Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **526,** 2278–2300 (2023) https://doi.org/10.1093/mnras/stad2824
Advance Access publication 2023 September 19

Tracing Evolution in Massive Protostellar Objects – I. Fragmentation and emission properties of massive star-forming clumps in a luminosity-limited ALMA sample

A. Avison [1,2,3★] G. A. Fuller [1,2,4,5] N. Asabre Frimpong [1,6] S. Etoka [1] M. Hoare [7] B. M. Jones [1,5] N. Peretto [8] A. Traficante [9] F. van der Tak,[10,11] J. E. Pineda,[12] M. Beltrán,[13] F. Wyrowski,[14] M. Thompson [15] S. Lumsden,[7] Z. Nagy [16,17] T. Hill,[18] S. Viti,[19,20] F. Fontani [13] and P. Schilke[5]

Affiliations are listed at the end of the paper

Accepted 2023 August 31. Received 2023 August 21; in original form 2023 January 13

ABSTRACT

The role of massive ($\geq 8\,M_\odot$) stars in defining the energy budget and chemical enrichment of the interstellar medium in their host galaxy is significant. In this first paper from the *Tracing Evolution in Massive Protostellar Objects* (TEMPO) project we introduce a colour-luminosity selected ($L_* \sim 3 \times 10^3$ to $1 \times 10^5\ L_\odot$) sample of 38 massive star-forming regions observed with ALMA at 1.3 mm and explore the fragmentation, clustering, and flux density properties of the sample. The TEMPO sample fields are each found to contain multiple fragments (between 2 and 15 per field). The flux density budget is split evenly (53 per cent–47 per cent) between fields where emission is dominated by a single high flux density fragment and those in which the combined flux density of fainter objects dominates. The fragmentation scales observed in most fields are not comparable with the thermal Jeans length, λ_J, being larger in the majority of cases, suggestive of some non-thermal mechanism. A tentative evolutionary trend is seen between luminosity of the clump and the 'spectral line richness' of the TEMPO fields; with 6.7 GHz maser associated fields found to be lower luminosity and more line rich. This work also describes a method of line-free continuum channel selection within ALMA data and a generalized approach used to distinguishing sources which are potentially star-forming from those which are not, utilizing interferometric visibility properties.

Key words: techniques: interferometric – stars: formation – stars: protostars – ISM: clouds – submillimetre: ISM – submillimetre: stars.

1 INTRODUCTION

Despite the importance of high-mass stars ($M > 8\ M_\odot$) on the galactic scale, due to their prodigious chemical and energetic feedback, our understanding of their formation and early evolution remains poorly understood (e.g. Tan et al. 2014). Answering the unresolved issues of massive star formation is not only important for the study of our Galactic environment but also has implications for the modelling of star formation and the evolution of the interstellar medium in extra-galactic sources throughout the star forming life-time of the Universe (Kennicutt & Evans 2012).

Current discussion within the literature centres around two scenarios under which protostars may acquire the necessary mass to form high-mass stars; these are commonly termed the *clump-fed* and *core-fed* scenarios (following e.g. Wang et al. 2010). The *core-fed* scenario posits that a stars final mass is correlated with the mass in the core from which is formed (McKee & Tan 2003; Tan et al. 2014) thus requiring the presence of both low and high mass protostellar cores to create the distribution of stellar masses seen on the main sequence, with some core to final mass efficiency relating initial core mass to final stellar mass. However, currently there is little evidence for cores of sufficient mass to create the most massive stars of 10 $M_\odot$ and greater (e.g. Nony et al. 2018; Sanhueza et al. 2019). Conversely, under the *clump-fed* scenario the final mass of a star is not determined purely by material available within its natal core, but instead on its position within and the material available to it from the larger scales of the host clump. Such multiscale hierarchical collapse removes the need for any relation between the initial mass of a protostellar core and the final mass of the star it forms, as the final mass is instead determined by the dynamical properties of the material on much larger scales and interaction/competition with other protostars in the protoclusters (Bonnell & Bate 2006; Wang et al. 2010; Peretto et al. 2013; Williams et al. 2018; Vázquez-Semadeni et al. 2019).

An important observational indicator which can allow the discrimination between proposed evolutionary scenarios are the fragmentation of star-forming clumps at early times within their evolution and the distribution (both spatially and in terms of the mass) of fragments within them.[1] Specifically, thermal Jeans fragmentation

★ E-mail: adam.avison@skao.int

[1]Throughout this paper we combine the nomenclature seen commonly within the literature (e.g. Zhang et al. 2009; Traficante et al. 2023) when referring to

is considered to be consistent with global hierarchical collapse and competitive accretion models (Sanhueza et al. 2019; cf. *clump-fed* models) whereas the need for turbulence or other mechanisms to support massive protostellar cores under the *core-fed* scenario may indicate the presence of fragmentation on non-thermal scales.

There is some evidence for fragmentation on the thermal Jeans length ($\lambda_J \sim 0.1$ pc at $T = 25$ K and $n = 10^5$ cm^{-3}) as opposed to turbulent or filamentary fragmentation scales in samples of infrared (IR) dark (at 70 μm) star-forming clumps, when observed at high sensitivity and angular resolution (Pillai et al. 2011; Sanhueza et al. 2019; Svoboda et al. 2019). Conversely, a number of authors have found evidence for filamentary, turbulently, or magnetically supported fragmentation scales (Wang et al. 2014; Beuther et al. 2015; Fontani et al. 2016; Henshaw et al. 2016; Lu et al. 2018; Sokolov et al. 2018; Traficante et al. 2023) when studying high mass star forming IR dark clouds (IRDCs) with Traficante et al. (2023) finding evidence for an evolutionary relation of Jean's length as a function of *L*/*M*.

The discrepancies between observational results may be attributable to a combination of factors such as differing sensitivities within observations or evolutionary differences in the samples of sources observed. The latter issue will be resolved over time as larger samples with varying sample selection criteria are published. It may also be the case that there is no 'one true' model for high-mass star formation and that attributes of different models are represented in different regions and at different times in their evolution depending on the environment and starting conditions.

This paper represents the first in a series from the *Tracing Evolution in Massive Protostellar Objects* (TEMPO) project. TEMPO has undertaken a systematic high resolution and high sensitivity survey using the world leading capabilities of ALMA to simultaneously study the chemistry, structure, and fragmentation of a luminosity and colour selected sample of young high mass embedded objects.

The two initial key goals of TEMPO are:

(i) Investigating how the mass and fragmentation of material in high-mass star-forming regions changes with luminosity and temperature.

(ii) Investigating how the observed molecular gas chemical composition evolves (e.g. number of complex organic molecules present, high gas density tracer abundance) as a function of luminosity and spectral energy distribution (SED) properties. Asabre Frimpong et al. (in preparation) will provide the first detailed analysis of the molecular emission recovered from the TEMPO data.

The current paper begins to address the first goal and presents the population, clustering, flux density budget, and fragmentation properties of our high-mass protostellar cluster sample as well as introducing and characterizing the observations of the TEMPO project. Section 2 introduces the sample, the ALMA observations undertaken and data processing. Section 3 provides an overview of the observation results for continuum emission and the characteristics of this emission. In Section 4 the clustering, fragmentation, and flux density budgets of the sample of observed fragments are discussed. Section 4 also comments on two properties of the TEMPO sample which relate to evolutionary characteristics. Using visibility analysis Section 4 also addresses whether the detected fragments are

structures of differing size is used. As such, objects of several to hundreds of pc are referred to as clouds, objects of ∼1 pc as clumps and objects ≲0.1 pc as fragments, unless the are known to be star forming in which case they are termed cores.

