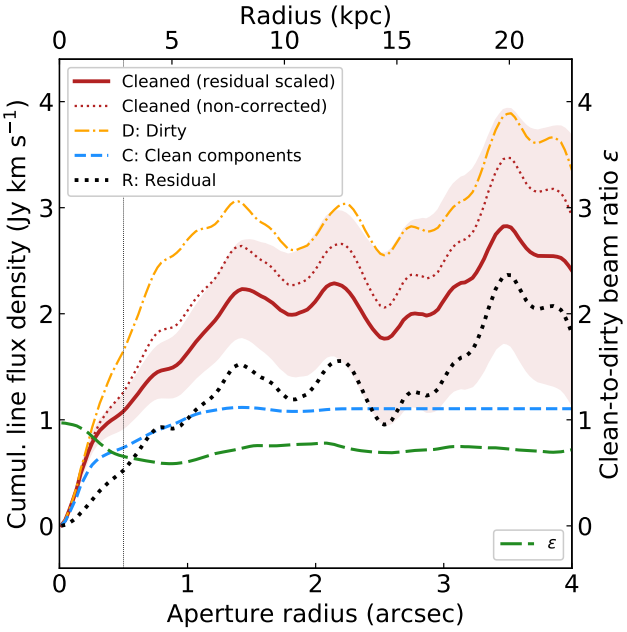

Figure 2. An example of the curves of growth used in the residual scaling method. As in Figure 1, we show the results for J1306+0356. The corrected flux density is obtained by scaling the uncleaned flux with the proper beam size, and its 1σ uncertainty is shown with the shaded region. The vertical line shows the manually chosen aperture (radius of $0\farcs5$), used to derive the total flux of the quasar host galaxy. Additional flux at larger radii is due to the presence of a companion galaxy. The clean-to-dirty beam ratio is roughly constant at larger radii. Within the chosen aperture, more than a third of the measured flux is contributed by the residual. The residual-scaled cleaned flux density is 85% of the one obtained from the standard cleaned map.

sion is positive, therefore the cleaning should not progress too deep, in order to avoid cleaning both positive and negative noise peaks. After the sky model is subtracted from the visibilities, the residual remains. Any flux measured from the residual (i.e. flux below the cleaning threshold) will be incorrectly assigned the clean beam size (defined to be a Gaussian in the restoration process, and written in the image header of the final map).

To mitigate this issue we scale the flux collected from the residual map (for details see Jorsater & van Moorsel 1995; Walter et al. 2008; Novak et al. 2019). In practice, this is performed by measuring aperture fluxes from three maps: the dirty (D), the residual (R), and the clean components[9] (C) map (see Figure 1). The flux density measurement is obtained by summing the values of pixels contained within the aperture, and dividing the sum with the number of pixels in the clean beam: $F_{\rm aper}[{\rm Jy}] = \sum_{\rm pixels} {\rm map}[{\rm Jy\,beam^{-1}}]/(\Omega_{\rm beam}/{\rm pixelsize}^2)$, where the beam axes sizes and the pixel size are all expressed in the same units (e.g. arcseconds). Corrected flux inside an aperture can be obtained as $F_{\rm correct} = C + \epsilon R = \epsilon D$, where $\epsilon = C/(D-R)$ is the clean-to-dirty beam area ratio defined inside a specific aperture (see Appendix B). The drawback of this method is that it requires some cleaned flux, and becomes numerically unstable when R approaches D. The error on the aperture flux can be approximated by ${\rm rms} \times \sqrt{N}$, where N is the number of beams contained inside the aperture, and the root-mean-square (rms) is the noise variation measured in the entire map.

[9] The clean components map can be obtained by either convolving the sky model map (units of Jy/pixel) with the clean beam, or by subtracting the residual map from the final map output by CASA.

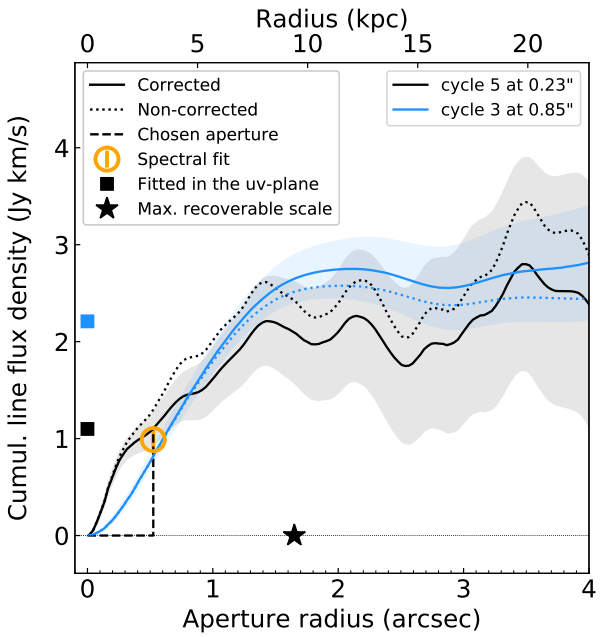

Figure 3. An example of the curves of growth measured from [C II] intensity maps of two different observation cycles (3 and 5, as shown in the legend) targeting J1306+0356. The shaded area indicates the 1σ uncertainty of the residual-scaled flux density. The dashed line marks the aperture size chosen to encompass the quasar host galaxy only (see Figure 1). The squares on the zero radius line show total flux densities derived from single Gaussian fits in the uv-plane (colored by cycle, same as in Figure 4). The star shows the MRS, which is larger than the plotted range in the case of cycle 3 observations. The orange circle shows the flux density obtained from a spectrum fit of the high-resolution cycle data (see Figure 6), multiplied by 0.84 to remove flux beyond $1.2 \times$ FWHM. This figures demonstrates that the companion is blended with the quasar host galaxy in the lower resolution data, and that residual scaling corrections differ based on the beam shape, and can go in either positive or negative direction.

The amount of the residual scaling depends on the exact shape of the dirty beam (defined by the uv-coverage, applied imaging weights, and any additional uv-tapering) and the aperture size, and can either boost or decrease the measured flux density. If the scaling is significant, not applying the correction is likely to produce larger systematic uncertainties in the case of extended sources that are resolved over multiple beams, sources with low surface brightness (where significant amount of emission remains in the residual), and in the case of a non-Gaussian dirty beam (more likely to happen with naturally weighted imaging). Image-based stacking is also likely to be affected, especially if maps with different dirty beams are being stacked, because the knowledge of the beam shape and the fraction of cleaned flux is lost in the stacking process. Furthermore, the units become increasingly non-trivial after 2D image convolution, often employed for beam matching[10].

In the first panel of Figure 1, we show an example of a high-resolution [C II] intensity map, produced by imaging averaged channels in total width of $1.2 \times$ FWHM (range of $\pm 1.4\sigma_{\rm line}$) from the continuum subtracted measurement set. The measured aperture flux density can be corrected for the missing tails beyond the $1.2 \times$ FWHM, assuming a Gaussian profile, by multiplying by 1.19. In Figure 1, a second 4σ peak is apparent 5 kpc away (see also Neeleman et al. 2019). Due to the large separation we are able to define an aperture containing the full source emission, with no significant contribution from the companion or the connecting bridge.

In Figure 2 we show flux densities measured inside a growing aperture (hereafter referred to as the curves of growth) for all of the maps from Figure 1. The clean beam size can only be used to measure the emission in the clean components map. If used on other maps, it will introduce systematic errors if the clean-to-dirty beam ratio defined inside the aperture is different from one. In this example, the ratio is 0.65 at the chosen aperture of $0\farcs5$, therefore any uncleaned flux is being over-counted by a factor of $1/0.65 \approx 1.5$. Because of this, the corrected flux density is 85% of the one measured in the cleaned map without residual scaling.

In Figure 3 we show measurements from two observation cycles of the same source, taken at different resolutions. The quasar host galaxy and its companion are blended together in the low resolution data (beam $\approx 0\farcs85$). The total flux density of the system, i.e. the value at which the curves of growth flatten, are consistent within the errors. The residual scaling correction is smaller for the low resolution data.

3.1.2. *uv-plane analysis*

Due to discrete and sparse uv-coverage, the map produced by the clean algorithm is not uniquely defined, i.e. missing Fourier components must be extrapolated. To avoid such issues altogether, the measurement can be directly performed in the uv-plane. Approaches vary from simpler ones, such as examining 1D amplitudes vs. uv-distance plots (e.g., Hodge et al. 2016), to more complex ones, such as using specialized multi-component fitting software, e.g. UVMULTIFIT (Martí-Vidal et al. 2014), as demonstrated in e.g. Rujopakarn et al. (2019), or fitting morphology models based on Bayesian statistics (see e.g., Pavesi et al. 2016). We employ two techniques: 1) we attempt to fit a 2D Gaussian in the uv-plane using the UVMODELFIT task in CASA, and 2) we analyze amplitudes in radially-averaged annuli of uv-distances (units of kλ). For high S/N data, a poor fit to the visibilities indicates a more complex source morphology (i.e. non-Gaussian), or it can imply additional sources/companions in the field (i.e. the second source is blended together with the main one on the shortest baselines only). One downside of this simple

[10] The convolution kernel is computed from the clean beam only, therefore making its effect on the dirty beam unknown, requiring further quantification.

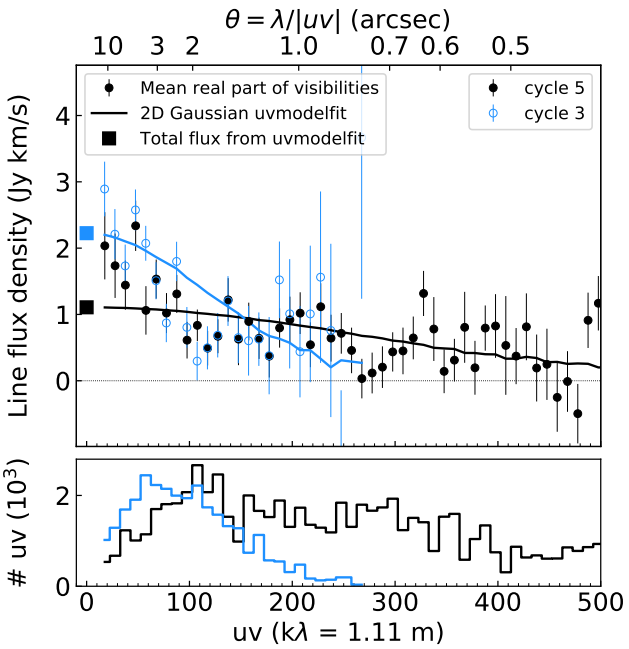

Figure 4. Example flux density measured in annuli of uv-distances of two observation cycles of J1306+0356. The error bars represent the standard error on the mean of visibilities (real part) in a given annulus. The solid lines are single Gaussian fits performed in the uv-plane. Histograms in the lower panel show the number of visibilities available in a given annulus after data averaging. The values on the x-axis are measured in kilowavelengths, where we also provide the value of 1 kλ in meters. The upper x-axis shows the fringe resolution of a given projected baseline length (defined as the observing wavelength divided by the uv-distance). A single Gaussian is a poor model choice in this example, exacerbated by the presence of the companion, resulting in significantly different flux estimates depending on available baseline lengths (i.e. cycle 5 data is dominated by longer baselines and fitting a compact source is preferred).

uv-plane analysis method is that it does not account for the complex morphology encoded in the visibilities' phases, so it is best applied in parallel with the map analysis.

In Figure 4 we show the flux densities measured at specific uv-distances (i.e. projected baseline lengths) of two different observation cycles. Only the real part of the complex visibilities is shown as a proxy for the visibility amplitude[11]. For a source in the phase center, the imaginary part of the visibilities will always have a mean value of zero. We have confirmed that this to be the case for all of our datasets. Histograms in the lower panel show the number of visibilities from different cycles that depend on the number of antennas and the total integration time (all data was averaged to the same interval of 30 s, see Section 2.2.3).

[11] Using only the real part helps to see oscillations around zero flux densities at larger uv-distances, whereas amplitudes are by definition always positive.

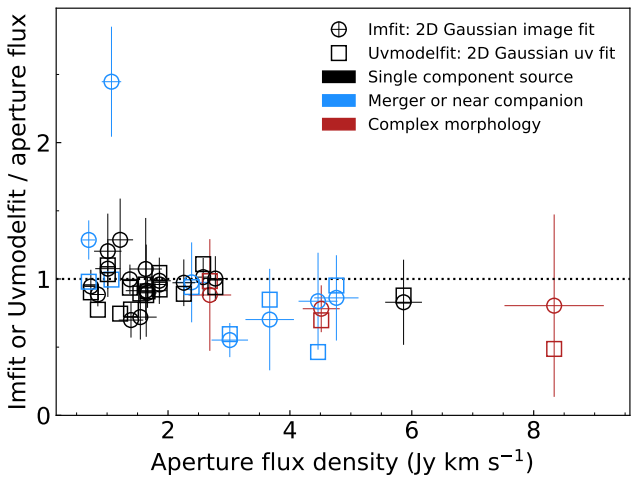

Figure 5. Systematic differences between three different flux measurements (IMFIT, UVMODELFIT and the aperture integration) for all sources. The horizontal dashed line corresponds to our best estimate using the residual scaled aperture measurements. The colored points indicate sources with complex morphologies.

It is expected that a single Gaussian would be a poor representation of this system, which contains two galaxies at close ($\sim$ 5 kpc) separation. Therefore, an attempt to fit a single 2D Gaussian to the visibilities, using the UVMODELFIT task in CASA, yields significantly different models and total fluxes (shown as squares in Figure 4) for the two cycles. The low-resolution data confirms the upturn in visibilities that is only hinted at in the high-resolution data below 100 kλ. In this example it is clear that there is a second galaxy that is causing this upturn. In a different scenario, where no such companion is obvious from the image plane analysis, the interpretation becomes more difficult, as the faint spatially extended emission can have the same signature.

3.1.3. *Comparing flux measurements*

We apply the different methods of measuring the flux density, outlined above, to our high-resolution sample of quasars (sample 1 from Section 2.4), and compare the results in Figure 5. Aperture sizes for individual sources were chosen manually (after visual inspection) to be at a radius where the curve of growth begins to flatten, approaching values consistent with short baseline uv-amplitudes. We limit the 2D Gaussian fitting in the image plane to a circle of a 2″ radius, and apply no additional constraints to uv-plane fits.

Disregarding the outlier with a nearby companion at $\sim$ 1 Jy km s^{-1} (in example source J1306+0356), all of the measurements are consistent within a factor of two, and generally consistent within the errors. Two tentative trends can be observed in Figure 5. Compared to the aperture measurements, image plane fitting using IMFIT (circles in Figure 5) produces larger values at the fainter end (i.e. Gaussian tails apparently account for positive noise peaks), while both image plane based IMFIT fits (circles), and uv-plane based UVMODELFIT (squares) single Gaussian fits, yield generally lower fluxes at

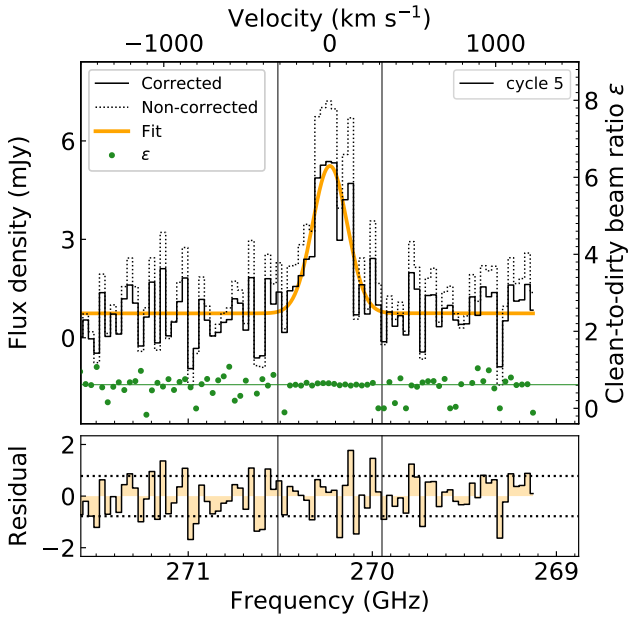

Figure 6. Example of an integrated spectrum, for J1306+0356, extracted from individual channels, using a manually chosen aperture with a radius of 0.″5. The orange line is a single Gaussian plus a constant fit to the corrected spectrum. Green points indicate the clean-to-dirty beam ratios measured in individual channels. The green line shows the best estimate of $\epsilon = 0.61$ (see Section 3.1.1), used to obtain the final corrected spectrum. Residuals from the fit are shown in the lower panel. The dotted horizontal lines outline $\pm 1\sigma$ variations in the residual. Vertical lines mark $\pm 3\sigma_{\rm line}$. With smaller S/N of emission in individual channels, compared to the [C II] intensity map, the effect of the residual scaling is more pronounced (the difference between dotted line and full line spectra), and is larger than 20% in this specific case.

the brighter end, possibly indicating the presence of more extended faint emission. In some of the cases, fitting a single 2D Gaussian is obviously a poor description of the observations, nevertheless we show this simple approach to demonstrate the resulting scatter.