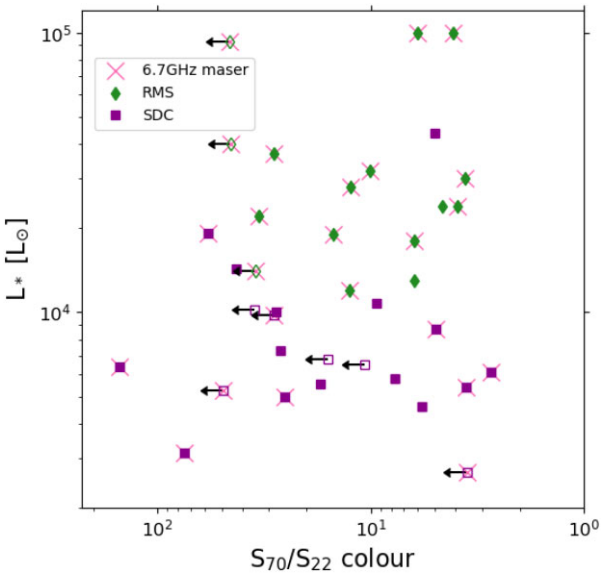

Figure 1. $S_{70\,\mu m}/S_{22\,\mu m}$ colour–luminosity plot for the fields in the sample, with RMS fields as green diamonds and SDC fields as purple squares. Luminosity derived from SED fitting to *Herschel* data by Mottram et al. (2011) for the RMS sources and Traficante et al. (2015) for the SDC sources. $S_{70\,\mu m}/S_{22\,\mu m}$ values are derived from Herschel (70 μm) and WISE (22 μm) measurements as presented in Lumsden et al. (2013; for RMS sources) and Traficante et al. (2015). Unfilled markers denote those fields which are not detected at 22 μm, listed with †s in Table 1 and represent 22 μm upper limits. Fields with a pink '×' have an associated 6.7 GHz methanol maser detection from the Methanol MultiBeam survey (Green et al. 2009).

likely currently star-forming or simply transient conglomerations of material, and association with other star forming tracers. Section 5 discusses our initial TEMPO findings and Section 6 provides a summarized conclusion.

2 OBSERVATIONS

2.1 The sample

The TEMPO sample comprises 38 luminosity and IR colour selected fields known to host young high-mass embedded protostellar sources, selected from both the Red MSX Source (RMS) survey (Lumsden et al. 2013) and the Spitzer Dark Cloud (SDC) sample (Peretto & Fuller 2009) to cover a range of $S_{70\,\mu m}/S_{22\,\mu m}$ colours and exhibit luminosities above 3×10^3 $L_\odot$, as seen in Fig. 1, a value which allows the sample to focus only on the most massive regions, i.e. those harbouring OB-type high-mass (proto)stars. The 70 μm data were taken from Herschel as part of the Hi-GAL survey (Molinari et al. 2010). The selection criteria were used to ensure the presence of high-mass protostars (high $L_\odot$ values) and cover a range of evolutionary stages from mid-IR 22 μm non-detections to $S_{70\,\mu m}/S_{22\,\mu m} \sim 1$.

The choice of colour [22–70 μm] was made as the similar [24–70 μm] colour has been found to provide a good discrimination between sources with SEDs which are well fitted by embedded Zero Age Main Sequence (ZAMS) star models (and are thus relatively more evolved objects) and those which are best fitted by a single optically thin greybody peaking at longer wavelengths than the ZAMS models (less evolved, relatively; Molinari et al. 2008), and bears a strong relation with source bolometric luminosity (Molinari et al. 2019). Similarly, Hughes & MacLeod (1989) used the [60–25 μm] colour to define the colour space occupied by highly evolved infra-red sources which

display H II regions at optical wavelengths. The WISE 22-μm data are used here rather than the *Spitzer* MIPS 24-μm data as the latter is saturated toward a number of the TEMPO fields.

The RMS Survey (Lumsden et al. 2013) was constructed using a subset of the v2.3 *MSX* point source catalogue (Egan, Price & Kraemer 2003), to generate a mid- and near-IR colour selected sample of massive protostellar objects. The colour-selection criteria was complemented by additional higher resolution infra-red and radio observations to remove ultra compact H II regions (UCH II) and planetary nebulae, which exhibit similar colours, from the sample. As such the RMS is 90 per cent complete for massive protostellar objects within the survey's observed area $10^\circ < l < 350^\circ$, $b < 5^\circ$.

The SDC sources from Peretto & Fuller (2009) are drawn from an initial sample of >11 000 IRDCs seen in absorption at 8 μm ($\tau_{8\,\mu m} > 0.35$) in the GLIMPSE (Churchwell et al. 2009) data from the *Spitzer Space Telescope*. Such 8 μm opacities mean all the SDC IRDCs have column densities above 10^{22} cm^{-2}. The selected SDC sources as targets for the TEMPO sample are from the 'starless and protostellar clumps embedded in the IRDCs' catalogue of Traficante et al. (2015) and we use the mass and luminosity properties for the selected sources from this work. All SDC sources selected for this current work have core masses > 500 $M_\odot$.

Additionally, the TEMPO fields (both RMS and SDC) were chosen to be isolated across a range of IR wavelengths to avoid confusion and to have distances less than ∼6 kpc. The range of distances to our target fields covers 1.8–6.3 kpc (a factor of 3.5)[2] which limits the lower range of observable spatial scales common within the data. There are 28 fields in the TEMPO sample (74 per cent) in which a 6.7 GHz class-II methanol maser detected within the Methanol MultiBeam survey (MMB; Green et al. 2009) is located with the observed ALMA primary beam. The 6.7 GHz class-II methanol maser is known to be uniquely associated with high-mass protostellar objects (Minier et al. 2003; Xu et al. 2008; Breen et al. 2013). The selection criteria properties for each field in the TEMPO sample are given in Table 1.

Throughout this work fields drawn from the RMS survey are prefixed with 'RMS-' (normally named simply after their Galactic coordinates i.e. Glll.lll±bb.bbb) to differentiate them from the sources from the SDC sample (preceded with 'SDC').

2.2 ALMA observations

The observations were conducted in ALMA Band 6 during Cycle 3 under project code 2015.1.01312.S. The project consisted of six separate scheduling blocks each requiring a single execution to meet the requested sensitivity. The observations were made on the dates 2016 March 7, 12, and 21. The telescope was setup to observe 4 × 1.875 GHz spectral windows (SPWs) with central frequencies of 225.2, 227.1, 239.8, and 241.9 GHz (equivalent to wavelengths of 1.33, 1.32, 1.25, and 1.24 mm, respectively). Each SPW consisted of 1920 channels giving a frequency resolution of 976.562 kHz, equivalent to a velocity resolution of ∼1.25 km s^{-1}. During each observation the array was configured with minimum and maximum baseline lengths of 15.1 and 460.0 m, respectively. These values give an average resolution of ∼0.7–0.8 arcsec and maximum recoverable scale[3] (MRS) within the data of 10.5 arcsec. At the average distance to the TEMPO fields, the average angular resolution gives a physical scale of 0.01 pc and the MRS is 0.2 pc. Table 2 gives the observing properties of the data set. The data used within this work was extracted from the ALMA Archive and calibrated using scripts provided in the CASA (McMullin et al. 2007) data reduction software (versions 4.7 for calibration and 5.4 for analysis).

2.3 Continuum determination and imaging

2.3.1 *Line emission*

The TEMPO target fields are young high-mass embedded protostellar objects meaning that all fields show some level of molecular line emission within the observations. Fig. 2 shows sample spectra from SPW 1 for a molecular linfe 'quiet' field and a line-dominated field.

To extract continuum emission information about the sample we must remove channels containing molecular line emission from the spectra. To do this, a new CASA based task, LUMBERJACK,[4] was developed and used to process these data. LUMBERJACK was used to process each field in the following way.

(i) The user selects the required ALMA measurement set and the target field within the measurement set to process.

(ii) LUMBERJACK then generates an image cube of the whole target field at full spectral resolution in each SPW.

(iii) The position of peak emission within each cube is located. This position is a single voxel (i.e. a position with a RA, Dec., and velocity value. The spectrum along the velocity axis at this position (in RA and Dec.) is extracted.

(iv) The returned spectrum is analysed to locate spectral lines using two complementary methods.

To analyse the spectrum, first, a sigma clipping analysis is used. This analysis derives the median and standard deviation values within the spectrum. Next, all channels with values which are either greater than the median value *plus* the spectrum standard deviation multiplied by a clip factor, or less than the median value *minus* the spectrum standard deviation multiplied by a clip factor, are excluded in iterative steps. The iterative analysis stops when either (i) the signal-to-noise of the current spectrum is greater than in the previous iteration (here the signal-to-noise is defined as the maximum value in the current spectra divided by the spectrum median value) or (ii) the percentage change in the standard deviation of the spectrum between iterations is greater than a user defined tolerance. For the TEMPO sample the clip level was set to twice the standard deviation and the tolerance set to a percentage of 95.5 per cent.