Measurements of both the aperture flux density, and amplitudes at the shortest baselines, can be sensitive to large scale structure, whereas fitting a 2D Gaussian in the image plane will unlikely provide good estimates for the extended faint emission. For this reason we refrain from fitting 2D Gaussians in the image plane in the remainder of the paper.

3.2. *Measuring the spectrum*

3.2.1. *Extracting aperture spectra*

To measure the spectrum, we apply residual scaling on the aperture flux densities obtained from individual channels before a spectrum is extracted. The clean-to-dirty beam ratio, ϵ, can be estimated from the channels with the highest S/N, and applied on the remaining channels, as the beam shape is not expected to vary significantly between neighboring channels. This is demonstrated in Figure 6, where the factor ϵ exhibits the least scatter when measured across the line. If the continuum is absent, ϵ can be measured exclusively over the line (as it requires some clean flux C to calculate). The value of $\epsilon = 0.61$, obtained here, is slightly smaller compared to $\epsilon = 0.65$, derived from the [C II] intensity map (see Figure 2), because the shape of the dirty beam will vary slightly due to different visibilities being averaged in both cases (varying frequency ranges are being averaged). The S/N of the emission is lower inside individual channels compared to a broader intensity map collapse, and with less cleaned emission in smaller velocity bins, the contribution of the residual (uncleaned) flux is larger. This results in a more significant residual-scaling effect, as the corrected value now corresponds to 78% of the non-corrected one (compared to the previous 85% in the intensity map). In this example, a single Gaussian plus a constant fit yields the [C II] line FWHM of $246 \pm 26\,{\rm km\,s^{-1}}$. The residual spectrum after the subtraction of the fit (the lower panel of Figure 6) is consistent with pure noise. There is no indication of an additional broad spectral component.

3.2.2. *Stacking spectra*

To increase the S/N of the spectra and recover possible fainter emission we also perform spectral stacking. Due to the range of [C II] linewidths in our sample (see Table 1) it is necessary to rescale the spectra by their respective widths before stacking. This step ensures that we do not misinterpret the contribution from broader lines as potential outflows in the mean spectrum. Therefore, we resample all spectra before stacking, and present the stacked spectra in the units of velocity divided by $\sigma_{\rm line}$, which has an average value of $150\,{\rm km\,s^{-1}}$.

We perform spectral stacking in two ways. First, we stack individually measured and residual-scaled aperture spectra from the data cube (aperture sizes were chosen manually), where we consider the underlying continuum to be flat, within the bandwidth of interest, and estimated from the fitting of a Gaussian plus constant profile. This constant value[12] for the continuum is subtracted before the addition of the spectra. No additional weights to individual spectra are applied. Second, we perform spectral uv-stacking (as described in Section 2.2.3). To obtain the full spectral cube, we selected velocity slices of $0.2\sigma_{\rm line}$ from individual datasets and imaged them together. This velocity corresponds to a width of $30\,{\rm km\,s^{-1}}$ on average. The uv-data was continuum subtracted prior to stacking and, in contrast to the first method, this continuum subtraction allowed for a spectral slope in the fit. Having two different methods of continuum estimation largely mitigates any potential bias of subtracting a broad component as continuum.

[12] The dust continuum spectral energy distribution is not flat. However, in our case this slope would result in only a few percent difference in flux between the two ends of the observed bandwidth. In a single spectral setup, ALMA provides up to 3.75 GHz of a contiguous bandwidth coverage, which corresponds to $\sim$4000 km s^{-1} at the observed frequency of $\sim$270 GHz, where the [C II] line of $z \sim 6$ galaxies is redshifted.

4. RESULTS

In the following sections we present the recovered spatial extent of both [C II] and the dust continuum, as well as evidence for/against outflows.

4.1. *Spatial extent of the* [C II] *emission*

Using the analysis template presented in the previous section, we performed a visual inspection for each of the 27 quasar host galaxies in our sample (for details see Appendix G). We find that, in most cases, the full extent of the source is difficult to quantify. Our chosen aperture sizes have a mean radius of 4 kpc, but the curves of growth usually hint at additional larger scale emission. Most sources for which the amplitudes are well-fit by a single Gaussian exhibit slightly higher amplitudes at the lowest baselines that probe the largest spatial scales. Fluxes measured between different cycles of individual sources are usually consistent, although discrepancies can reach up to 50% in some cases. This may be attributed to the low S/N of the particular observation, but some remaining calibration issues may also be present. Unfortunately, the S/N available at short baselines is generally insufficient to make any significant claims regarding extended emission in individual sources. Three sources with additional 7 m data show at least 10% larger amplitudes on short baselines compared to the 12 m data. High-resolution imaging ($\sim$1 kpc), can still be successfully used to recover emission on significantly larger scales because the MRS (see Section 2.1) is approximately an order of magnitude larger than the synthesized beam. In summary, visual inspection of individual sources show tentative evidence for extended emission, with apertures larger than the usual 2σ outlining contours necessary to collect this emission.

We investigate the mean extent of this [C II] emission via uv-stacking. The stack was created as described in Section 2.2.3 using the 19 quasar host galaxies in the clean sample (sample 4 from Section 2.4, totaling 32 observation runs). In short, from each dataset, which was first continuum subtracted, we selected frequency channels that encompass a total of $1.2 \times$ FWHM of the [C II] line, and then imaged them all combined, thus obtaining a single uv-stacked [C II] intensity map. The results are presented in Figure 7, where the maps highlight the quality of the combined data and the success of the deconvolution process. The anisotropic 2D structure is the result of stacking galaxies with random inclinations, that might also have faint undetected companions, and is therefore not meaningful.

From the curve of growth in Figure 7, we conclude that the emission extends up to 10 kpc ($\sim 2''$), beyond which we do not recover any significant additional flux. The value at which the cumulative flux saturates (2.5 ± 0.1 Jy km s^{-1}), when corrected for missing tails, is consistent with the mean of the line fluxes measured individually from the spectra. The corrected and the non-corrected flux densities are consistent within the errors, because most of the flux was cleaned, and the clean-to-dirty beam ratio is 0.75–1, up to a radius of $3\farcs5$. The dirty beam sidelobes levels are 4% at most. We report a simple one parameter size estimate, the half-light radius, defined as the radius where the curve of growth equals to 50% of the value at 10 kpc (where the curve reaches saturation). The [C II] emission half-light radius is 1.6 ± 0.1 kpc, and the region within 2, 3, 4, 5 and 10 kpc contains 55%, 65%, 75%, 80% and 100% of the flux, respectively.

We also examine the surface brightness profile of the [C II] emission, shown in Figure 7. The azimuthally-averaged profile measured in the cleaned map (red line) indicates the presence of two components of emission; a steep core, and a broad fainter component. The extended component is detected at $> 3\sigma$ significance up to a radius of 5 kpc (and $> 2\sigma$ up to 8 kpc), and is well described by an exponential function with a scale length of 2 kpc (also shown in Figure 7). The emission drops to 1% of the peak beyond 6 kpc. Here, we did not perform any residual scaling, but instead calculated the mean value of pixels within some annulus, as the corrections are negligible.

We ascertained to what extent the multi-scale cleaning algorithm may bias the shape of the profile by repeating the imaging process with a different and simpler algorithm. The results are presented in Appendix C, and show that, regardless of the clean procedure, we recover the same shape of the surface brightness profile. Furthermore, in Appendix D we fit the brightness profile models directly to the uv-data. The results obtained with this method are in agreement with our conclusions drawn from the image plane analysis. We investigate further possible selection biases by repeating this analysis for several different subsamples in Appendix E and find no systematics.

4.2. *Spatial extent of the dust continuum emission*

We investigate whether the extended emission is specific to the [C II] line by replicating the analysis for the dust-continuum emission. The stacking method is fully described in Section 2.2.3. In short, from each full dataset (no continuum subtraction performed) of the clean sample, we removed a total bandwidth of $2.5 \times$FWHM around the [C II] line, and imaged the remaining channels together, thereby creating the uv-stacked dust continuum map. We show the results in Figure 8. Based on the cumulative aperture flux, the total flux density saturates at 10 kpc, as was the case with the [C II] emission. The final value of 2.05 ± 0.07 mJy is consistent with the mean of individual continuum measurements (see Venemans et al. 2020, accepted) indicating that no object is biasing our stacked results at a significant level. The dirty beam sidelobes levels are 3% at most. The curve of growth is steeper in the core for the continuum, compared to the [C II]. The dust continuum half-light radius, defined as in the previous section, is 0.86 ± 0.03 kpc, and the emission contained within 2, 3, 4, 5 and 10 kpc corresponds to 75%, 80%, 85%, 90% and 100% of the total, respectively.

The compactness of the dust continuum emission is evident in the surface brightness profile, shown in Figure 8. The emission drops to 1% of the peak beyond 4 kpc (compared to 6 kpc in case of [C II]). However, at radii beyond 2 kpc we also find an extended component with a significance of $> 3\sigma$ up to a radius of 5 kpc (and $> 2\sigma$ up to 7 kpc). The expo-

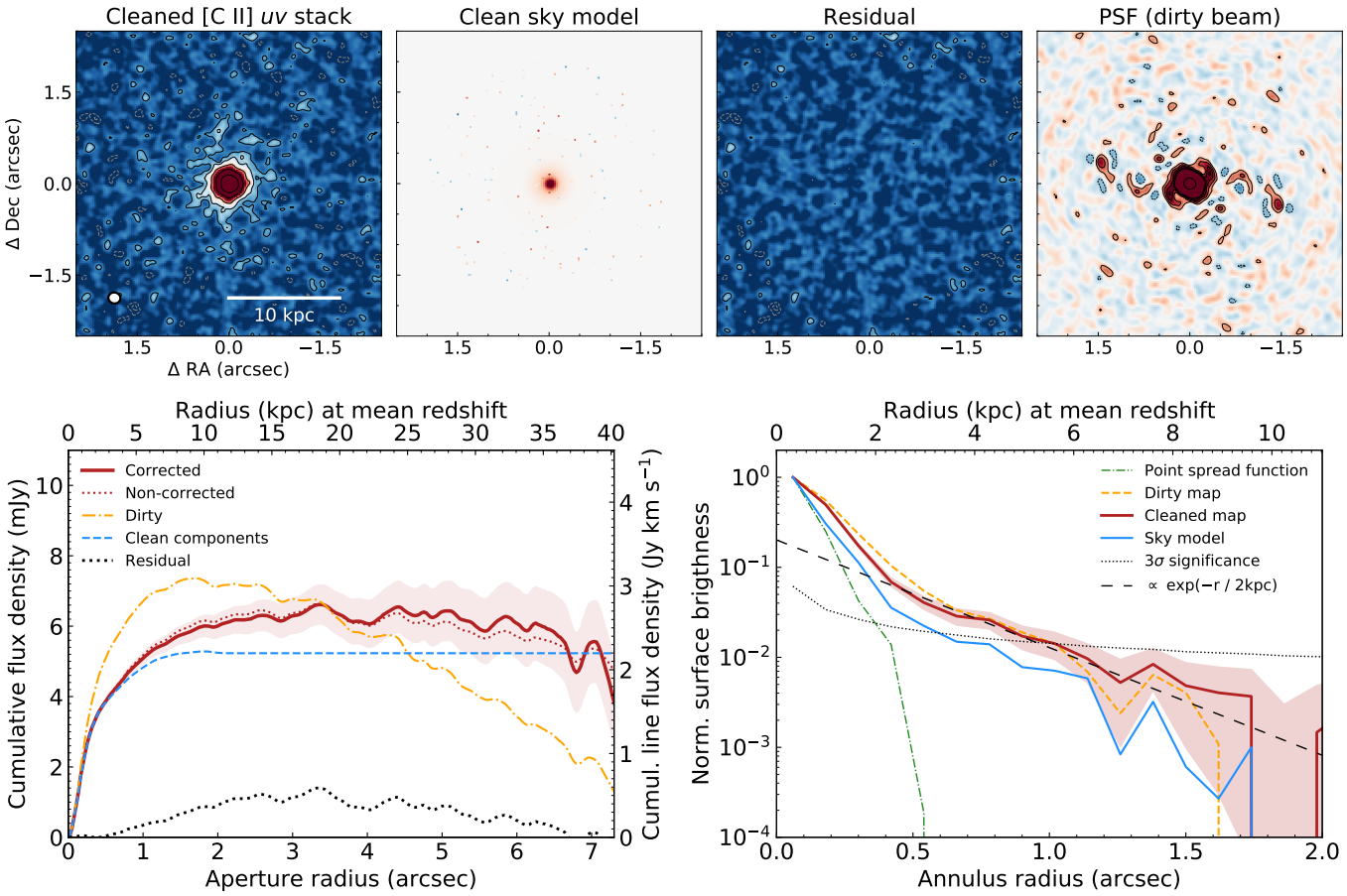

Figure 7. The emission of the [C II] line measured in the uv-stack of the clean sample of $z \gtrsim 6$ quasar host galaxies. *The upper four panels* display the final cleaned map, as well as the multi-resolution clean model, the associated residuals, and the dirty beam (a different deconvolution method is discussed in Appendix C). Contours for the cleaned map and the residual are logarithmic in powers of 2 starting with 2σ, where $1\sigma = 18.8\,\mu\mathrm{Jy\,beam^{-1}}$, and dashed lines represent negative values. The peak of the emission is detected at $\mathrm{S/N} = 60$. Contours for the PSF indicate 1, 2, 3, 4, 5 and 50% levels (the last one defining the beam FWHM of $0\farcs20 \times 0\farcs18$, also shown in the corner of the first panel). *The lower left panel* shows the curves of growth measured in different maps as specified in the legend. The cleaning was performed within a circular mask of $2''$ radius. The right-hand y-axis is scaled assuming an average bandwidth of $420\,\mathrm{km\,s^{-1}}$. The flux values are not corrected for the missing flux outside the bandwidth used (i.e. total fluxes need to be multiplied by 1.19, see Section 3.1.1). The shaded area indicates the 1σ uncertainty on the cumulative flux. *The lower right panel* shows surface brightness in annuli measured in different maps, as specified in the legend, and normalized to the maximum value. The shaded area indicates 1σ error on the mean, computed as the $\mathrm{rms}/\sqrt{N}$, where N is the number of beams contained in the annulus. An exponential profile with a scale length of 2 kpc is shown with the long-dashed line for comparison.

nential function has a scale length consistent with the one we find for the [C II] emission.

In summary, both the [C II] and the dust continuum emission show two-component profiles, where the scale lengths of the extended components are consistent between the two. The central component, which dominates the half-light radius measurement, is more compact in the case of the dust continuum emission.

4.3. *High-velocity outflows*

We turn our attention to the [C II] spectra to investigate the possible presence of broad components that may be indicative of outflows.