Secondly, a gradient analysis is used to calculate the channel to channel gradient, G. G is calculated as

$$G = \frac{S_{ch} - S_{ch-1}}{\Delta ch}, \quad (1)$$

where ch represents a channel number, S, the flux density in that channel, and Δch the channel width in units of channel (which here has a value of 1). Channels with $G > 3\sigma$, where σ is the theoretical *rms*-noise per channel of the data calculated by the LUMBERJACK

[2]For reference, at these distance 1 arcsec corresponds to a physical distance of 0.009 to 0.03 pc, respectively.

[3]The MRS for an interferometer is the scale at which an interferometer can reliably recover all emission from a coherent object. The MRS does not relate to the scale over which an interferometer can recover any emission. Objects observed with an interferometer above this size scale are likely to have missing flux, and any associated images suffering from imaging artefacts, e.g. negative bowling, due to this. All recovered fragments in the TEMPO sample are below the MRS and no imaging artefacts are seen in the TEMPO image data.

[4]See https://github.com/adam-avison/LumberJack for more information.

Table 1. Selection criteria properties of fields in the sample.

Field	Pointing RA [h:m:s]	Pointing Dec [°:′:″]	L_* [10^4 $L_\odot$]	$S_{70\,\mu m}/S_{22\,\mu m}$	D [kpc]	CH_3OH Maser? [Y/N]	*rms* [mJy]	percentage BW	N -	R_{cl} [pc]	FOV [pc]	X_{mean} [pc]	λ_J [pc]	M_{clump} [$M_\odot$]	R_{clump} [pc]	T_{clump} [K]
RMS-G013.6562-00.5997	18:17:24.40	−17:22:15.000	1.4	19.33	4.1	Y	0.15	34.2	6.0	0.2	0.44	0.13	0.05	7250.1	0.38	26.3
RMS-G017.6380 + 00.1566	18:22:26.40	−13:30:12.000	10.0	4.09	2.2	Y	0.63	17.7	9.0	0.08	0.24	0.05	0.02	143.6	0.05	40.0
SDC18.816-0.447_1	18:26:59.00	−12:44:45.000	0.5	5.72	4.29	N	0.19	23.3	2.0	0.06	0.46	0.2	0.04	373.5	0.14	19.6
SDC20.775-0.076_1	18:29:16.30	−10:52:09.000	0.6	†	3.95	N	0.23	18.8	14.0	0.22	0.42	0.08	0.03	880.6	0.15	20.3
SDC20.775-0.076_3	18:29:12.20	−10:50:35.000	0.6	7.71	3.95	N	0.11	53.6	4.0	0.06	0.42	0.09	0.08	2006.1	0.35	28.1
SDC22.985-0.412_1	18:34:40.10	−09:00:39.000	0.3	74.97	4.59	Y	0.33	14.8	7.0	0.23	0.49	0.16	0.02	622.2	0.08	36.3
SDC23.21-0.371_1	18:34:55.20	−08:49:15.000	1.0	†	3.84	Y	0.26	34.2	9.0	0.14	0.41	0.11	0.04	11135.0	0.46	22.1
RMS-G023.3891 + 00.1851	18:33:14.30	−08:23:57.000	2.4	3.93	4.5	Y	0.14	23.9	8.0	0.13	0.48	0.1	0.04	2138.3	0.24	23.5
SDC24.381-0.21_3	18:36:40.60	−07:39:14.000	0.6	17.19	3.61	N	0.16	21.2	11.0	0.27	0.39	0.1	0.02	1606.3	0.16	18.0
SDC24.462 + 0.219_2	18:35:11.60	−07:26:23.000	0.7	26.53	6.27	N	0.12	41.7	5.0	0.11	0.67	0.13	0.02	1098.0	0.12	21.3
SDC25.426-0.175_6	18:37:30.20	−06:41:16.000	1.0	†	3.98	N	0.14	40.2	3.0	0.05	0.43	0.13	0.07	663.8	0.2	35.5
SDC28.147-0.006_1	18:42:42.50	−04:15:34.000	0.5	25.23	4.49	Y	0.14	7.8	6.0	0.17	0.48	0.11	0.03	1078.4	0.16	19.9
SDC28.277-0.352_1	18:44:21.90	−04:17:39.000	0.5	3.55	3.12	Y	0.12	33.1	4.0	0.26	0.33	0.26	0.22	280.9	0.44	15.6
SDC29.844-0.009_4	18:46:13.00	−02:39:01.000	0.3	†	5.38	Y	0.69	7.4	8.0	0.18	0.58	0.08	0.03	7092.3	0.36	15.2
RMS-G029.8620-00.0444	18:45:59.60	−02:45:07.000	2.8	12.4	4.9	Y	0.23	27.3	6.0	0.1	0.52	0.1	0.04	1055.5	0.18	22.6
SDC30.172-0.157_2	18:47:08.20	−02:29:58.000	0.7	†	4.16	N	0.24	4.6	2.0	0.04	0.45	0.11	0.33	79.8	0.28	40.0
RMS-G030.1981-00.1691	18:47:03.10	−02:30:36.000	3.0	3.62	4.9	Y	0.15	6.1	3.0	0.13	0.52	0.2	0.06	372.8	0.18	25.1
SDC33.107-0.065_2	18:52:08.20	+ 00:08:13.000	1.9	58.3	4.54	Y	0.21	50.2	13.0	0.19	0.49	0.08	0.03	4405.5	0.26	26.4
RMS-G034.7569 + 00.0247	18:54:40.70	+ 01:38:07.000	1.2	12.56	4.6	Y	0.14	19.7	6.0	0.19	0.49	0.12	0.04	348.4	0.12	22.6
RMS-G034.8211 + 00.3519	18:53:37.90	+ 01:50:31.000	2.4	4.58	3.5	N	0.15	46.4	9.0	0.23	0.37	0.11	0.04	616.6	0.16	22.3
SDC35.063-0.726_1	18:58:06.00	+ 01:37:07.000	0.5	†	2.32	Y	0.29	39.3	9.0	0.13	0.25	0.06	0.01	379.8	0.06	25.5
SDC37.846-0.392_1	19:01:53.50	+ 04:12:51.000	4.4	5.03	4.08	N	0.66	36.0	9.0	0.18	0.44	0.09	0.06	4386.0	0.38	29.1
SDC42.401-0.309_2	19:09:49.90	+ 08:19:47.000	0.6	2.72	4.48	Y	0.09	63.0	4.0	0.1	0.48	0.14	0.06	210.2	0.12	40.0
SDC43.186-0.549_2	19:12:08.90	+ 08:52:08.000	1.4	42.68	4.15	N	0.24	54.1	12.0	0.26	0.44	0.11	0.05	3911.3	0.32	25.3
SDC43.311-0.21_1	19:11:17.00	+ 09:07:30.000	1.1	9.39	4.25	N	0.23	51.4	9.0	0.25	0.45	0.12	0.03	611.5	0.12	25.6
SDC43.877-0.755_1	19:14:26.20	+ 09:22:35.000	0.9	4.92	3.22	Y	0.25	70.2	11.0	0.17	0.34	0.08	0.04	6239.8	0.34	19.6
SDC45.787-0.335_1	19:16:31.20	+ 11:16:12.000	0.6	151.39	4.54	Y	0.36	2.1	5.0	0.14	0.49	0.14	0.02	237.6	0.08	25.9
SDC45.927-0.375_2	19:16:56.20	+ 11:21:54.000	1.0	27.69	4.21	N	0.11	39.7	5.0	0.16	0.45	0.13	0.06	885.0	0.24	23.4
RMS-G050.2213-00.6063	19:25:57.80	+ 15:03:00.000	1.3	6.22	3.3	N	0.15	24.1	11.0	0.13	0.35	0.06	0.03	221.9	0.09	22.2
RMS-G326.6618 + 00.5207	15:45:02.80	−54:09:03.000	1.4	†	1.8	Y	0.24	30.6	8.0	0.09	0.19	0.07	0.02	334.4	0.09	23.7
RMS-G327.1192 + 00.5103	15:47:32.80	−53:52:39.000	3.7	28.55	4.9	Y	0.24	24.2	7.0	0.48	0.52	0.27	0.03	443.7	0.1	36.8
RMS-G332.0939-00.4206	16:16:16.50	−51:18:25.000	9.3	†	3.6	Y	0.27	43.3	10.0	0.16	0.39	0.1	0.02	812.8	0.09	27.2
RMS-G332.9636-00.6800	16:21:22.90	−50:52:59.000	2.2	33.34	3.2	Y	0.44	6.6	10.0	0.26	0.34	0.12	0.02	1723.6	0.16	23.2
RMS-G332.9868-00.4871	16:20:37.80	−50:43:50.000	1.8	6.27	3.6	Y	0.21	16.2	4.0	0.1	0.39	0.1	0.04	602.1	0.16	22.4
RMS-G333.0682-00.4461	16:20:49.00	−50:38:40.000	4.0	†	3.6	Y	0.4	21.7	15.0	0.29	0.39	0.1	0.01	2291.5	0.1	23.7
RMS-G338.9196 + 00.5495	16:40:34.00	−45:42:08.000	3.2	10.09	4.2	Y	0.44	35.7	6.0	0.23	0.45	0.15	0.12	109.5	0.16	39.1
RMS-G339.6221-00.1209	16:46:06.00	−45:36:44.000	1.9	15.07	2.8	Y	0.25	29.4	9.0	0.22	0.3	0.12	0.03	321.0	0.1	24.1
RMS-G345.5043 + 00.3480	17:04:22.90	−40:44:24.000	10.0	6.05	2.0	Y	0.58	35.9	8.0	0.11	0.21	0.05	0.01	404.7	0.05	32.3