A single Gaussian describes the [C II] spectrum well in all of the individual cases (see Appendix G), except in a few that are known for hosting a merger or a companion system (see Section 2.3). The stacked spectrum is also well fit by a single Gaussian, as shown in Figure 9. Here, the peak of the stacked [C II] line is detected at a S/N of 32 in velocity bins of $0.2\sigma_{\mathrm{line}}$. There is no evidence of any broad spectral component. Fitting two Gaussians would collect marginally more emission, only at velocities between $-6\sigma_{\mathrm{line}}$ and $-2\sigma_{\mathrm{line}}$, yielding 3% larger integrated line flux density. One caveat of this stacking approach is that it relies on manually-chosen aperture sizes, which means that all stacked objects do not

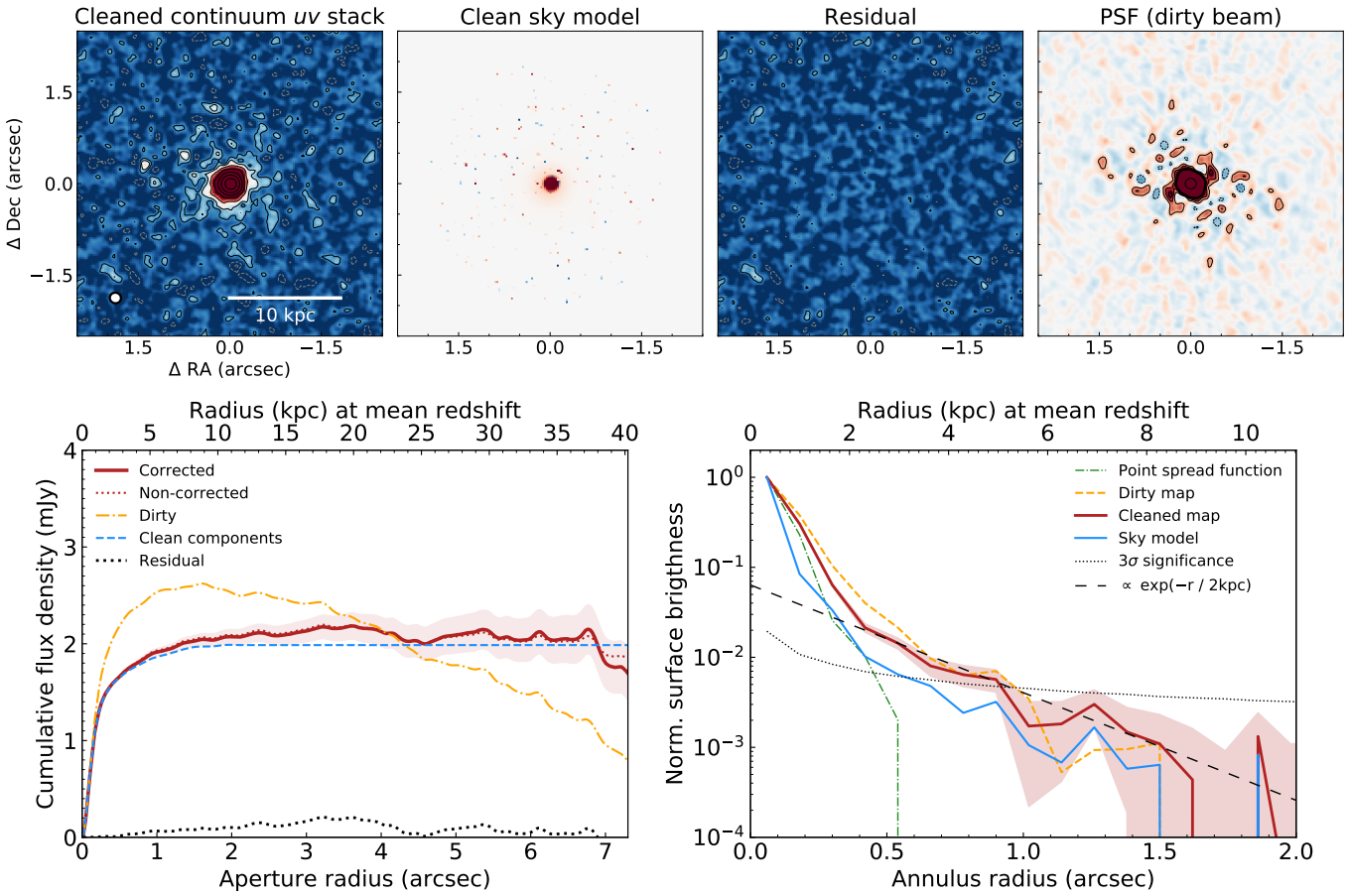

Figure 8. Same as Figure 7, but for the dust continuum emission. The mean observed continuum frequency is 256 GHz ($\approx$ 1.2 mm), the rms in the cleaned map is $1\sigma = 4\,\mu$Jy beam^{-1}, the peak is observed at the S/N of 200, and the resolution is $0\farcs18 \times 0\farcs17$.

contribute equally on all spatial scales. We further improve on the result by performing a uv-plane spectral stack next.

We also find no evidence for a broad spectral component, indicative of a high-velocity outflow, in the [C II] spectrum of the uv-stacked data. The [C II] spectrum drawn from the on the uv-stacked and imaged data cube of the clean sample (sample 4 from the Section 2.4) is presented in the left panels of Figure 10. The spectral uv-stacking procedure is described in detail in Section 2.2.3. In short, from each continuum subtracted dataset we selected channels that accumulate to slices of $0.2\sigma_{\rm line}$. We then imaged together slices that correspond to the same offset from the line peak. There are 100 such slices, covering velocities between $\pm 10\sigma_{\rm line}$.

The rms noise in the velocity range between $\pm 6\sigma_{\rm line}$ is 60 μJy beam^{-1} per channel of $0.2\sigma_{\rm line}$ (which on average corresponds to 30 km s^{-1}). We measured the 2″ radius aperture spectrum (residual corrected) from the uv-stack imaged cube, which should encompass the entire [C II] emission, according to our previous spatial scale analysis. Additionally, we performed a single beam (i.e. single pixel) measurement at the central [C II] peak position following Decarli et al. (2018). These two measurements are both necessary as we do not know the spatial scale of the potential broad spectral component. Outflows are more likely to be dominant at small angular offsets from the quasar, however this is not the case if the high velocity component is due to companion galaxies. The residuals that remain after subtracting a single Gaussian from the spectrum show no evidence of a broad spectral component, either in the core, or on large scales (see Figure 10). The total [C II] flux contained within the single Gaussian fit, taking the $\sigma_{\rm line} = 150$ km s^{-1}, is equal to 2.9 ± 0.1 Jy km s^{-1}, consistent with the value measured from the uv-stacked intensity map (see Section 4.1). The difference in the cumulative flux density across the spectrum between the observed data and the fit is less than 1%. If we measure the [C II] spectrum in an annulus between 1″ and 2″, the spectral shape is equivalent to the single beam and the full aperture measurement.

Even when maximising the available signal, we recover no broad component of emission. We therefore selected and stacked velocity ranges between $\pm(3$ to $6)\sigma_{\rm line}$ in the uv-plane. This velocity range corresponds to 450 – 900 km s^{-1} on average. If a broad spectral component were to exist, its S/N would be maximized in such a map, which is shown in Figure 10. This map has an rms noise level of 10.8 μJy beam^{-1}, and can be considered as a collapse over 900 km s^{-1}. No significant emission is detected, i.e. the map is consistent with pure random noise. To be certain that we

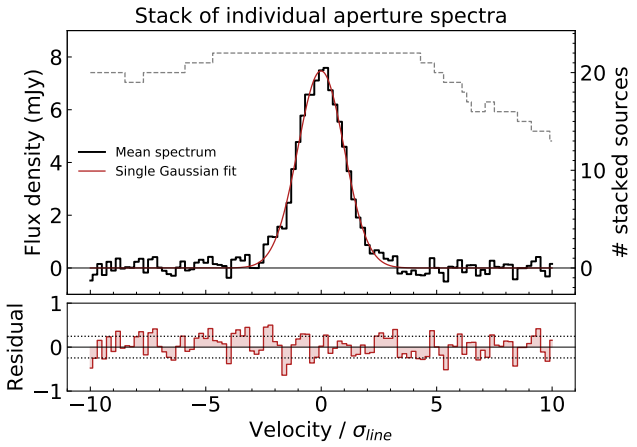

Figure 9. *The upper panel* shows the stack of individual [C II] spectra of the quasar host galaxies, drawn from the clean high-resolution sample (sample 2 from Section 2.4). Prior to stacking, all spectra were rescaled in velocities by their respective linewidths (mean $\sigma_{\rm line} = 150\,{\rm km\,s^{-1}}$), and corrected for a constant continuum. The dashed gray line indicates the number of sources available per channel bin. *The lower panel* shows the residual (in units of mJy) after the subtraction of a single Gaussian. Horizontal dotted lines indicate the 1σ standard deviation in the residuals. A single Gaussian is a good fit to the stacked spectrum.

have not missed any fainter, more extended emission, we repeated the analysis on the full uv-stacking data sample (sample 3 from the Section 2.4). The spectrum and map of these wings are shown in the right panels of Figure 10. The rms noise in the cube is $48\,\mu{\rm Jy\,beam^{-1}}$ per channel, which is 20% lower compared to the more restricted clean sample. The final result remains the same.

We conclude that, on average, $z \gtrsim 6$ quasar host galaxies do not show evidence of broad-linewidth [C II] outflow, in either the aperture spectrum, single beam spectrum, or the extended wings map. However, we cannot rule out the possibility of a very faint outflow, below our stack detection limit. For that purpose we define an upper limit on the outflow emission as follows. By taking the rms level of the wings map, and its bandwidth, we estimate a 3σ upper limit on the outflow in the very core (single $\sim$ 1 kpc beam) of $0.03\,{\rm Jy\,km\,s^{-1}}$, which corresponds to less than 5% of the core peak emission. Alternatively, the spatially integrated [C II] line of average ${\rm FWHM} \sim 350\,{\rm km\,s^{-1}}$ detected at peak ${\rm S/N} \sim 30$ in velocity bins of $\sim 30\,{\rm km\,s^{-1}}$ shows no outflow signatures. If outflows are ubiquitous in high-redshift quasar host galaxies, the flux contained in such a broad component is negligible compared to the main narrow component. Although individual objects may exhibit significant outflows, these galaxies should then be considered as being outliers (we found no such object in our studied sample). We note that a secondary broad spectral feature can be obtained if one does not employ linewidth normalization, as discussed in Appendix F.

5. DISCUSSION

In the following sections we interpret and discuss our results in the context of published work.

5.1. *Spatial extent of the* [C II] *and dust emission*

We find that the dust continuum and the interstellar gas, traced by the [C II] emission, follow the same exponential surface brightness profiles extending between 2 and 10 kpc, as shown in Figure 11, implying that the [C II]-to-FIR ratio[13] is constant at large radii. These similar extents imply that the [C II] emission is tracing the ISM of the galaxies, rather than halos or outflows.

The central ISM component dominates the dust and [C II] emission. The emission within 2 kpc contains 55% (75%) of the total [C II] (dust continuum) flux density with a 20% (10%) contribution from beyond 5 kpc. Such observations are rare for high-redshift quasar host galaxies. For the $z \sim 6.4$ quasar SDSS J1148+5251 Walter et al. (2003, 2009) found a compact [C II] emission embedded in a more extended gas reservoir (traced by the CO line). However, these observations could not recover the faint component probed here via stacking due to sensitivity constraints. Recently, more high-resolution observations were conducted targeting additional quasar host galaxies. These showed that the dust emission is also centrally concentrated, even more so than the [C II] emission (e.g., Wang et al. 2019b; Venemans et al. 2019).

Although their properties differ from those of our $z \sim 6$ quasar host galaxies, the majority of previously observed and simulated star-forming galaxies also appear to exhibit a dominant, compact component of [C II] and/or dust emission (e.g., Olsen et al. 2015; Hodge et al. 2016; Chen et al. 2017; Calistro Rivera et al. 2018; Gullberg et al. 2018, 2019; Cochrane et al. 2019; Rybak et al. 2019). Some high-redshift observations also recover extended faint components of [C II] or dust emission (e.g., Rybak et al. 2019; Gullberg et al. 2019; Fujimoto et al. 2019, 2020b).

The extent of the ISM in our quasar host galaxies, traced by [C II] and dust emission, appears to be consistent with the expected extent of the stellar component. For high-redshift quasar host galaxies, such as our sample, no stellar emission has been detected likely due to the presence of a strong central point source and dust obscuration (see e.g., Mechtley et al. 2012). We therefore compare the [C II] and dust sizes measured here to the predictions from the ZFOURGE survey, see Figure 12. Based on the simulations of Marshall et al. (2019), which probe the properties of quasar host galaxies, we assume that galaxies in our sample have stellar masses of $10^{10.5-11}\,{\rm M_\odot}$. For galaxies of these stellar masses at $z \sim 7$, the stellar mass - size evolution fit by Allen et al. (2017) indicates that our sample should have stellar half-light radii of

[13] This refers to the surface-brightness ratio only. Computing the luminosity ratio would require assuming a dust SED, and a spatial dust temperature distribution, which is unknown for these sources (see accompanying paper, Venemans et al. 2020, accepted).

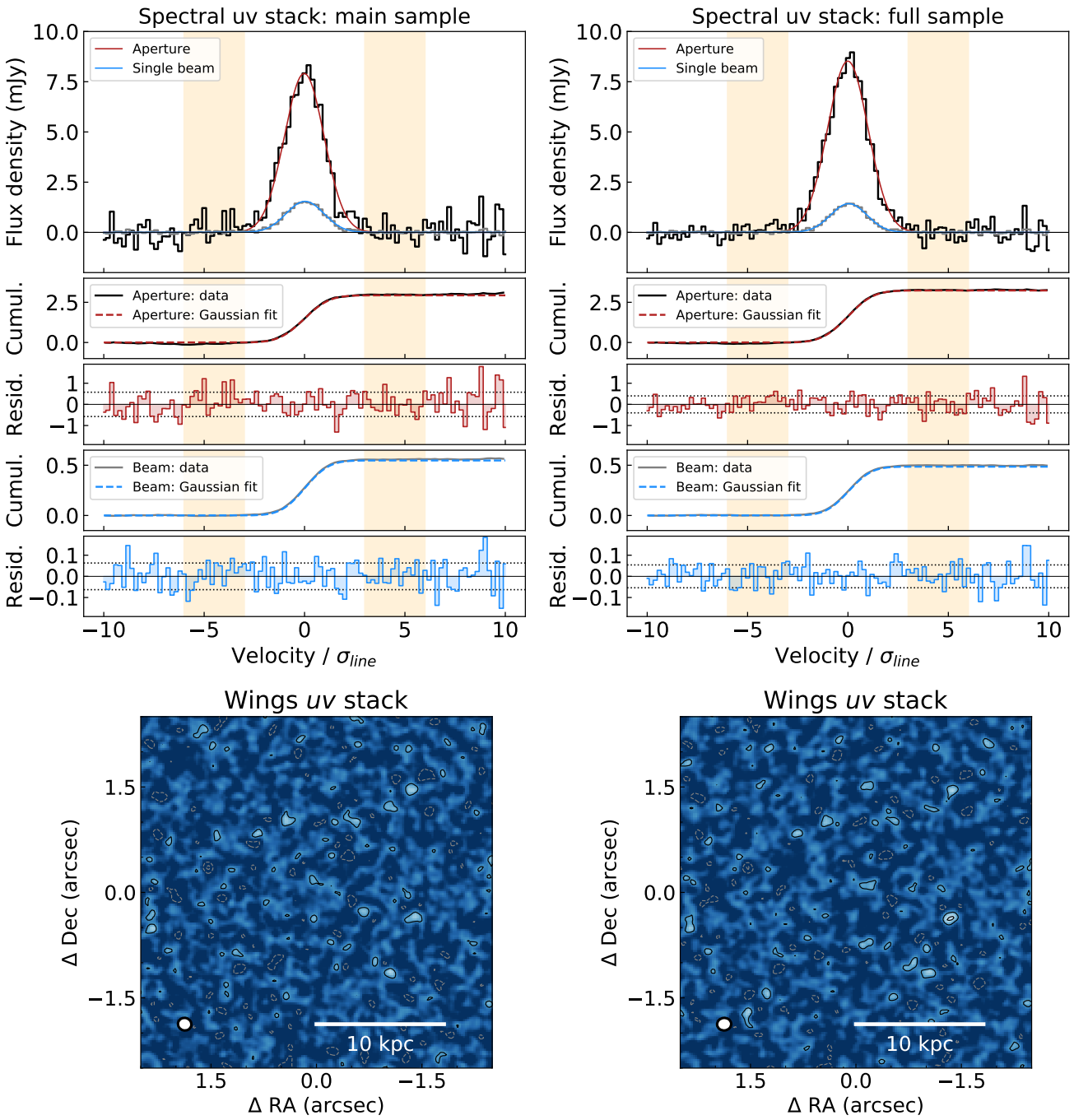

Figure 10. The mean [C II] spectrum of a sample of $z \gtrsim 6$ quasar host galaxies, measured from the uv-stacked and imaged data. *The top panels* show the [C II] spectra in the main clean sample (left) and the full sample (right). The mean linewidth in the sample is $\sigma_{\rm line} = 150\,{\rm km\,s^{-1}}$. The residual-corrected $2''$ radius aperture spectrum is shown with a black line, with the corresponding single Gaussian fit in red. The gray line shows the single beam (i.e. single pixel, the central $\sim$1 kpc) spectrum, with the corresponding single Gaussian fit in blue. The cumulative distribution (in Jy km s^{-1}) and the fitting residual (in mJy) can be seen in the subsequent panels for the aperture and the single beam spectrum, respectively. Horizontal dotted lines in the residual panels show 1σ standard deviation. The orange shaded regions outline the velocity ranges imaged to produce the wings map shown below. We observe no evidence for a broad-line component. *The bottom panels* show the maps obtained by imaging velocity range between $3\sigma_{\rm line}$ and $6\sigma_{\rm line}$ (both positive and negative velocity components combined), to maximize the sensitivity to detect a broad emission line feature. The rms level in the map is 10.3μJy beam^{-1} for the main sample (left) and 8.3μJy beam^{-1} for the full sample (right). Solid (dashed) contours outline $\pm2\sigma$ and $\pm4\sigma$ significant positive (negative) emission. The beam size is shown in the lower-left corner. Again, we find no evidence for a broad emission, that could manifest as, e.g., a peak in the center of the map, or an excess of positive emission at larger annuli.

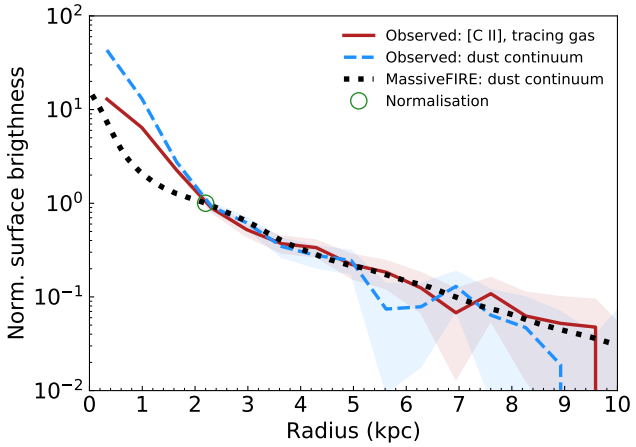

Figure 11. Comparison of average surface brightness profiles, normalized in the region of the extended faint emission. Observed profiles are measured from the uv-stacked sample of $z \gtrsim 6$ quasar host galaxies (clean sample, same profiles as in Figures 7 and 8). The simulated profile is measured in the stack of a sample of massive galaxies at $z \sim 6$, where no SMBH modeling was performed (see Feldmann et al. 2016, 2017). The tails of the profiles have comparable exponential scale lengths in all three cases.

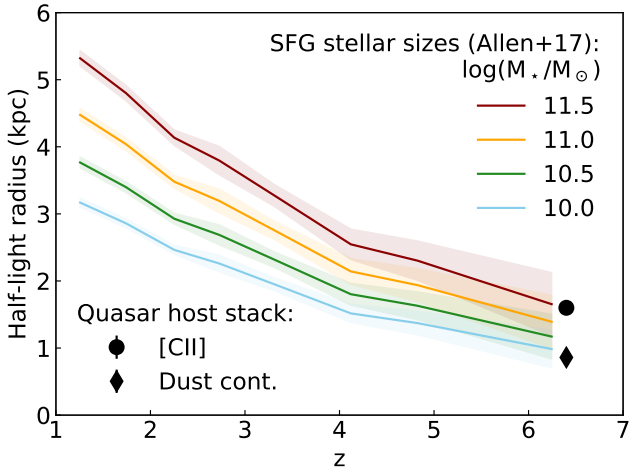

Figure 12. Comparison between the stellar half-light radii of star-forming galaxies (SFGs) from a mass-complete sample (Allen et al. 2017), with different colors referring to specific stellar mass bins, and the two measurements from our stacks of $z \gtrsim 6$ quasar host galaxies. The uncertainties of our measurements correspond only to the errors on the mean half-light radii, and are smaller than the symbols.

1 – 2 kpc, consistent with the [C II] and dust half-light radii. Note that for the $z \sim 4.3$ star-forming galaxy, for which the [C II] was probed at high sensitivity, Neeleman et al. (2020) measure consistent [C II] and stellar half-light radii. Moreover, studies of local galaxies typically find consistent stellar and dust half-light radii, albeit with some scatter depending on the wavelength of the two tracers (e.g., Leroy et al. 2009; Hunt et al. 2015; Casasola et al. 2017).

5.2. *Dust emission from the MassiveFIRE simulation*

We compare our observational data with predictions from the MassiveFIRE suite of cosmological zoom-in simulations (Feldmann et al. 2016, 2017), which is a part of the Feedback in Realistic Environments (FIRE) project (Hopkins et al. 2014). Galaxy formation simulations offer the prospect of studying the properties of massive, high-redshift galaxies with high spatial resolution and from multiple viewing angles.

All simulations start from cosmological initial conditions and account for star formation and various stellar feedback channels such as momentum injection from supernovae, stellar winds, photo-heating, and radiation pressure. Feedback from SMBHs is not included. The physical model employed by these FIRE simulations is described in detail in Hopkins et al. (2014). The spatially resolved UV-to-mm spectral energy distributions (SEDs) are calculated in post-processing via self-consistent dust radiative transfer calculations with the help of SKIRT (Baes et al. 2011; Baes & Camps 2015). A cosmic microwave background temperature floor is included in the calculation. ALMA broadband fluxes are computed by convolving the redshifted SEDs with the ALMA transmission functions. Further details are described in Liang et al. (2019).

In Figure 13, we show several examples of galaxies from the D and E series of the MassiveFIRE suite at $z = 6$. These simulated galaxies reside in halos with virial masses of $(2-6) \times 10^{12}\,\mathrm{M}_\odot$, have star-formation rates (SFRs) of $70 - 400\,\mathrm{M}_\odot\,\mathrm{yr}^{-1}$, and total IR luminosities ($8 - 1000\,\mu$m) of about $10^{12}\,\mathrm{L}_\odot$, comparable to our sample of quasar hosts. We calculate the radial profile of the sub-mm emission by stacking the predicted 1.2 mm ALMA flux density maps of five simulated galaxies, each viewed from 18 random directions, and by averaging the flux densities in radial annuli. As for the observational data, all intrinsic structure (different disk orientations etc.) are smoothed out by stacking the simulated galaxies' emission. The stacked 1.2 mm flux density profile of MassiveFIRE galaxies closely follows both the SFR surface density as well as the dust column density profiles.

The radial dust profile of the stacked MassiveFIRE galaxies closely follows the radial profile of the dust and [C II] emission of our quasar host galaxies at radii beyond 2 kpc (shown in Figure 11), while the core emission differs, most likely due to selection bias. The observed sources host an actively accreting SMBH at their centers, which may be associated with enhanced levels of star formation activity, whereas the simulated galaxies do not, and are simply matched to our sample in FIR luminosity. Although the simulations and observations do not pertain to the same population of galaxies, we believe that the comparison of the extended profiles is valid, assuming that the feedback from the central SMBH does not severely affect the host (gas and dust) at radii beyond 2 kpc. Our findings suggest that the observation of ex-

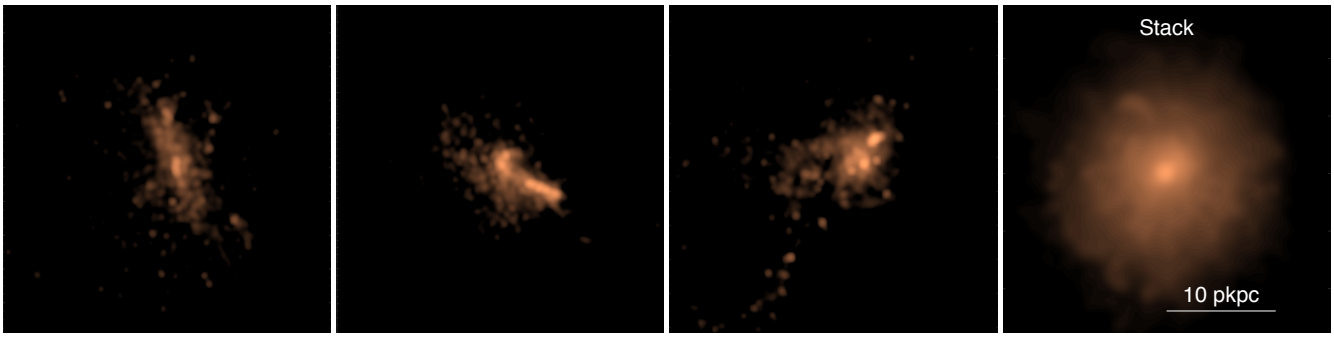

Figure 13. *The first three panels* show the observed FIR emission in ALMA Band 6 (1.2 mm) of three examples galaxies at $z \sim 6$ from the MassiveFIRE simulation. These galaxies have IR luminosities comparable to our quasar host galaxy sample. The maps are smoothed with a Gaussian kernel with FWHM of 1 kpc, and are 30 kpc on the side. *The last panel* shows the stack of 5 galaxies at 18 random viewing angles.

tended sub-mm emission around IR-luminous, high-redshift galaxies may be a consequence of an extended distribution of dust, gas, and active star formation, partially due to the large but irregular morphology of some of these luminous objects. The agreement between simulations and observations suggests that the extended emission seen in the ALMA data are part of the main galaxy's ISM. Furthermore, the lack of correlation between the quasar's bolometric luminosity and the host's FIR emission (see accompanying paper by Venemans et al. 2020, accepted) suggests that the feedback from the active galactic nucleus is not dominating the FIR emission.

5.3. *Outflows*

We find no evidence for a broad [C II] emission line component in the mean spectrum of either the clean or the full quasar host galaxy sample, using multiple analysis approaches, including a novel technique of spectral uv-stacking. Individual objects also show no evidence for broad-line outflows. The same spectral shape, consistent with a single Gaussian, is measured in the central beam (< 1 kpc), the larger aperture spectrum (< 10 kpc), and the extended annulus (5 – 10 kpc) (see also Neeleman et al. 2020, in prep.). A single Gaussian collects 99% of the observed emission, based on the cumulative flux integral. Imaging a wider velocity range, where the broad feature may be expected, also yielded a non-detection. Our results therefore suggest that high-velocity outflows are atypical for $z > 6$ quasar host galaxies. Objects in our sample are the brightest emitters at $z \gtrsim 6$, and we may expect strong quasar feedback based on their bolometric luminosities. However, the lack of outflow signatures in our results suggest that the ISM is not significantly affected by the wind-like feedback in these systems.

The observational evidence for [C II] outflows in $z \sim 6$ quasar host galaxies remains contested. Decarli et al. (2018) found no broad spectral feature in their spectral stack of 23 quasar host galaxy population, although only a relatively short ~ 8 min integration time was available per source. Another stacking analysis, utilizing 26 quasar host galaxies at $z \sim 6$, was performed by Stanley et al. (2019). They normalized all line widths prior to stacking and report a broad component of $> 700\,\mathrm{km\,s^{-1}}$, albeit at less than 1.5σ significance, which reaches a maximum of 2.5σ from a favourably-selected subsample. Their result is therefore formally consistent with a non-detection, which is also supported by the deeper data used in our analysis.

Maiolino et al. (2012) and Cicone et al. (2015) performed a study of the quasar host galaxy SDSS J1148+5251 and presented evidence for a significant outflow identified as a broad spectral component of $\sim 900\,\mathrm{km\,s^{-1}}$, extending out to spatial scales of 30 kpc. To date, these studies represent the only cases where such strong features are reported for a single high-redshift object, which makes SDSS J1148+5251 an outlier in the $z \gtrsim 6$ quasar host galaxy population. A spectral stack of 48 quasar host galaxies beyond $z > 4.5$, performed by Bischetti et al. (2019), also appears to reveal a very broad $\sim 1730\,\mathrm{km\,s^{-1}}$ mean spectral feature (ranging from $700 - 2500\,\mathrm{km\,s^{-1}}$, depending on the subsample), with an integrated flux ratio between the broad and the narrow component of ~ 0.2 (and 0.05 for the peak ratio).

Our findings appear to be in tension with the results of Bischetti et al. (2019), although the studies are of comparable depths. Given the high significance of their detection, and the large percentage of the total flux being in the broad component, we would also expect to see a broad line. This discrepancy could be explained by one or a combination of three factors. The first explanation is that stacking spectra of different linewidths in velocity space without width normalization may change the spectral shape. This effect could be exacerbated in the sample studied in Bischetti et al. (2019), as it contains additional objects at $4.5 < z < 6$ with much larger linewidths of up to a FWHM of $800\,\mathrm{km\,s^{-1}}$ (other aforementioned studies limit themselves to $z \gtrsim 6$ objects). In Appendix F, we demonstrate that the broad wings are identified in the spectra stacked without line width renormalization. However, from the shape of the residual, we argue that this is, at least in part, an artifact resulting from the linear addition of Gaussians with various widths. The second explanation for the discrepancy is that faint or blended companion galaxies that are unresolved at the available resolution could contribute to the broadening of the [C II] line (with the recent addition of new high resolution data, three additional quasar host galaxies in their sample now have confirmed nearby companions, see Section 2.3, besides those already considered by the authors). The third possible issue is

that when computing the continuum for its subtraction, Bischetti et al. (2019) use a zeroth order polynomial on a quarter of the available bandwidth, which ensures that the [C II] line is properly excluded from the fit, but reduces the S/N of the fit.

The quasar host galaxy samples used in Decarli et al. (2018), Bischetti et al. (2019), and Stanley et al. (2019), largely overlap with our work (our studied sources constitute 15/27, 22/48, and 18/26 of the sample in the three mentioned papers, respectively), which makes the range of different reported results puzzling. We emphasize that our study is building upon the above mentioned studies by analyzing previously unavailable $\sim$1 kpc resolution data, which allows us to better identify merging or companion systems, as well as to resolve extended features, if present. However, a weak link of the analysis in all of the present studies lies in securely separating the continuum from the potential broad spectral component. The trade-off between selecting velocities far enough from the [C II] line, and the resulting S/N of the continuum fit is likely the greatest culprit causing inconsistent results. In this paper we employed two methods of continuum subtraction. The first one depends on extracting the spectrum from the full (line plus continuum) cube, and fitting it with a Gaussian plus a constant line. The second one relies on the UVCONTSUB task in CASA to fit the continuum component in the visibilities using a slope and disregarding on average $\sim 900\,\mathrm{km\,s^{-1}}$ worth of data centered on the [C II] line. Both methods yielded the same result, i.e. a non-detection. The only way to properly address this issue would be to observe much wider bandwidths around the line. In addition, there are multiple effects inherent to interferometric data that could bias the measurements if not properly accounted for. Image plane based stacking will result in a beam that is difficult to interpret, and the final map will be a complex combination of dirty residuals, each with a different PSF, and cleaned components. We attempted to mitigate these issues by performing stacking in the uv-plane, and applying the residual scaling method to account for the subtleties (see Section 3.1.1) inherent to any cleaned map.

6. CONCLUSIONS

We have investigated the extent of the interstellar gas and dust in $z \gtrsim 6$ quasar host galaxies using ALMA observations of the [C II] emission line and the underlying continuum. Our sample consists of 27 objects observed at the high resolution of $\sim$ 1 kpc, with lower resolution archival data available for 20 objects. We apply the novel approach of uv-plane spectral stacking in order to mitigate the issue of beam matching and uncleaned flux, which can become a significant hindrance in image-based stacking methods. This improved the S/N and the fidelity of our analysis. Depending on the subsample under investigation, we can accumulate up to 34 h (with 25.5 h on the 12 m array) of on-source time. From the individual and stacked analysis we conclude the following.

1. Quasar host galaxies at $z \gtrsim 6$ are characterized by a central (< 2 kpc) concentration of [C II] and dust emission, with mean half-light radii of 1.6 ± 0.1 kpc and 0.86 ± 0.03 kpc, respectively. By stacking we recover a region of fainter [C II] and dust emission that extends out to 10 kpc, which is well described by an exponential function with a scale length of 2 kpc. Beyond 5 kpc, the mean surface brightness is observed at values less than 1 – 2% of the peak, and contributes 10 – 20% to the total measured flux density.

2. The large spatial extent of the dust continuum matches that of the [C II], suggesting that we are observing the gas and dust of the interstellar medium of the quasar host galaxies extending out to 10 kpc. The extended dust surface brightness profile follows the same trend as obtained from the simulations of massive galaxies (MassiveFIRE) at similar far-infrared luminosities as our sample, further strengthening our interpretation.

3. We find no evidence of a high-velocity ($\gtrsim 500\,\mathrm{km\,s^{-1}}$) [C II] component, that may be indicative of outflows, in our $z \gtrsim 6$ quasar host galaxy sample. The mean velocity profile of the [C II] line is consistent with a single Gaussian, with less than 1% flux excess on both small (< 2 kpc), and large (< 10 kpc) spatial scales. This finding suggests that the ISM in these extremely active objects is not significantly affected by wind-like feedback.

Our findings imply that the extended emission is due to the extended ISM in the host galaxies, and that no additional component (e.g. a galaxy halo, or a broad-line outflow) needs to be invoked to interpret the measurements.

We thank the anonymous referee for providing helpful suggestions that improved the clarity of the paper. Ml.N., B.P.V., F.W., Ma.N., and A.B.D. acknowledge support from the ERC Advanced Grant 740246 (Cosmic Gas). M.K. acknowledges support from the International Max Planck Research School for Astronomy and Cosmic Physics at Heidelberg University (IMPRS-HD). R.F. acknowledges financial support from the Swiss National Science Foundation (grant no 157591). MassiveFIRE simulations were run with resources provided by the NASA High-End Computing (HEC) Program through the NASA Advanced Supercomputing (NAS) Division at Ames Research Center, proposal SMD-14-5492. Additional computing support was provided by HEC allocations SMD-14-5189, SMD-15-5950, and NSF XSEDE allocations AST120025, AST150045.

This paper makes use of the following ALMA data:
ADS/JAO.ALMA#2012.1.00240.S,
ADS/JAO.ALMA#2012.1.00882.S,
ADS/JAO.ALMA#2013.1.00273.S,
ADS/JAO.ALMA#2015.1.00399.S,
ADS/JAO.ALMA#2015.1.00606.S,
ADS/JAO.ALMA#2015.1.00692.S,
ADS/JAO.ALMA#2015.1.00997.S,
ADS/JAO.ALMA#2015.1.01115.S,

ADS/JAO.ALMA#2016.1.00544.S,
ADS/JAO.ALMA#2016.A.00018.S,
ADS/JAO.ALMA#2017.1.00396.S,
ADS/JAO.ALMA#2017.1.01301.S,
ADS/JAO.ALMA#2018.1.00908.S.
ALMA is a partnership of ESO (representing its member states), NSF (USA) and NINS (Japan), together with NRC (Canada), NSC and ASIAA (Taiwan), and KASI (Republic of Korea), in cooperation with the Republic of Chile. The Joint ALMA Observatory is operated by ESO, AUI/NRAO and NAOJ.