Fields which have † in the $S_{70\,\mu m}/S_{22\,\mu m}$ column indicate a non-detection at 22 μm. Distances and luminosities are taken from Lumsden et al. (2013; and online material from the RMS survey) for RMS sources and Traficante et al. (2015) for SDC sources. A 6.7-GHz CH_3OH maser is deemed associated with the field if it lays within one ALMA primary beam (at our observing frequency) of the field pointing position, maser postions were taken from the MMB survey catalogues (Caswell et al. 2010, 2011; Green et al. 2012; Breen et al. 2015). The value 'percentage BW' indicates the percentage of the total observed bandwidth (7.5 GHz) which was used in imaging once spectral line emission was removed (see Section 2.3). N, gives the number of sources detected in each field. FOV is the field of view in parsec, at the field distance. X_{mean} is the mean edge length of the minimum spanning tree of sources in each field and λ_J is the thermal Jeans length at T_{clump}. The values M_{clump} and R_{clump} are the clump mass, radii and, temperature as derived by Elia et al. (2021) from *Herschel* data.

Table 2. Observing properties of the ALMA data.

SPW	Central Freq. [GHz]	Freq. range [GHz]	Channel width [km s^{-1}]	Synthesized beam[a] [arcsec × arcsec]	PA[a] [°]	MRS[b] [arcsec]
0	239.8	238.86–240.74	1.22	0.77 × 0.64	58	10.2
1	241.9	240.96–242.84	1.21	0.77 × 0.64	46	10.2
2	227.1	226.16–228.04	1.29	0.81 × 0.67	55	10.8
3	225.2	224.26–226.14	1.30	0.82 × 0.68	56	10.9

[a]Average value of the synthesized beam across all fields. [b]MRS in data, defined as MRS $= \frac{0.6\lambda}{b_{min}}$ where $b_{\rm min}$ is the minimum baseline in the array.

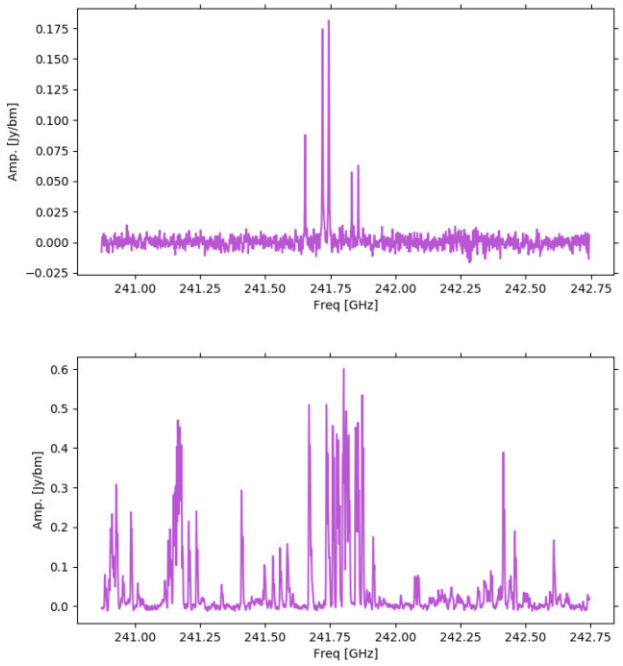

Figure 2. Examples of whole SPW (1.875 GHz bandwidth) spectra from the ALMA data. These examples present SPW 1 from a line 'quiet' field (*top:* SDC20.775−0.076_1, $L_* = 6.5 \times 10^3$ L$_\odot$) and line dominated source (*bottom:* SDC35.063−0.726_1, $L_* = 5.2 \times 10^3$ L$_\odot$). The spectra are taken at the position of the strongest continuum detection in the respective fields. The two brightest lines in the top panel are CH_3OH (5_K-4_K) transitions.

algorithm from the measurement set metadata[5] are rejected as line contaminated. The combination of the line contaminated channels found using the sigma-clip and gradient analysis are combined to give a conservative first pass at the line free channels in the data set.

(v) Following these steps a first pass 'line-free' continuum image is made for the combined (e.g. all SPW) data.

The user then defines continuum sources within this field for a second pass of line-free channel extraction. In the current work this was done using dendrogram analysis from the ASTRODENDRO PYTHON package[6] to find all the candidate continuum sources in the field. The parameters used during the continuum determination are the same as used during the final source extraction and discussed fully in Section 3.1.

(vi) Using the position of these candidate continuum sources, additional spectra are extracted (for the fitted source sizes) and then step (iv) repeated for all spectra, with the line-free channels from each source in each SPW concatenated to created a final list of line-free channels for the target field. The final channel list comprises only channels determined as line free for all sources in that field which ensures, as far as possible, no line contamination remains within the final images.

(vii) The final line-free channel lists for each SPW are created as the output product of the LUMBERJACK process.

There are three potential limitations of note with the LUMBERJACK analysis. First, typically the theoretical *rms*-noise used in the gradient analysis will be smaller than the measured *rms*-noise in an image as calibration errors are not accounted for when calculating the theoretical *rms*-noise. The implication of this is that some low intensity spectral lines may be overlooked in the gradient analysis; however, using a factor $>3\sigma$ should tend to counteract this, as should the cross comparison with the sigma-clipping analysis. Secondly, using the positions of continuum sources within the field may lead to spectral line emission from e.g. molecular outflows not being fully excluded as this type of emission would tend to be offset from the position of the continuum sources. The use of a first pass continuum image and a second round of spectral line analysis acts to mitigate this. Visual inspection of the spectra, cubes and continuum images suggests that the effect of this latter limitation is minimal. The third limitation would occur in very line-rich objects within which there was a lot of velocity components or velocity gradients from the molecular material. This would give broad and potentially overlapping spectral line profiles across the observed spectrum and exclude possibly all channels within the observed frequency range. This case does not occur within the TEMPO sample.