This research makes use of CASA (McMullin et al. 2007), Astropy,[14] a community-developed core Python package for Astronomy (Astropy Collaboration et al. 2013; Price-Whelan et al. 2018), Matplotlib (Hunter 2007), and Topcat (Taylor 2005) .

REFERENCES

Allen, R. J., Kacprzak, G. G., Glazebrook, K., et al. 2017, ApJL, 834, L11, doi: 10.3847/2041-8213/834/2/L11

Arnaboldi, M., Neeser, M. J., Parker, L. C., et al. 2007, The Messenger, 127, 28

Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33, doi: 10.1051/0004-6361/201322068

Bañados, E., Venemans, B. P., Decarli, R., et al. 2016, ApJS, 227, 11, doi: 10.3847/0067-0049/227/1/11

Bañados, E., Venemans, B. P., Mazzucchelli, C., et al. 2018, Nature, 553, 473, doi: 10.1038/nature25180

Bañados, E., Novak, M., Neeleman, M., et al. 2019a, ApJL, 881, L23, doi: 10.3847/2041-8213/ab3659

Bañados, E., Rauch, M., Decarli, R., et al. 2019b, ApJ, 885, 59, doi: 10.3847/1538-4357/ab4129

Baes, M., & Camps, P. 2015, Astronomy and Computing, 12, 33, doi: 10.1016/j.ascom.2015.05.006

Baes, M., Verstappen, J., De Looze, I., et al. 2011, ApJS, 196, 22, doi: 10.1088/0067-0049/196/2/22

Bischetti, M., Maiolino, R., Carniani, S., et al. 2019, A&A, 630, A59, doi: 10.1051/0004-6361/201833557

Calistro Rivera, G., Hodge, J. A., Smail, I., et al. 2018, ApJ, 863, 56, doi: 10.3847/1538-4357/aacffa

Carilli, C. L., & Walter, F. 2013, Annual Review of Astronomy and Astrophysics, 51, 105, doi: 10.1146/annurev-astro-082812-140953

Casasola, V., Cassarà, L. P., Bianchi, S., et al. 2017, A&A, 605, A18, doi: 10.1051/0004-6361/201731020

Chambers, K. C., Magnier, E. A., Metcalfe, N., et al. 2016, arXiv e-prints, arXiv:1612.05560. https://arxiv.org/abs/1612.05560

Chen, C.-C., Hodge, J. A., Smail, I., et al. 2017, ApJ, 846, 108, doi: 10.3847/1538-4357/aa863a

Cicone, C., Maiolino, R., Gallerani, S., et al. 2015, A&A, 574, A14, doi: 10.1051/0004-6361/201424980

Cochrane, R. K., Hayward, C. C., Anglés-Alcázar, D., et al. 2019, MNRAS, 488, 1779, doi: 10.1093/mnras/stz1736

De Rosa, G., Decarli, R., Walter, F., et al. 2011, ApJ, 739, 56, doi: 10.1088/0004-637X/739/2/56

De Rosa, G., Venemans, B. P., Decarli, R., et al. 2014, ApJ, 790, 145, doi: 10.1088/0004-637X/790/2/145

Decarli, R., Walter, F., Venemans, B. P., et al. 2017, Nature, 545, 457, doi: 10.1038/nature22358

—. 2018, ApJ, 854, 97, doi: 10.3847/1538-4357/aaa5aa

Decarli, R., Dotti, M., Bañados, E., et al. 2019, ApJ, 880, 157, doi: 10.3847/1538-4357/ab297f

Fan, X., Strauss, M. A., Becker, R. H., et al. 2006, AJ, 132, 117, doi: 10.1086/504836

Farina, E. P., Arrigoni-Battaia, F., Costa, T., et al. 2019, ApJ, 887, 196, doi: 10.3847/1538-4357/ab5847

Feldmann, R., Hopkins, P. F., Quataert, E., Faucher-Giguère, C.-A., & Kereš, D. 2016, MNRAS, 458, L14, doi: 10.1093/mnrasl/slw014

Feldmann, R., Quataert, E., Hopkins, P. F., Faucher-Giguère, C.-A., & Kereš, D. 2017, MNRAS, 470, 1050, doi: 10.1093/mnras/stx1120

Ferrara, A., Vallini, L., Pallottini, A., et al. 2019, MNRAS, 489, 1, doi: 10.1093/mnras/stz2031

Fujimoto, S., Oguri, M., Nagao, T., Izumi, T., & Ouchi, M. 2020a, ApJ, 891, 64, doi: 10.3847/1538-4357/ab718c

Fujimoto, S., Ouchi, M., Ferrara, A., et al. 2019, ApJ, 887, 107, doi: 10.3847/1538-4357/ab480f

Fujimoto, S., Silverman, J. D., Bethermin, M., et al. 2020b, arXiv e-prints, arXiv:2003.00013. https://arxiv.org/abs/2003.00013

Gullberg, B., Swinbank, A. M., Smail, I., et al. 2018, ApJ, 859, 12, doi: 10.3847/1538-4357/aabe8c

Gullberg, B., Smail, I., Swinbank, A. M., et al. 2019, MNRAS, 490, 4956, doi: 10.1093/mnras/stz2835

Hodge, J. A., Swinbank, A. M., Simpson, J. M., et al. 2016, ApJ, 833, 103, doi: 10.3847/1538-4357/833/1/103

Hollenbach, D. J., & Tielens, A. G. G. M. 1999, Reviews of Modern Physics, 71, 173, doi: 10.1103/RevModPhys.71.173

Hopkins, P. F., Kereš, D., Oñorbe, J., et al. 2014, MNRAS, 445, 581, doi: 10.1093/mnras/stu1738

Hunt, L. K., Draine, B. T., Bianchi, S., et al. 2015, A&A, 576, A33, doi: 10.1051/0004-6361/201424734

Hunter, J. D. 2007, Computing In Science & Engineering, 9, 90, doi: 10.1109/MCSE.2007.55

[14] http://www.astropy.org

Jiang, L., McGreer, I. D., Fan, X., et al. 2016, ApJ, 833, 222, doi: 10.3847/1538-4357/833/2/222
Jorsater, S., & van Moorsel, G. A. 1995, AJ, 110, 2037, doi: 10.1086/117668
Lawrence, A., Warren, S. J., Almaini, O., et al. 2007, MNRAS, 379, 1599, doi: 10.1111/j.1365-2966.2007.12040.x
Leroy, A. K., Walter, F., Bigiel, F., et al. 2009, AJ, 137, 4670, doi: 10.1088/0004-6256/137/6/4670
Liang, L., Feldmann, R., Kereš, D., et al. 2019, MNRAS, 489, 1397, doi: 10.1093/mnras/stz2134
Maiolino, R., Gallerani, S., Neri, R., et al. 2012, MNRAS, 425, L66, doi: 10.1111/j.1745-3933.2012.01303.x
Marshall, M. A., Ni, Y., Di Matteo, T., et al. 2019, arXiv e-prints, arXiv:1912.03428. https://arxiv.org/abs/1912.03428
Martí-Vidal, I., Vlemmings, W. H. T., Muller, S., & Casey, S. 2014, A&A, 563, A136, doi: 10.1051/0004-6361/201322633
Matsuoka, Y., Strauss, M. A., Kashikawa, N., et al. 2018, ApJ, 869, 150, doi: 10.3847/1538-4357/aaee7a
Mazzucchelli, C., Bañados, E., Venemans, B. P., et al. 2017, ApJ, 849, 91, doi: 10.3847/1538-4357/aa9185
McMullin, J. P., Waters, B., Schiebel, D., Young, W., & Golap, K. 2007, in Astronomical Data Analysis Software and Systems XVI, ed. R. A. Shaw, F. Hill, & D. J. Bell, Vol. 376, 127
Mechtley, M., Windhorst, R. A., Ryan, R. E., et al. 2012, ApJL, 756, L38, doi: 10.1088/2041-8205/756/2/L38
Morganson, E., De Rosa, G., Decarli, R., et al. 2012, AJ, 143, 142, doi: 10.1088/0004-6256/143/6/142
Mortlock, D. J., Warren, S. J., Venemans, B. P., et al. 2011, Nature, 474, 616, doi: 10.1038/nature10159
Neeleman, M., Prochaska, J. X., Kanekar, N., & Rafelski, M. 2020, Nature, 581, 269, doi: 10.1038/s41586-020-2276-y
Neeleman, M., Venemans, B., Walter, F., & Novak, M. 2020, in prep.
Neeleman, M., Bañados, E., Walter, F., et al. 2019, ApJ, 882, 10, doi: 10.3847/1538-4357/ab2ed3
Novak, M., Bañados, E., Decarli, R., et al. 2019, ApJ, 881, 63, doi: 10.3847/1538-4357/ab2beb
Olsen, K. P., Greve, T. R., Narayanan, D., et al. 2015, ApJ, 814, 76, doi: 10.1088/0004-637X/814/1/76
Pavesi, R., Riechers, D. A., Capak, P. L., et al. 2016, ApJ, 832, 151, doi: 10.3847/0004-637X/832/2/151
Price-Whelan, A. M., Sipőcz, B. M., Günther, H. M., et al. 2018, AJ, 156, 123, doi: 10.3847/1538-3881/aabc4f
Rujopakarn, W., Daddi, E., Rieke, G. H., et al. 2019, ApJ, 882, 107, doi: 10.3847/1538-4357/ab3791
Rybak, M., Calistro Rivera, G., Hodge, J. A., et al. 2019, ApJ, 876, 112, doi: 10.3847/1538-4357/ab0e0f
Stanley, F., Jolly, J. B., König, S., & Knudsen, K. K. 2019, A&A, 631, A78, doi: 10.1051/0004-6361/201834530
Taylor, M. B. 2005, Astronomical Society of the Pacific Conference Series, Vol. 347, TOPCAT & STIL: Starlink Table/VOTable Processing Software, ed. P. Shopbell, M. Britton, & R. Ebert, 29
Vallini, L., Ferrara, A., Pallottini, A., & Gallerani, S. 2017, MNRAS, 467, 1300, doi: 10.1093/mnras/stx180
Venemans, B., Walter, F., Neeleman, M., et al. 2020, accepted, ApJ
Venemans, B. P., McMahon, R. G., Warren, S. J., et al. 2007, MNRAS, 376, L76, doi: 10.1111/j.1745-3933.2007.00290.x
Venemans, B. P., Neeleman, M., Walter, F., et al. 2019, ApJL, 874, L30, doi: 10.3847/2041-8213/ab11cc
Venemans, B. P., Walter, F., Decarli, R., et al. 2017a, ApJ, 845, 154, doi: 10.3847/1538-4357/aa81cb
—. 2017b, ApJ, 851, L8, doi: 10.3847/2041-8213/aa943a
Walter, F., Brinks, E., de Blok, W. J. G., et al. 2008, AJ, 136, 2563, doi: 10.1088/0004-6256/136/6/2563
Walter, F., Riechers, D., Cox, P., et al. 2009, Nature, 457, 699, doi: 10.1038/nature07681
Walter, F., Bertoldi, F., Carilli, C., et al. 2003, Nature, 424, 406, doi: 10.1038/nature01821
Wang, F., Wang, R., Fan, X., et al. 2019a, ApJ, 880, 2, doi: 10.3847/1538-4357/ab2717
Wang, F., Yang, J., Fan, X., et al. 2018, ApJL, 869, L9, doi: 10.3847/2041-8213/aaf1d2
Wang, R., Wagg, J., Carilli, C. L., et al. 2013, ApJ, 773, 44, doi: 10.1088/0004-637X/773/1/44
Wang, R., Shao, Y., Carilli, C. L., et al. 2019b, ApJ, 887, 40, doi: 10.3847/1538-4357/ab4d4b
Willott, C. J., Bergeron, J., & Omont, A. 2015, ApJ, 801, 123, doi: 10.1088/0004-637X/801/2/123
Willott, C. J., Delorme, P., Reylé, C., et al. 2010, AJ, 139, 906, doi: 10.1088/0004-6256/139/3/906
Yang, J., Wang, F., Fan, X., et al. 2020, ApJL, 897, L14, doi: 10.3847/2041-8213/ab9c26
York, D. G., Adelman, J., Anderson, John E., J., et al. 2000, AJ, 120, 1579, doi: 10.1086/301513

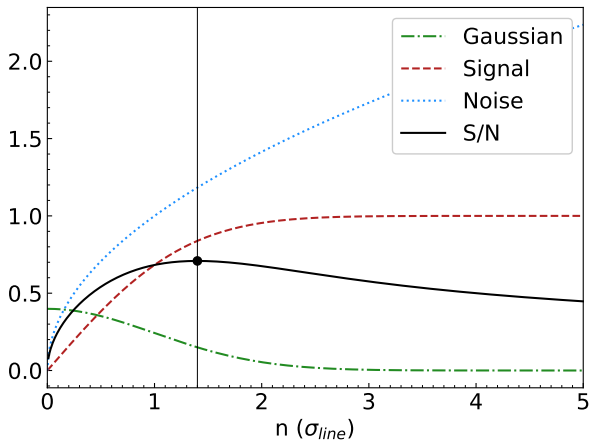

Figure 14. The cumulative signal and noise obtained by integrating a Gaussian profile over an interval of $[-n\sigma_{\rm line}, +n\sigma_{\rm line}]$. The y-axis scaling for both the signal and the noise curve is arbitrary, and the value of n corresponding to the S/N peak does not depend on it.

APPENDIX

A. OPTIMIZING GAUSSIAN LINE DETECTION

In order to calculate the optimal bandwidth across which to integrate a line, we first assume that the emission line has a single Gaussian profile with a line width $\sigma_{\rm line}$, and that the noise in individual velocity bins is Gaussian and not correlated between neighboring channels. Integrating from the line peak outwards, the collected flux grows as $\mathrm{Signal} = F_{\rm line} \times \mathrm{erf}(n/\sqrt{2})$, where $F_{\rm line}$ is the total line flux and n represents the number of $\sigma_{\rm line}$ to integrate over in both positive and negative velocity directions. The uncertainties from individual velocity channels are added in quadrature, therefore the noise increases with the number of channels in the sum as $\mathrm{Noise} = \sqrt{(2n\sigma_{\rm line})/\Delta v} \times \sigma_{\mathbf{rms}}$, where $2n\sigma_{\rm line}$ is the total integration width and $\sigma_{\mathbf{rms}}$ is the noise measured in a velocity bin $\Delta v\,(\mathrm{km\,s^{-1}})$ wide. From this, it follows that $S/N \propto \mathrm{erf}(n/\sqrt{2})/\sqrt{n}$. Numerical evaluation of this function yields the maximum value at $n = 1.4$, see Figure 14. Therefore, the total integration width that optimizes the S/N of a Gaussian line is equal to $2.8\sigma_{\rm line}$. For a Gaussian, the $\mathrm{FWHM} = 2\sqrt{2\ln 2}\sigma_{\rm line}$, and the total width translates to $\approx 1.2 \times \mathrm{FWHM}$, which accounts for 84% of the line flux.

B. RESIDUAL SCALING

The derivation of the residual scaling method is rooted in the fact that we should be able to measure the true emission independent of the cleaning threshold. We denote this true flux as F. If we clean the map down to two different thresholds, the following must hold $F = C + \epsilon R = C_2 + \epsilon R_2$, where C and R refer to aperture fluxes measured in the clean components and the residual maps, respectively (the subscript refers to a different cleaning threshold). The factor ϵ is the clean beam to dirty beam volume ratio which ensures that the summands have equal units (hence the name residual scaling). A special case is obtained if no cleaning is applied, then $C_2 = 0$ and $R_2 = D$, where D is the aperture flux measured inside the dirty map (i.e. the residual after zero iterations of clean). Solving the two equations yields $\epsilon = C/(D - R)$ and $F = \epsilon D$. The essence of the method lies in measuring the proper dirty beam volume, valid only inside a specific aperture. Because the integral of the dirty beam approaches zero (sum of finite number of cosine waves over all area), its volume is only meaningful inside a finite spatial region. This is not the case with the clean beam, which is a 2D Gaussian whose integral always converges to a finite non-zero value.

C. CLEANING WITHOUT MULTI-SCALES

The process of deconvolution, i.e. cleaning, relies on extrapolating the sky model from available visibilities, where the missing Fourier components must be somehow filled in by the algorithm. Therefore, the final sky model is only one of many possible representations that are consistent with the observed data. Throughout this work we utilize the multi-scale clean algorithm as described in Section 2.2.3. If we are using extended source sky model components, the question arises whether we are actually forcing the algorithm to indeed create extended structures. To quantify this effect we re-imaged the uv-stack from Section 4.1 without multi-scales, using only delta functions to populate the sky model. The results are shown in Figure 15.