To inspect the reliability of the LUMBERJACK continuum extraction within the TEMPO sample, a sub-sample of eight ($\sim$ 20 per cent) of the TEMPO fields were selected. The fields chosen were amongst the line richest of the RMS and SDC targets (four of each) and have been compared to the ARI-L continuum images available in the ALMA Archive (Massardi et al. 2021). Considering all four SPWs this gives a sample of 32 data points of comparison. The TEMPO and ARI-L continuum image peak flux density pixel values were used for the comparison as this tended to be toward the line richest source in a given field. The primary beam corrected images were used from both ARI-L and TEMPO (prior to self-calibration for TEMPO to ensure a fairer comparison). From this comparison we find all data points

[5]The theoretical *rms*-noise is calculated by extracting the time on-source, Δt, the median system temperature, $T_{\rm sys}$, channel width in Hertz, $\Delta\nu$ and number of antennas, N used during observation from the measurement set metadata. These values are then combined as

$$\Delta S = \frac{2kT_{sys}}{A_{eff}\eta\sqrt{N(N-1)\Delta\nu\Delta t}}, \quad (2)$$

where k is the Boltzmann constant, A_{eff} the effective area of an ALMA antenna at the observing frequency and η the aperture efficiency parameter ($\sim$0.7; Remijan et al. 2020)

[6]http://www.dendrograms.org/

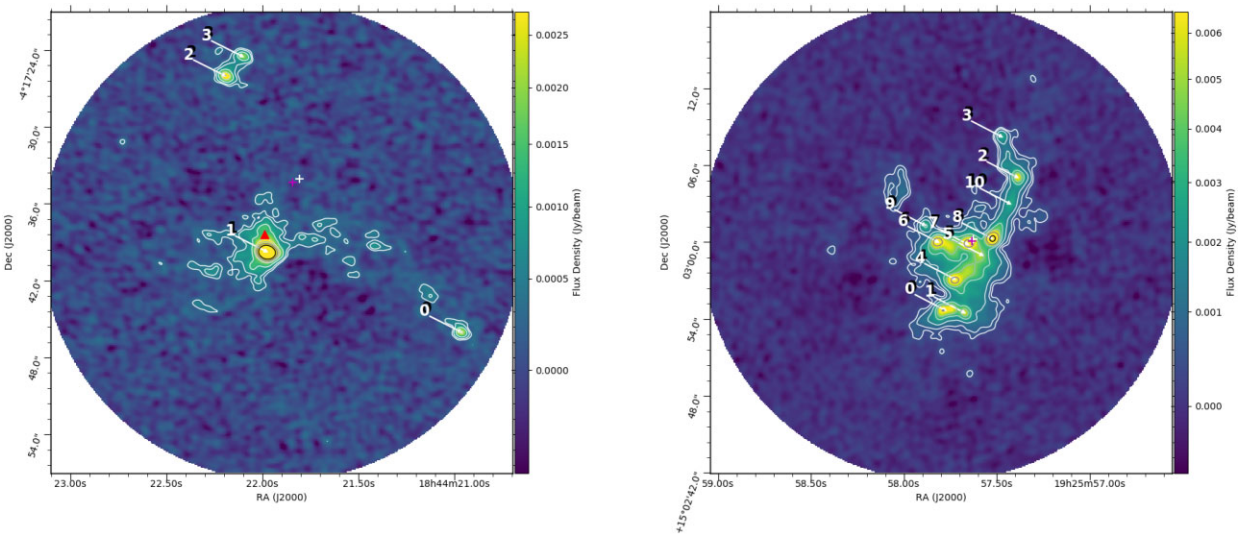

Figure 3. Example maps of the combined aggregate bandwidth continuum images for fields SDC28.277−0.352_1 (left) and RMS-G050.2213−00.6063 (b, right). Contours are at 3, 5, 10, 20, 30, 50, and 100× the fields *rms*-noise level, 0.19 and 0.46 mJy, respectively. The red triangle in (a) indicates the position of the 6.7GHz methanol maser in that field with position from Breen et al. (2015). The white and magenta '+' symbols give the average and normalized flux density-weighted average position of sources in the field. Numbers and arrows indicate the detected sources in each field. Maps of all target fields can be found in Appendix A (available online).

are within ±20 per cent of one another with the exception of three, showing a mean of 12 per cent difference with a standard deviation of 18 per cent (reducing to 8 per cent and 4 per cent when excluding the three outliers).

Given the absolute flux density calibration accuracy of ALMA being at the 10 per cent level in Band 6 (e.g. Remijan et al. 2020), the amount of line emission removed, differences in CASA version used in calibrating and imaging the data and differences in imaging parameters (e.g. cell size, 0.13 arcsec ARI-L and 0.093 arcsec TEMPO) we believe that this constitutes a good matching between the TEMPO/LumberJack line extraction and that implemented by the ARI-L project. For the three data points beyond this range, one shows 25 per cent discrepancy between ARI-L and TEMPO which is considered marginal. The remaining two are for sources RMS-G013.6562 − 00.5997 in SPW0 (239.8 GHz) at +42 per cent (ARI-L greater than TEMPO) and G326.6618+00.5207 in SPW1 (241.9 GHz) at + 82 per cent (again ARI-L greater than TEMPO). For these two objects the spectra are extremely line rich making continuum extract very difficult. We do note that in both cases, comparing the continuum values across all SPWs the TEMPO values are more consistent with a typical smoothly sloping spectral index than the ARI-L data.

The LUMBERJACK derived line-free channel lists were used to create continuum images of each field in each SPW and as a single aggregate bandwidth (i.e. combined line free channels across all SPWs) continuum image using all line-free channels. The data were imaged in CASA using the task `tclean`, using '*briggs*' weighting with the robust parameter set to 0.5. The `tclean` parameter *deconvolver* was set to *multiscale* as the data exhibit extended structure and this algorithm allows for the best quality images in such cases, *scales* of 0, 6, 18, 26, and 43 pixels were used. These values correspond to a delta function, one- and three-times the beam size in pixels and approximately, 0.25- and 0.4-times the MRSs of the data, respectively. The last two scales were found by manual inspection to produce the best images with the TEMPO data. The default *smallscalebias* value of 0.6 was used throughout.

2.4 Self-calibration and noise characteristics

To ensure the highest dynamic range continuum maps for the TEMPO sample, an initial set of continuum images for the TEMPO fields (both combined continuum from all SPWs and continuum from each individual SPWs) were inspected to check if the respective signal-to-noise ratio was sufficient to undertake self-calibration of the data. For sources where self-calibration was possible (35/38 sources),[7] up to three rounds of phase-only calibration were used to correct the phase solutions and produce the final maps used in our analysis. Amplitude self-calibration was not attempted as amplitude base calibration artifacts were not obvious within the data set. The single SPW images were made with *nterms* = 1 which assumes a flat spectrum due to fractional bandwidth considerations, whereas the combined SPW images used *nterms* = 2. The cleaning masks for each source were created using CASA's auto-masking capabilities. Images of all fields were created both with and without primary beam correction. Fig. 3 give example images of the generated maps, with the rest of the sample shown in Appendix A (available online).

Following the LUMBERJACK processing described in Section 2.3 the number of channels determined to be 'line-free' and thus the total aggregate bandwidth in each SPW and each field is different. This results in the final continuum maps having a non-uniform sensitivity from field to field. An additional factor in the sensitivity achieved in each field is the spatial distribution of extended emission and any associated 'missing' flux which is resolved out by the interferometer. Missing flux leads to artifacts such as negative 'bowling' in the maps and has a significant effect on the determination of the noise characteristics of the images.

Table 3 gives characteristic values for the data set as a whole, with the final two rows giving the equivalent mass sensitivities for the combined SPW images at $T = 15$ K and $T = 30$ K at the average distance to our target fields, D = 3.9 kpc. The sensitivity by field is

[7]The exceptions being SDC18.816−0.447_1, SDC30.172−0.157_2, and SDC45.927−0.375_2.

Table 3. Characteristics *rms*-noise and mass sensitivity properties by SPW across the sample.

SPW	*rms*-noise [mJy]			
	Mean	*Median*	*Max.*	*Min.*
0	0.47	0.33	1.80	0.17
1	0.56	0.37	3.65	0.17
2	0.50	0.30	3.17	0.16
3	0.44	0.30	2.28	0.15
All	0.26	0.23	0.69	0.09
	Mass sensitivity [$M_\odot$]			
	Mean	*Median*	*Max.*	*Min.*
$T = 15$K	2.5	2.2	6.5	0.9
$T = 30$K	1.0	0.9	2.7	0.4

'All' row gives the *rms*-noise properties for the combined SPW images. The mass sensitivities are calculated using the 'All' noise values at temperatures of $T = 15$ K and $T = 30$ K, the average distance to our target fields D = 3.9 kpc and using the dust opacities from Ossenkopf & Henning (1994) for protostellar cores. The model used was that including grains with ice mantels at a column density of 10^6 cm^{-3} (sixth column, including wavelength column), following e.g. van der Tak et al. (1999). Opacity value used was therefore $\kappa = 8.99 \times 10^{-1}$ g^{-1}cm^2.

listed in column 8 of Table 1 with column 9 giving the percentage of line free channels (across all four SPWs) found by the analysis described in the previous subsection as an indicator of the wealth of lines found in the sample.

3 RESULTS

The spectral line free ALMA continuum maps are given in Figs 3 and A1 in Appendix A. The observed and derived properties for each field as a whole can be found in Table 1, which gives the *rms*-noise value, percentage line free bandwidth, number of sources, and protocluster radius ($R_{\rm cl}$), the field of view of the ALMA primary beam in parsecs at the used target distance, the mean edge length ($X_{\rm mean}$) of a minimum spanning tree (MST) in each field and the thermal Jeans fragmentation length ($\lambda_{\rm J}$), respectively. The derivation of $R_{\rm cl}$ and $\lambda_{\rm J}$ are discussed in Section 3.2 and $X_{\rm mean}$ in Section 3.3.

The positions and properties of each detected source (hereafter refered to as a fragment) are given in Table 4. Column 1 lists the target field (as found in Table 1), column 2 the fragment ID in that field (from 0 to the *n*th), columns 3 and 4 the RA and Dec. of the source. Column 5 gives the measured continuum flux density in the map combining data from all SPWs. Column 6 gives an indication, the active star-formation candidates score (ASC $_{\rm score}$), of the likelihood the source is actively star forming (as discussed in Section 4.5), column 7 denotes which fragment is the brightest in the field, columns 8 and 9 indicate the most central fragment in the cluster for both an arithmetic and normalized flux density-weighted average cluster centre, respectively.

3.1 Source extraction

To generate the lists of fragments for each field a dendrogram analysis (Rosolowsky et al. 2008) was run on the final continuum maps for each SPW and on the combined SPW map using the ASTRODENDRO PYTHON package. The dendrogram analysis used the following parameters *min_value* = 5.0 × *rms*, *min_delta* = 1.0 × *rms* and a *min_pix* equivalent to the number of pixels within the synthesized beam area (approximately 21 pixels). These parameters were selected after experimentation with the TEMPO data to yield

Table 4. Sample of the Catalogue Table for the first 10 fragments detected in the TEMPO sample, the full table for all sources is available as online material and on the CDS via anonymous ftp to cdsarc.u-strasbg.fr (130.79.128.5) or via https://cdsarc.unistra.fr/viz-bin/cat/J/MNRAS..

Field	Source No.	RA [h:m:s]	Dec. [°:′:″]	$\rm Axis_{maj}$ [arcsec]	$\rm Axis_{min}$ [arcsec]	PA [Deg]	$\rm S_{combined}$ [mJy]	$\rm ASC_{score}$	Brightest?	Central [mean]	Central [weighted]?
RMS-G013.6562-00.5997	0	18:17:24.374	−17:22:14.720	0.703	0.484	−177.99	4.707	0	0	0	0
	1	18:17:24.028	−17:22:14.907	0.854	0.234	−166.59	12.442	0	0	0	0
	2	18:17:23.878	−17:22:14.346	0.662	0.26	−171.96	6.785	0	0	0	0
	3	18:17:24.243	−17:22:12.852	0.434	0.377	−176.11	272.43	3	1	1	1
	4	18:17:24.348	−17:22:11.824	0.252	0.166	148.05	9.851	0	0	0	0
	5	18:17:24.152	−17:22:01.924	0.592	0.391	174.27	17.836	0	0	0	0
RMS-G017.6380 + 00.1566	0	18:22:26.976	−13:30:18.258	0.909	0.455	171.87	36.54	0	0	0	0
	1	18:22:26.778	−13:30:17.978	1.241	0.375	−179.81	38.38	0	0	0	0
	2	18:22:26.848	−13:30:16.016	0.733	0.623	−172.76	66.572	0	0	0	0
	3	18:22:26.573	−13:30:16.016	1.45	0.439	178.98	31.153	0	0	1	1

Column 1 gives the field name as specified in Table 1, column 2 gives the source number of the TEMPO fragment in that field, columns 3, 4, 5, 6, and 7 give the RA, Dec., fitted Major, and Minor axes and source position angle for the detected sources from the Dendrogram analysis, respectively. Column 8 gives the measured source continuum flux density in mJy. Columns 9 gives the fragments ASC score as defined in Appendix C and columns 10, 11, and 12 denote if the source is the Brightest in the field, the most central using the arithmetic mean (column 11) or weighted mean (column 12), respectively, with 1 indicating True and 0 indicating False.

realistic results and are consistent with those used by other authors on comparable data sets (e.g Henshaw et al. 2016).

The resulting lists of fragments per image are cross matched in position, with fragments which have a matching peak position (within half the ALMA synthesized beam Full Width Half Maximum (FWHM) for a given field) in all individual images retained. As an independent additional check the *GaussClumps* algorithm within the STARLINK software package was run on the combined continuum image and the final dendrogram fragment list cross matched with the *GaussClumps* list. The fragments retained from this cross comparison are our final fragment list for each field. The properties of these fragments are then extracted from each image.

During the dendrogram and *GaussClumps* processing the non-primary beam corrected images were used, as primary beam correction increases the noise toward the edge of each map and leads to both algorithms including spurious noise features in their respective source lists. Using the final fragment lists the flux densities were extracted from the primary beam corrected maps.

From the sample's 38 fields a total of 287 individual fragments were detected above 5σ (in the non-primary beam corrected maps). This gives an average of 7.6 fragments per observed field, with values ranging from 2 to 15 fragments in individual fields. The distribution of fragments per field is given in Fig. 4.

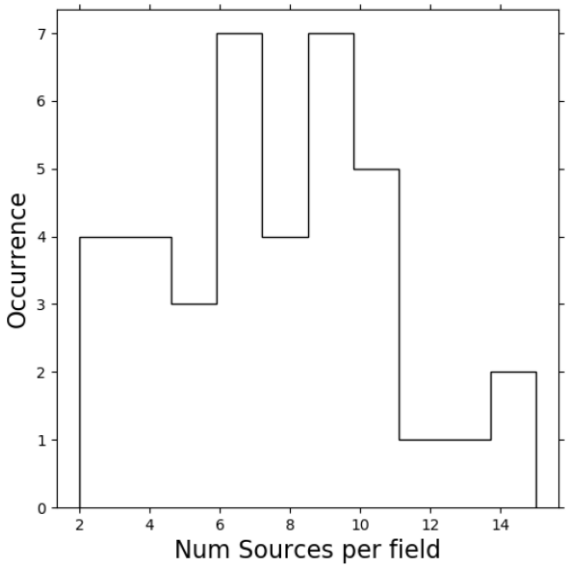

Figure 4. Distribution of fragment counts across the 38 TEMPO fields.

3.2 Protocluster radius and Jeans length

Using the extracted positions and flux densities of the fragments in each field, the protocluster radius and representative values of the Jeans length were derived.

The protocluster radius is defined here as the distance from the cluster centre to the furthest fragment position in that field, and makes the assumption that the whole cluster is observed within the ALMA primary beam of the TEMPO observations (∼23 arcsec). The cluster centre is defined in two ways, first as the average position of all fragments in each cluster and second as the average of the flux density-weighted fragment position (such that those with greater flux density are weighted more highly, this utilize the field normalized flux density, e.g. fragment flux density divided by the highest fragment flux density in the field). The distribution of cluster radii calculated using both methods can be seen in Fig. 5 and values for each field are given in column 11 of Table 1. Using either the arithmetic or weighted mean has little impact on the distribution of protocluster radii in this sample, both peaking between 0.1 and 0.2 pc, with a potential bimodality in the weighted case.

In the simplest case, i.e. with no magnetic or turbulent support against collapse, clump fragmentation is expected to occur on the scales of the Jeans length (λ_J). The λ_J values for the TEMPO fields were calculated following the approach used by the ASHES survey (Sanhueza et al. 2019):

$$\lambda_J = \sigma_{th} \sqrt{\frac{4\pi^2 R_{clump}^3}{3GM_{clump}}}, \quad (3)$$

where M_{clump} and R_{clump} are the clump masses and radius respectively (columns 13 and 14 in Table 1), for the TEMPO fields these values were taken from Elia et al. (2021). σ_{th} is the thermal velocity dispersion and is given by $\sigma_{th} = \sqrt{\frac{kT}{\mu m_H}}$, with k the Boltzmann constant, μ the molecular weight (here = 2.37) and m_H the mass of the Hydrogen atom. The temperatures, T, used here is the T_{clump} also from Elia et al. (2021); given in column 15 of Table 1. Fig. 6 provides a histogram of λ_J from the fields in the TEMPO sample, this value peaks at ∼0.025 pc, with a relatively narrow distribution throughout the sample excluding a few outliers at higher values.