The final cleaned map shows no significant differences compared to the multi-scale version, and the dirty beam does not change as imaging weights were kept the same. However, the clean sky model is now comprised of individual pixel sources (delta functions), and there is considerably more flux in the residual on large scales (all below 2σ). Nevertheless, both the curves

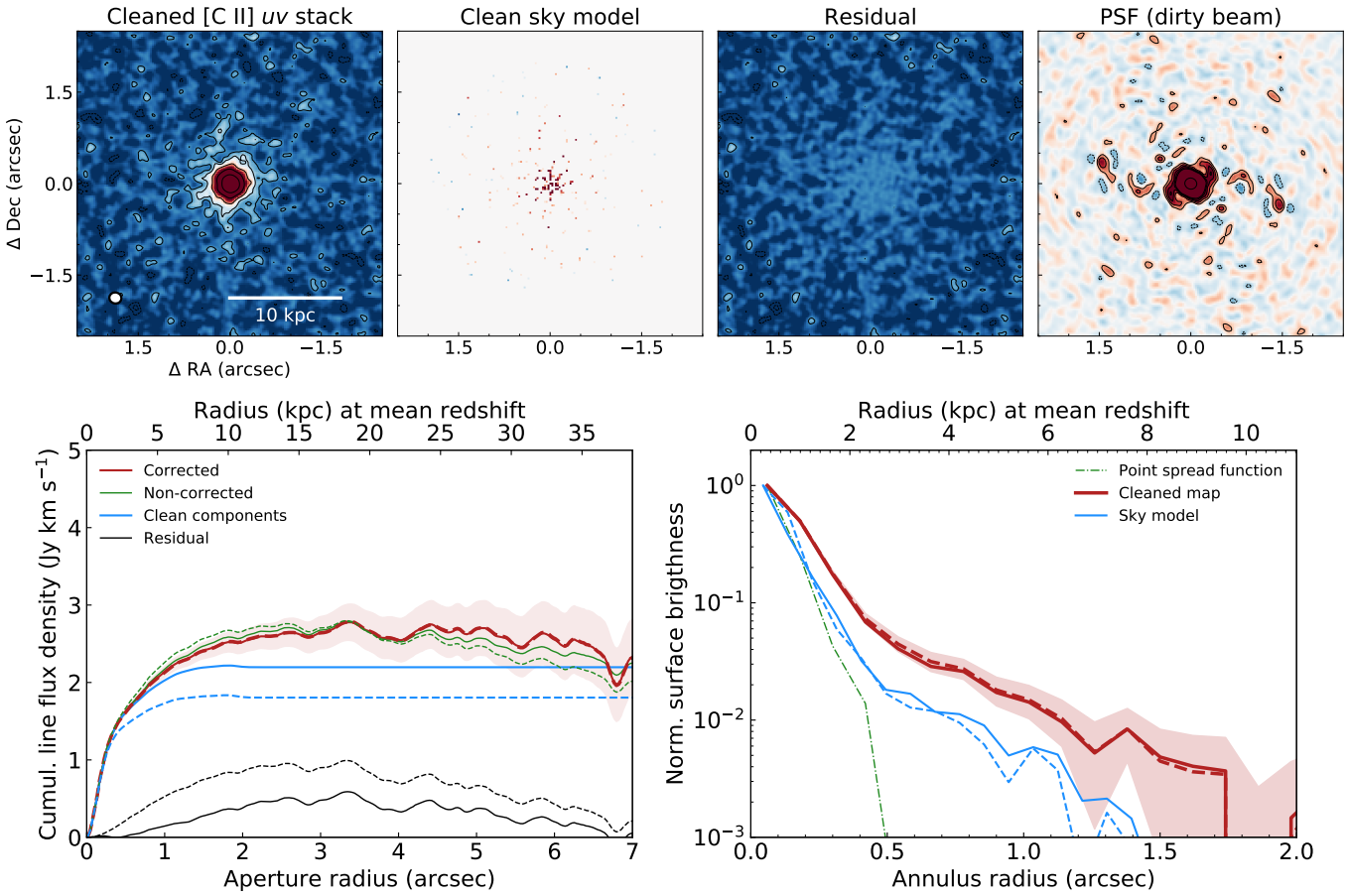

Figure 15. Emission of the [C II] line measured from the imaged uv-stack, similar to Figure 7, but with a different cleaning algorithm utilized. *Upper panels* show the results of deconvolution without multiple scales. *Lower panels* show the comparison between multi-scale cleaned maps (solid lines) and single-scale cleaned maps (dashed lines). Despite obviously different ratios of cleaned to non-cleaned flux, and different sky models, the final corrected flux measurements do not depend on the cleaning strategy.

of growth (after applying residual scaling correction) and surface brightness profiles are consistent between the two cleaning approaches.

D. FITTING THE STACKED DATA IN THE UV-PLANE

We fit the uv-data of our clean uv-stacking sample using the software package UVMULTIFIT (Martí-Vidal et al. 2014). The fitting model was a linear combination of two radially symmetric exponential brightness profiles (Lorentzians in the uv-plane), each with four free parameters: two positional offsets, the total flux, and the major axis FWHM (related to the effective or half-light radius as $R_{\rm eff} = 1.678 \times l$, where l is the scale length of the exponential profile, $l = {\rm FWHM}/(2\ln 2)$). We compare the visibilities binned in uv-distances of 5 kλ and the fitted model in Figure 16. There is no spatial offset between the two fitted components, but they are both shifted by ~ 17 mas with respect to the phase center, which was chosen to be the centroid position of the [C II] emission for each individual source during the uv-stacking process.

The fit yielded the following flux densities, F, and exponential scale lengths, l, for the compact and the extended components: $F^{\rm comp.}_{\rm [C\,II]} = 2.8 \pm 0.1$ mJy with $l = 0.43 \pm 0.02$ kpc, $F^{\rm ext.}_{\rm [C\,II]} = 3.4 \pm 0.2$ mJy with $l = 2.6 \pm 0.2$ kpc, $F^{\rm comp.}_{\rm dust} = 1.19 \pm 0.03$ mJy with $l = 0.20 \pm 0.01$ kpc, and $F^{\rm ext.}_{\rm dust} = 0.76 \pm 0.03$ mJy with $l = 1.3 \pm 0.1$ kpc. We fitted a second model with one less free parameter, by choosing to fix the scale length of one exponential function to 2 kpc, the value derived from image plane analysis (see Sections 4.1 and 4.2). This fit is shown in Figure 16 with the dashed line, and is still consistent with the observed data points. Fitting in the uv-plane circumvents issues with the clean/dirty beam, and yields the conclusions consistent with those presented in the main text.

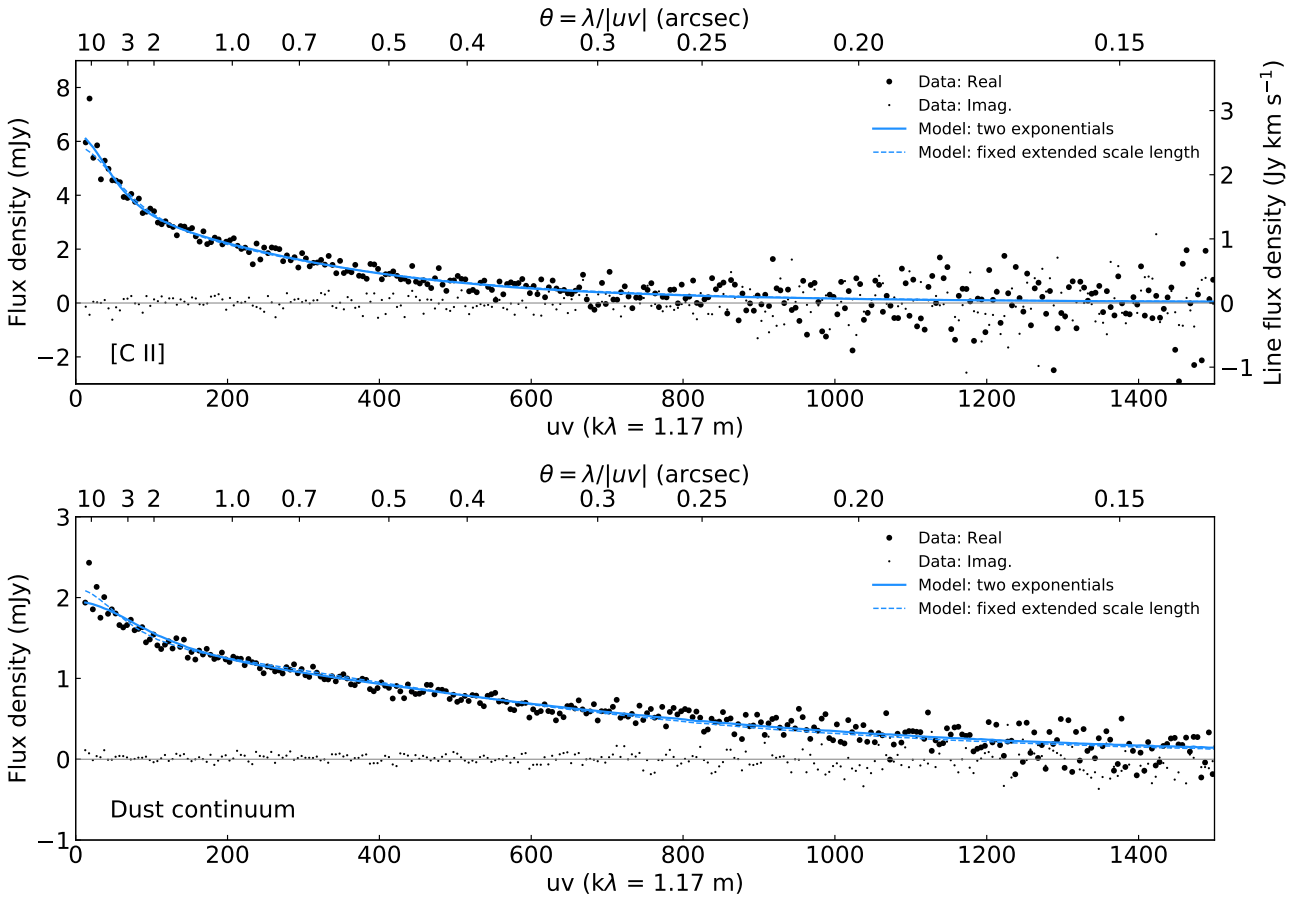

Figure 16. Weighted average flux densities in annuli of uv-distances, drawn from the stacked clean sample of $z \gtrsim 6$ suasar host galaxies (sample 4 from the Section 2.4). The average velocity bandwidth of 420 km s^{-1} is used to scale the right y-axis of the [C II] measurements. The real part of the visibilities is used as a proxy for the amplitudes, while the imaginary part is centered around zero, indicating that the emission is in the phase center. The solid blue line indicates the best fit two-component model obtained with the UVMULTIFIT package. The dashed blue line shows the second model, where an additional constraint was imposed, by fixing one exponential scale length to 2 kpc. Both models are consistent with the data within scatter.

E. SUBSAMPLES AND BIASES

In the main text, for uv-stacking, we have been using the clean sample, in which all systems with close companions and obvious mergers were removed, in addition to any 7 m array data. Here, we evaluate the robustness of our results by considering smaller and larger data samples.

We first enlarge our clean sample by considering systems with companion galaxies and available 7 m array observations. The 7 m array data are provided by the ALMA compact array (ACA), where smaller dishes allow for denser spacing of antennas, thus allowing measurements of visibilities on shorter baselines compared to the 12 m array. There are only three quasar host galaxies with such observations available in our sample, and one of them is designated as a companion system as well. Hence, the largest spatial scales data in the uv-stack will be provided by only two to three objects. Including them in the stack could potentially bias the mean emission measurement. To investigate this effect we define four different samples, which we analyze and present separately in Figure 17. The four tested samples are:

(a) Clean uv-stacking sample (19 objects, 32 datasets), as in the main text.
(b) Clean uv-stacking sample with 7 m data added (19 objects, 34 datasets).
(c) Clean uv-stacking sample with companion and merger systems added (26 objects, 45 datasets).
(d) Full dataset[15] containing companion and merger systems, and all the 7 m data (26 objects, 48 datasets).

[15] Six datasets are always excluded from uv-stacking (see Table 1).

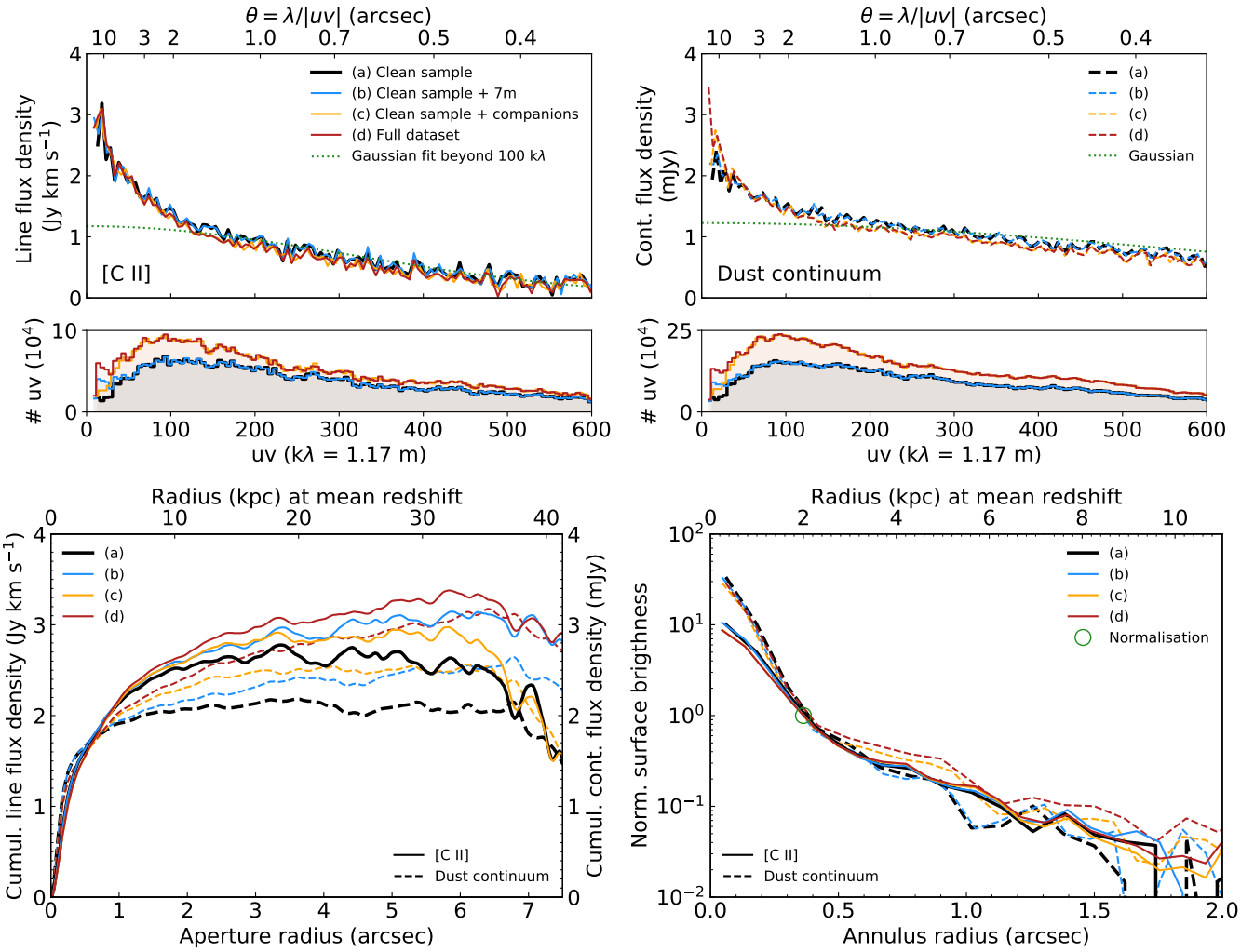

Figure 17. Comparison between the [C II] (solid lines) and the dust continuum $\nu_{\rm obs} = 256$ GHz emission (dashed lines) for different data samples. *The upper left panel* shows the mean [C II] flux in annuli of uv-distances, as well as the histogram of available number of visibilities in each bin. Excess amplitudes at short uv-distances indicate possible extended emission. *The upper right panel* is same as the left one, but measured on the dust continuum. *The lower left panel* shows the curves of growth. Both the continuum and [C II] saturate at similar radii (subsamples a and c). Including the 7 m data requires integrating over larger radii (subsamples b and d). *The lower right panel* shows normalized surface brightness profiles. All subsamples demonstrate a more compact dust continuum component, compared to [C II], and an extended tail in both cases. As expected, including companions and merger systems increases the flux measured in the 3 – 5 kpc range (more pronounced in the case of the dust continuum, see dashed lines of subsamples c and d).