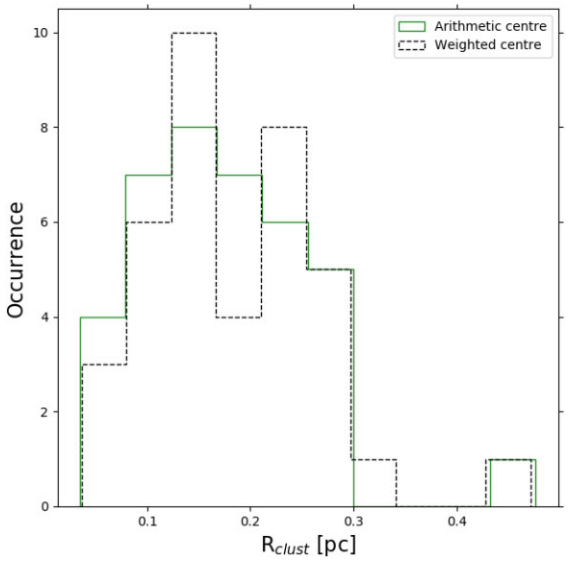

Figure 5. Distribution of measured cluster radii across the 38 fields in the TEMPO sample, as measured from the arithmetic mean position and weighted mean position.

3.3 Minimum Spanning Trees (MST)

Using the extracted fragment positions a set of MST were generated for each TEMPO field. The MSTs were created using the MINIMUM-SPANNING-TREE module within the PYTHON SCIPY module. MSTs provide a set of edges, which describe the minimized set of lines to connect points within a cluster of points. Within this analysis the MSTs are used to describe the mean edge length in the TEMPO clusters as part of the Fragmentation analysis 4.2 and in an investigation of the 'Q'-value metric used to described source distributions in Appendix B (available online). Example MSTs are given in Fig. 7.

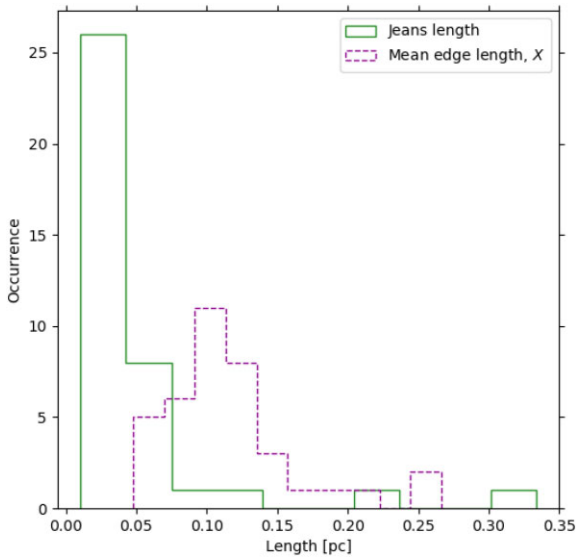

Figure 6. Distribution of calculated Jeans lengths, λ_J, for 38 fields in the TEMPO sample (green solid lined histogram) and measured mean edge length, X, from the MST analysis of the sample (purple dashed histogram).

From the MSTs the average mean edge length is 0.12pc (not accounting for projection effects). This value is similar to the fiducial core scale of 0.1 pc.The distribution of these values is shown in Fig. 5 as the purple dashed histogram. The implications of these measurements are discussed as part of the fragmentation analysis (Section 4.2).

4 ANALYSIS

The initial focus of the TEMPO analysis is on the structure/fragmentation and the distribution of flux density detected in each of the sample fields. At this stage (Sections 4.1–4.3) no attempt to categorize the detected fragments into star-forming cores and not star-forming fragments is made and, as such, all fragments are treated as potentially star forming. In Section 4.5 a potential interferometric classification into star forming core and non-star forming fragmented material is introduced.

4.1 Clustering properties

4.1.1 Nearest neighbours

Using the distance to each target field (given in Table 1) the projected physical separation between each fragment in a given field was calculated from the observed angular separation.[8] The number of neighbours per fragment within radial cut-offs of 0.03, 0.05, and 0.1 pc were inspected. These cut-offs were chosen to be representative as they are all within the fiducial protostellar core size scale (0.1 pc; e.g Zinnecker & Yorke 2007) and above the lowest angular separation detectable within our data. This lower limit on detectable angular separation arises from the angular resolution of our data, objects separated by less than this scale would be observed as a single object. Taking the major axis of the average synthesized beam (0.82 arcsec) this lower limit would be 0.015 pc ($\sim$ 3200 au) at the average field distance of 3.9 kpc and covers a range from 0.007 to 0.025 pc over the TEMPO sample's distance range of 1.8–6.3 kpc. Below this it is not possible to distinguish between objects with the current data.

Fig. 8 shows the distribution of the number of nearest neighbours within each cut-off interval, including those which do not have a neighbour within that interval in the 'Neighbours' equal to 0 bin. Fig. 8 shows that very few of the sources within our sample are solitary.

Over half of fragments (58.2 per cent) have a neighbour within 0.03 pc, increasing to 82.6 per cent of fragments with a neighbour within 0.05 pc and 96.9 per cent with a neighbour within our largest cut-off of 0.1 pc. Only nine sources (3.1 per cent of the total sample) do not have a neighbour within the 0.1 pc cut-off. Coupling this with the number of fragments detected per field, ranging from 2 to 15, would suggest that our detected fragments are densely distributed within the target fields (cf. the observing field of view which is $\sim$23 arcsec, equivalent to $\sim$0.4 pc at the average field distance of 3.9 kpc). Together these values would seem to suggest that in most cases we are seeing in each field the fragmentation of a single star forming core (under e.g. the core accretion scenario) assuming the fiducial 0.1 pc size scale.

4.1.2 Cluster radial profile properties

To examine the fragment density profiles of the protoclusters in the TEMPO fields, the positional offset for each fragment from their respective protocluster centre was calculated. Fig. 9 gives the number of fragments at increasing radial offsets from both calculated cluster centres. We use the distance to each field from Table 1 to give a physical offset and normalized by the cluster radius.

Fig. 9 shows as filled lines the equivalent distribution of field normalized offsets from 40 000 randomly created three-dimensional clusters. The randomly generated clusters have N sources/fragments (for N randomly selected between 3 and 13, to closely match the true field values without extremes cf. 2–15 is the true range) and radial profiles of $N(r) \propto r^{-\alpha}$, where $N(r)$ is the number of sources as a function of r given the exponent $\alpha = 0.0$, 1.0, 2.0, and 2.9 with 10 000 distributions per α value. To generate the cluster distributions the work of Cartwright & Whitworth (2004) was followed, using their formulae

$$r = \left(\frac{(3-\alpha)R}{3}\right)^{\frac{1}{3-\alpha}}, \tag{4}$$

$$\theta = \cos^{-1}((2\Theta) - 1), \tag{5}$$

$$\phi = 2\pi\Phi, \tag{6}$$

where for each cluster R, Θ and Φ are randomly selected values between 0 and 1. The resulting r, θ, and ϕ values are then converted to x, y, z positions and projected into two dimensions. The projected 2D positions are used to calculate the offset from the cluster central position. The width of the filled lines in Fig. 9 represent a ± 1 standard deviation at each histogram bin at a given normalized offset.

The observed data does not agree strongly with any of the plotted $r^{-\alpha}$ profiles, though visually both distributions appear closest to the $r^{-2.0}$ profile with exceptions of an excess between $\sim$0.2 and 0.5 for the normalized offset for both the averaged centre and normalized flux density-weighted centre histograms.