The two top panels of Figure 17 demonstrate that amplitudes cannot be described by a single Gaussian in any of the samples, for both [C II] and dust continuum emission. Between different samples, the dust continuum measurements show somewhat larger dispersion at short baselines than the [C II] values. The 7 m array data double the number of available visibilities below 30 kλ (blue vs. black and red vs. yellow lines). The curves of growth shown in the lower left panel of Figure 17 differ for the four sets of data considered. The inclusion of companion systems increases the mean flux, as one would expect, but the value still saturates at ∼10 kpc. The inclusion of the 7 m data requires integration over larger radii to encompass the data from the shortest baselines and results in ∼20% larger flux densities of both the [C II] and the dust continuum. Because this contribution arises from three or fewer sources, it likely does not represent the mean behavior of full sample of observed $z \sim 6$ quasar host galaxies. At small radii the residual scaling is numerically stable, but because of division with a small number (aperture flux densities measured in the residual and the dirty map approach each other), it becomes numerically unstable beyond 7″. The last panel of Figure 17 shows the surface brightness profiles, normalized at 2 kpc. All four samples exhibit similar functional forms.

For the remainder of this section we analyze smaller subsamples. Our full sample contains sources that span approximately one order of magnitude in [C II] line flux densities (∼1 – 10 Jy km s^{-1}) and galaxies on the fainter end of this range could have

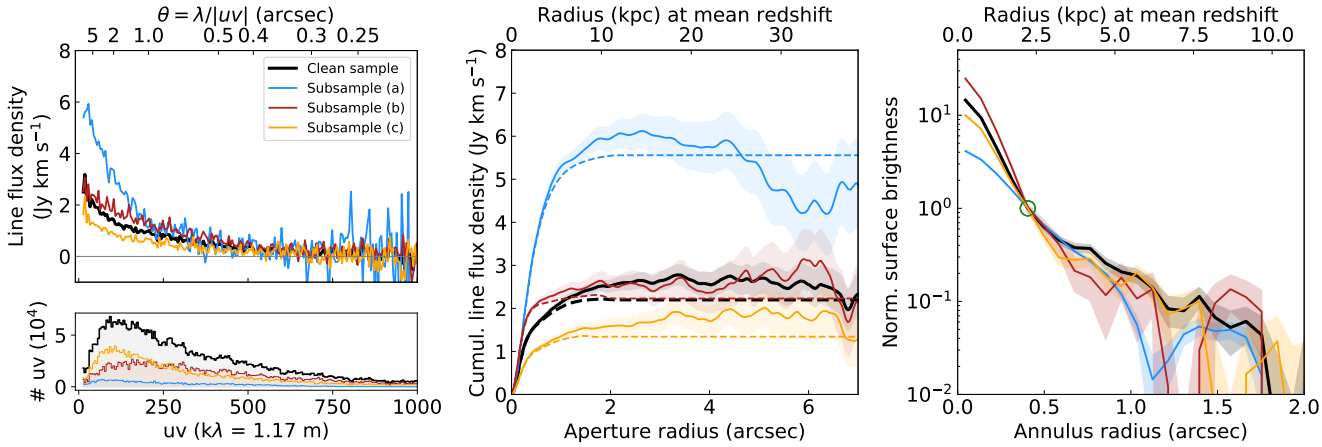

Figure 18. Analysis of [C II] emission for smaller subsamples (three colored lines as listed in the legend), compared to the the clean sample (black line). *The left panel* shows line flux as a function of uv-distance, and the number of visibility data points for each bin in the histogram below. *The middle panel* shows the curves of growth measured in the maps (solid lines), and the values from the clean components only (dashed lines). *The right panel* shows the surface brightness profile normalized at the radius indicated with the green circle. In all of the cases, the [C II] curves of growth saturate around ∼10 kpc, and a two component surface brightness profile is recovered.

intrinsically different properties from those on the brighter end. Also, the addition of objects at various flux levels in the uv-plane can translate into erroneous spatial structure. To quantify possible biases we split our clean sample into three smaller subsamples, based on the flux values measured at the lowest baselines (first available bin of 15 kλ):

(a) Bright subsample, [C II] flux density above 3 Jy km s^{-1} (3 objects, 5 datasets).
(b) Medium bright subsample, between 2 – 3 Jy km s^{-1} (8 objects, 14 datasets).
(c) Faint subsample, below 2 Jy km s^{-1} (8 objects, 13 datasets).

These are presented in Figure 18. Because we are significantly reducing our sample size this way, the S/N becomes poorer. Nevertheless, all our main conclusions hold within the errors. The spatial structure is non-Gaussian, the total [C II] flux saturates at ∼10 kpc, and surface brightness profiles exhibit extended exponential tails. The bright subsample is composed of only three sources, namely J025–33, P009–10, and P183+05, which all show extended morphology and significant emission within 5 kpc (for details on individual objects see Appendix G), thus the surface brightness profile at smaller radii is less steep then the remaining two subsamples.

As a final consistency check we ensured a consistent overlap in uv-coverage between all objects by selecting only the high-resolution cycles (sample 1 from Section 2.4). The analysis performed on such a selection yielded no new systematics. In summary, our main results regarding the spatial extent of [C II] and dust continuum emission remain unchanged across all considered samples.

F. VELOCITY STACKS

Stacking spectra that are individually well described by single Gaussians of different linewidths will result in a spectral shape with at least one narrow and one broad component. To demonstrate this effect, we repeated the [C II] stacking from Section 4.3 without line-width renormalization, in velocity space, and present the results in Figure 19. After subtracting a single fitted Gaussian from the stacked spectrum, positive residuals can be seen around velocities $\sim$ 400 km s^{-1} in both the central beam and the larger aperture measurement. The total observed flux density excess compared to a single Gaussian fit remains low, less than 4%, as evident from the cumulative distribution in Figure 19. The specific shape of the residuals (a peak at zero velocity, followed by one negative and one broader positive feature toward larger velocities) can also be reproduced by averaging pure Gaussian functions, one per object, that correspond to measured [C II] lines (in terms of their total fluxes and FWHMs), and then subtracting a single Gaussian fit from the average. The similar shape of the residual is also visible in Figure A.2 from Bischetti et al. (2019), indicating that at least some excess emission is potentially due to averaging spectra of various line widths.

G. INDIVIDUAL SOURCE ANALYSIS

In this section we provide more details on individual objects. We presented multiple [C II] flux diagnostics throughout Section 3. We show similar analysis for every object of our sample in Figure 20. We also highlight several particular galaxies as potential interesting follow-up candidates. Three objects, listed in Section 2.3 as having complex morphology, deserve a mention here. System J0100+2802 exhibits a factor of two increase in [C II] flux between 3 and 10 kpc. This source was excluded

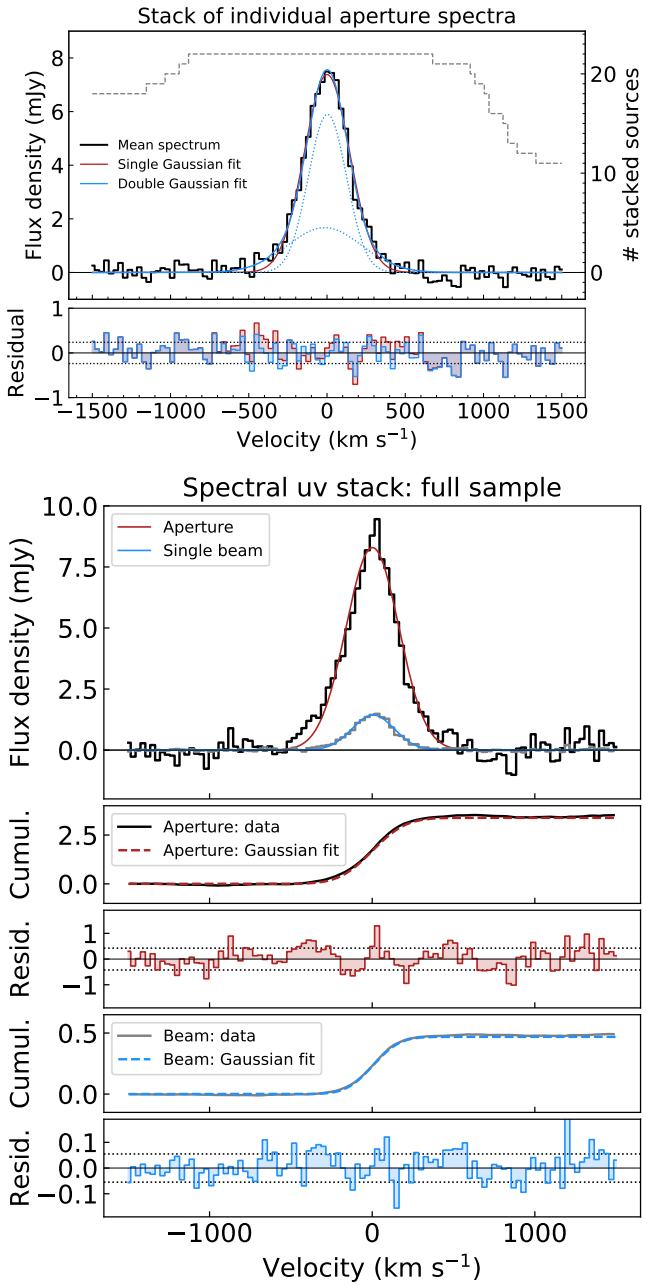

Figure 19. Stacked [C II] spectrum without linewidth normalization. *Top:* Aperture spectra stacking, same as Figure 9. *Bottom:* uv-plane spectral stacking, same as right panels in Figure 10, with the cumulative flux distribution units changed to Jy km s^{-1}, whereas the residuals remain in the units of mJy.

from our uv-stacking analysis and its interpretation is unclear. It was studied in detail by Wang et al. (2019a), and a possible lensing scenario was proposed by Fujimoto et al. (2020a). Systems J025–33 and P009–10[16] are both among the brightest in our sample and show a non-Gaussian amplitude distribution in the uv-plane. However, it is unclear whether a companion is present, or whether we observe more of the underlying extended emission due to higher surface brightness. Another object where we observe a factor of ~ 2 increase in flux densities between the central 3 kpc and extended 10 kpc is J0842+1218, as evident in two independent cycles. This system is already known for having a companion galaxy at a separation of 50 kpc (see Decarli et al.

[16] We note that the large discrepancy between fluxes measured in two cycles is due to partial line coverage in the earlier cycle, where only half of the [C II] line is observed.

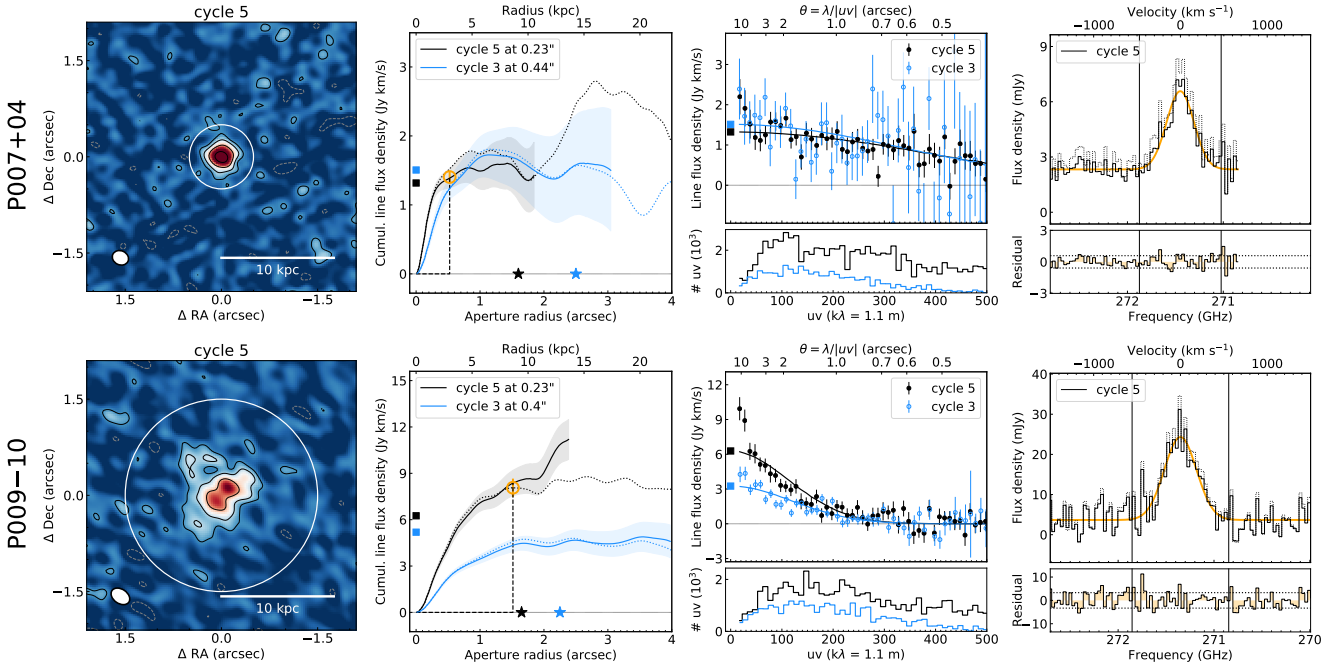

Figure 20. Emission line flux diagnostics of individual sources drawn from the full sample of 27 quasar host galaxies. *The first panel:* [C II] intensity map imaged from the nominal high-resolution dataset using $1.2 \times$ FWHM of the line width. Contours are logarithmic starting from 2σ in powers of 2. Synthesized beam FWHM is shown in the corner. Manually chosen aperture used to extract the spectrum is shown with the white circle. *The second panel:* Curves of growth measured from the intensity maps of various cycles as indicated in the legend. Full lines show residual corrected fluxes, dotted lines show uncorrected values. Dashed line corresponds to the chosen aperture size. Stars on the horizontal zero line indicate the MRS of the dataset (color coded by cycle). Squares on the zero radius line are the same as in the third panel (color coded by cycle). The orange circle is the flux measured from the 1D Gaussian fit of the spectrum from the fourth panel multiplied by 0.84 to offset the missing tails from the intensity map collapse. *The third panel:* Flux measured in annuli of uv-distances of various cycles is shown with points. Solid line represents the model corresponding to a single 2D Gaussian fit obtained with the UVMODELFIT task. Squares at $uv = 0$ show the total flux estimated from the fit. Histograms in the lower panel show the number of visibilities available in a given annulus after data averaging. The upper x-axis shows the resolution of a given baseline length. *The fourth panel:* Spectrum extracted from the manually chosen aperture (same as in the first panel). Full lines show residual corrected fluxes, dotted lines show uncorrected values. Orange line is a single Gaussian fit to the corrected spectrum, with the residual shown in the lower panel (also in units of mJy). Dotted horizontal lines outline $\pm 1\sigma$ variations in the residual. Vertical lines outline $\pm 3\sigma$ of the line width to assist in the search of possible broad components.

2017) and a second, fainter companion at 31 kpc (Neeleman et al. 2019). Similarly, J1044–0125 shows evidence for enhanced emission on large scales, which is particularly interesting as it is hinted at in the 7 m data. Moreover, P359–06 shows some evidence of extended emission in two separate cycles, although the effect is not as pronounced as in the previous two sources. Finally, we note that P183+05 and P007+04 are identified as proximate damped Lyα absorption systems (see Bañados et al. 2019b; Farina et al. 2019, respectively), which may have some effect on our measurements if the foreground galaxy is aligned with the quasar host.

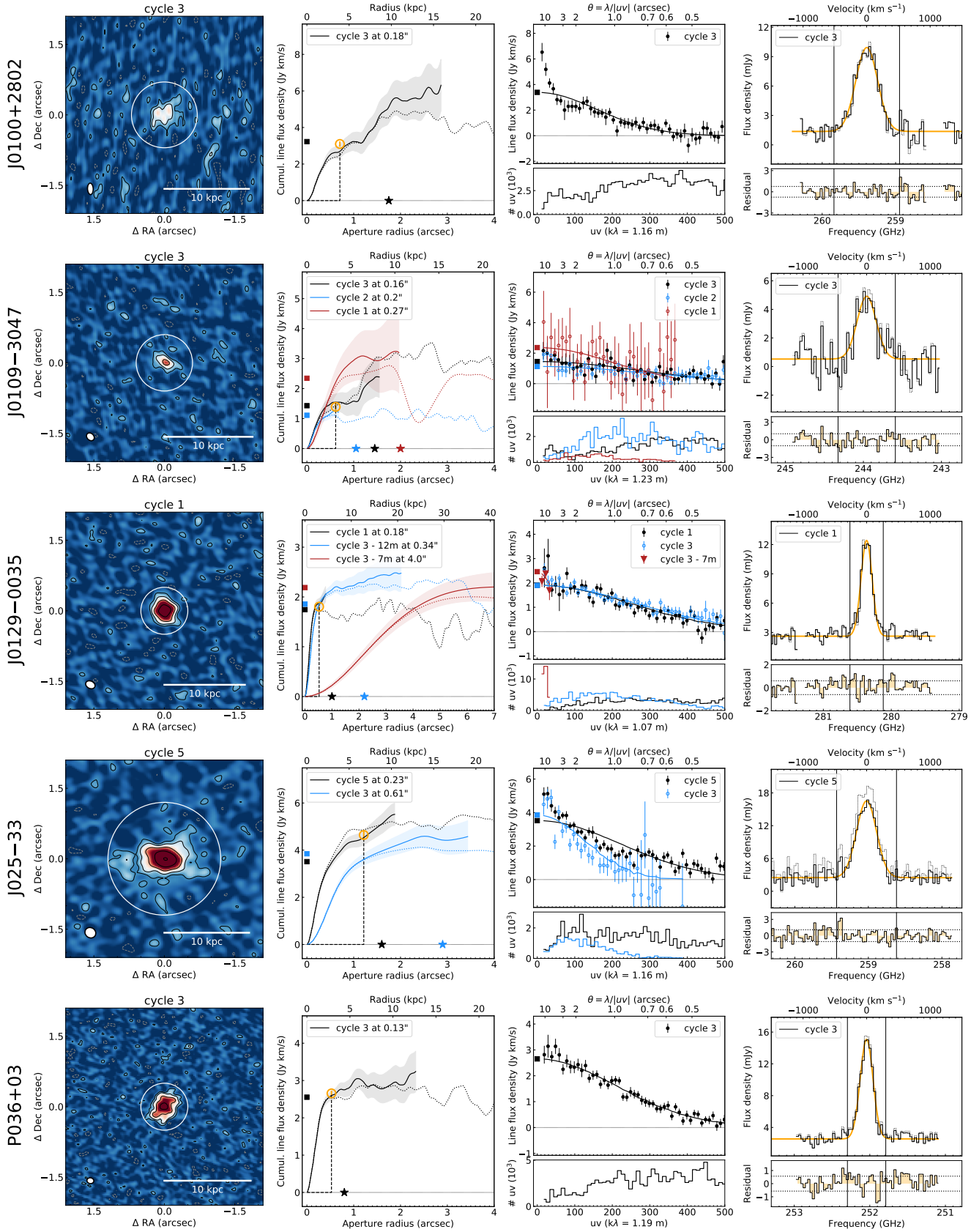

Figure 20. Continued.

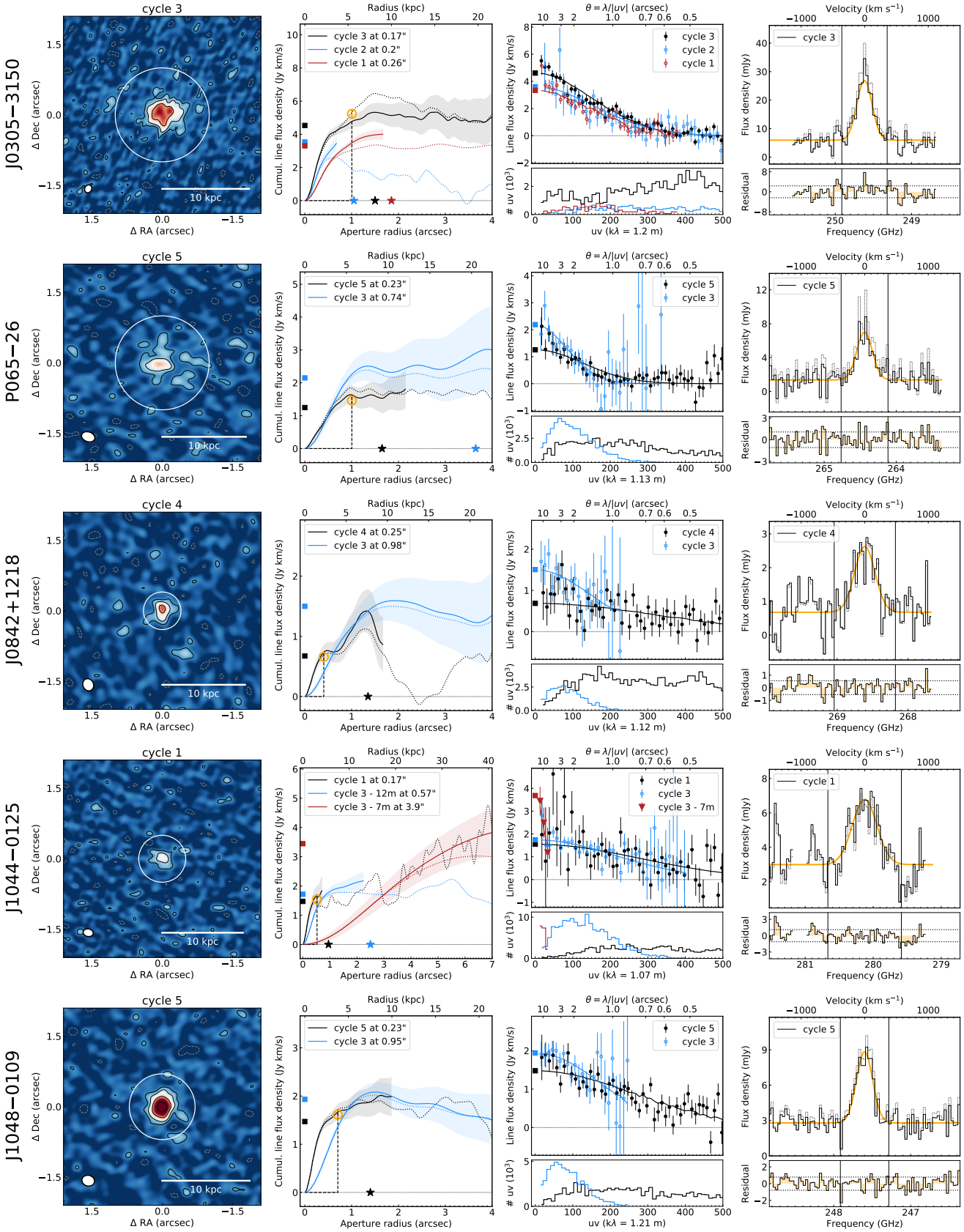

Figure 20. Continued.

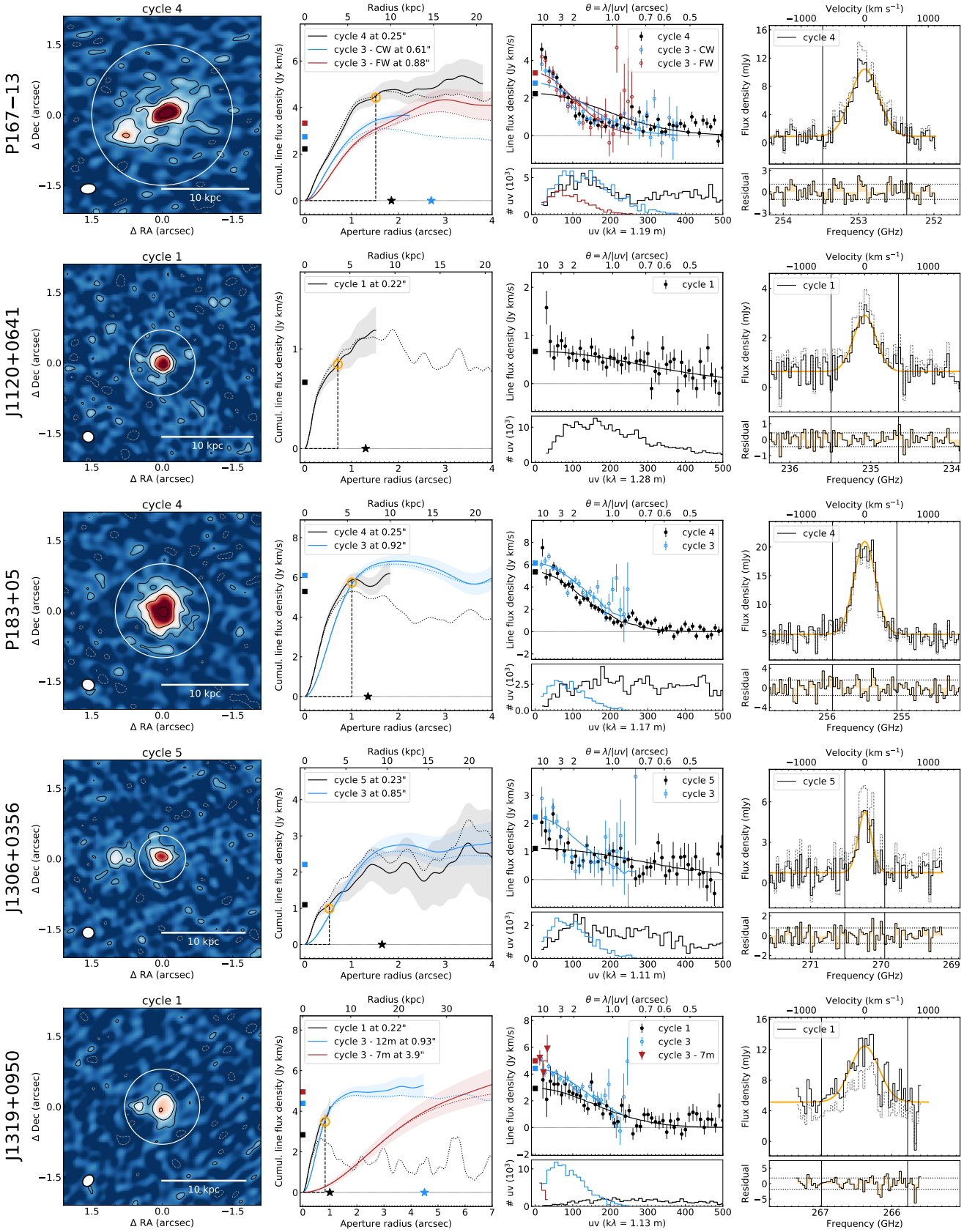

Figure 20. continued.

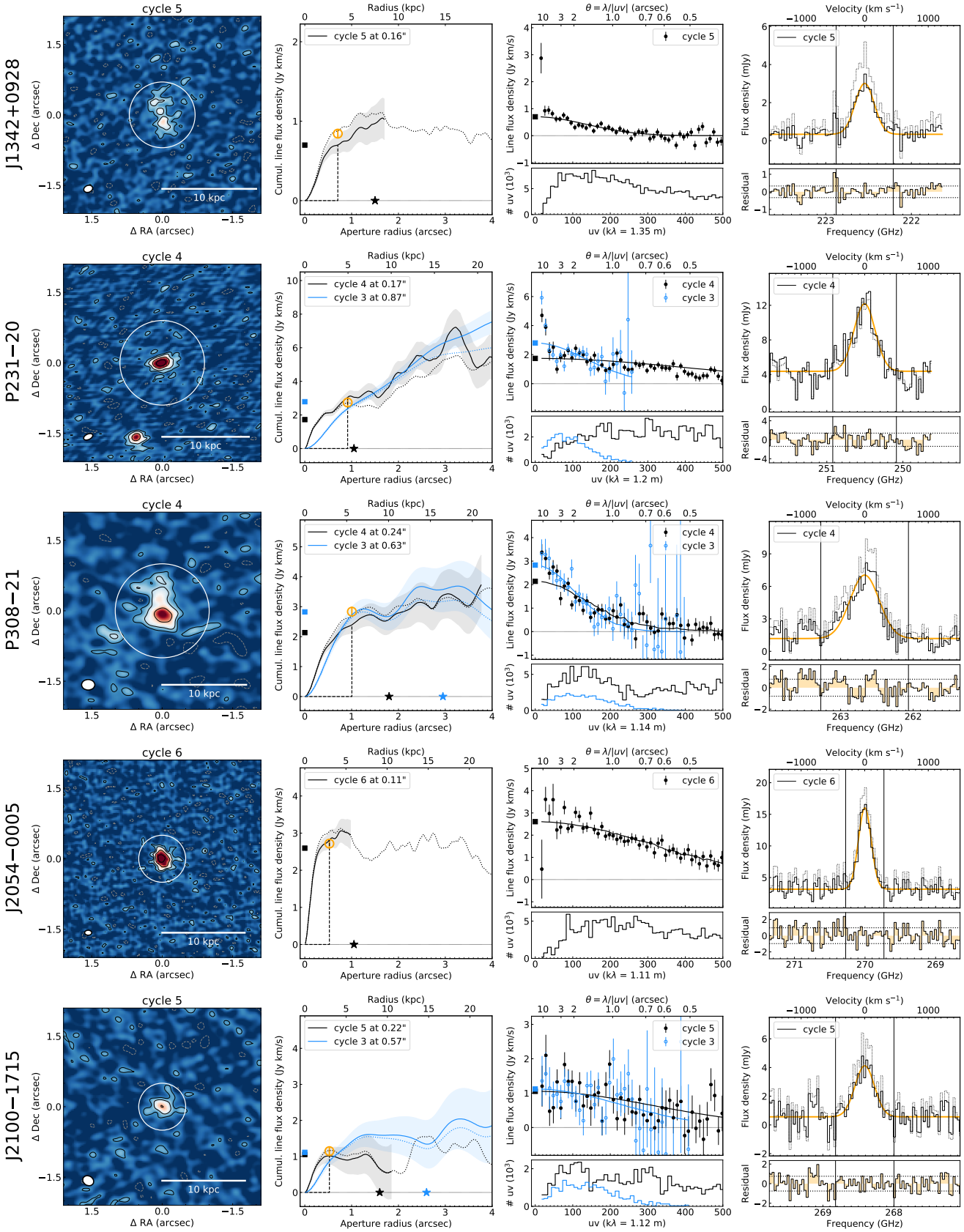

Figure 20. Continued.

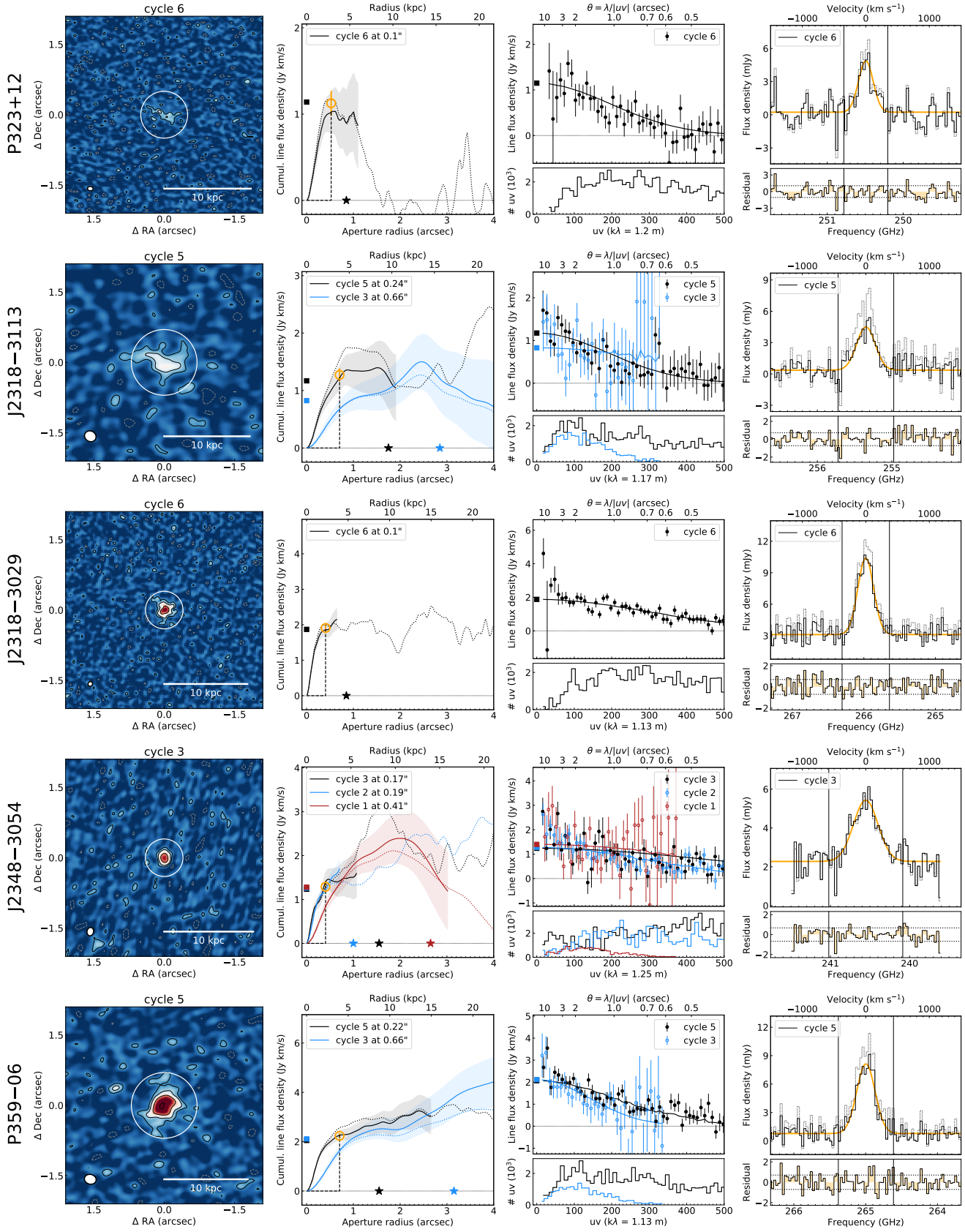

Figure 20. Continued.