As a more quantitative measure the observed data distributions were compared to the generated $r^{-\alpha}$ profiles using a two sample Kolmogorov–Smirnov test. With this method the null hypothesis

[8]The effects of projection are accounted for when converting from observed angular separation to physical separation by dividing by a factor of $\frac{2}{\pi}$. This of course assumes the cluster is spherical in nature which may not be true in all cases.

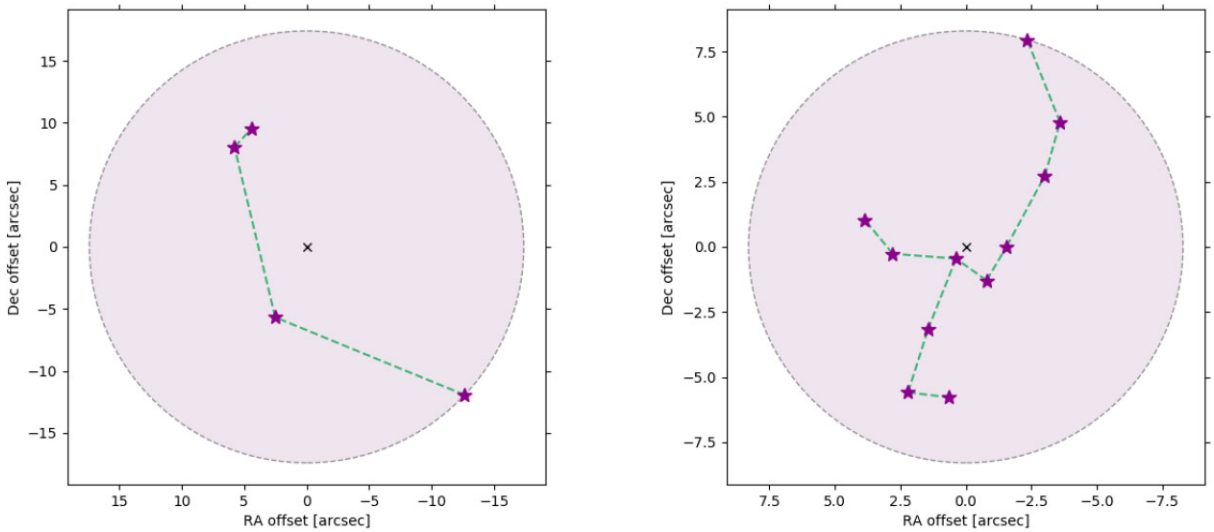

Figure 7. Example MST for fields SDC28.277−0.352_1 (left) and RMS-G050.2213−00.6063 (right), the same fields as shown in the maps in Fig. 3. Purple star icons denote source location, the green dashed lines are the edges of the MST. The purple shaded region is a circle of radius equal to the protocluster radius as defined in Section 3.2 centred at the average position of all sources in the field.

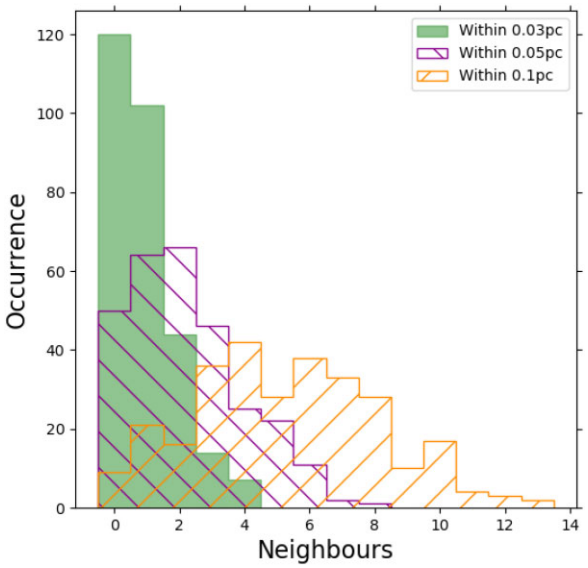

Figure 8. Number of nearest neighbours for each fragment in the sample at cutoffs of 0.03 pc (filled green histogram), 0.05 pc (purple '\' hatched histogram) and 0.1 pc (yellow '/' hatched histogram). Fragments in the Neighbours = 0 bin do not have a neighbour within that angular offset cut-off.

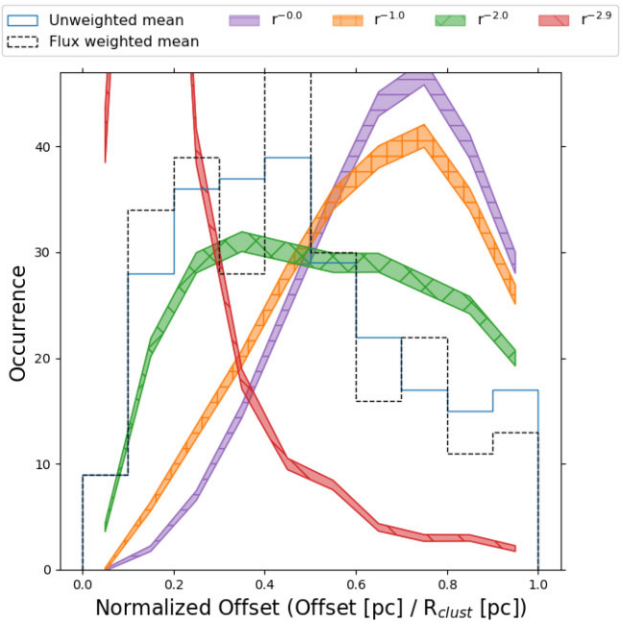

Figure 9. Combined distribution of source position as offset from the mean position of all fragments in their respective fields (blue histogram) and the normalized flux density-weighted mean position (black dashed histogram), normalized by the cluster radius $R_{\rm clust}$ or weighted cluster radius, for each field. The filled regions show the expected normalized radial profiles, $r^{-\alpha}$ for values of α = 0.0, 1.0, 2.0, and 2.9 (purple with horizontal hatching, orange with '+' hatching, green with '×' hatching, and red with '\' hatching, respectively). These profiles were drawn from 40 000 (10 000 per α value) randomly generated three-dimension clusters with between 3 and 13 sources within them. See text for further details.

that the observed data are drawn from the same distribution as the generated profiles is tested. Applying this test to the TEMPO data it is possible to reject the null hypothesis for TEMPO fields being drawn from an $r^{-2.9}$ profile with a p-value of 0.007 (0.031; with comparisons to the weighted average values in brackets) these values indicate that the null hypothesis is rejected with only a <0.7 per cent (<3.1 per cent) probability of rejecting a true null, typically a p-value of less that 0.05 is considered sufficient to reject the null hypothesis.

It is not possible to reject the null hypothesis for α values of 0, 1, or 2 with p-values of 0.68 (0.68), 0.97 (0.97), and 0.31 (0.11), respectively. This finding shows the TEMPO fields do not show a highly centrally condensed profile (α = 2.9) but beyond this it is not possible to not rule out that shallower radial profiles exist within our target fields. This may also suggest that different population distributions, e.g. fractal or broken power law, are present within the sample. The small source counts in the TEMPO sample limits the ability to conduct this analysis on a field by field basis.

The Q-parameter, introduced by Cartwright & Whitworth (2004), has proven within the literature to be a useful diagnostic of stellar distributions within clusters. However, in testing this parameter for fields in the TEMPO sample it was found that the fragment counts were too small for Q to be used robustly. A similar interpretation of the Q-parameter for small number clusters is seen in Parker (2018) in the case of L1622 for as many as 29 sources. Details of an investigation into the Q-value for small source/fragment counts conducted by the TEMPO team is presented in Appendix B.

4.2 Fragmentation scales

In addition to the the cluster profile characteristics, the scales upon which the material in each field is fragmenting was investigated by comparing the source separations to the Jeans fragmentation length.

Table 1 lists in column 13 the calculated values of λ_J for each observed field. The average λ_J value is 0.05 pc. These values are compared to the mean edge length, X, which gives the distance between sources along the MST (this is the same X as seen in equation B1 corrected for projection effects by division of a factor $\frac{2}{\pi}$ (Sanhueza et al. 2019). As can be see in Fig. 5, X peaks at ∼0.1 pc and covers a smaller range of values than the Jeans Lengths, but with typically higher values.

The ratio of λ_J to X-values throughout the sample range from 0.33× to 9.1×, with only one field (SDC30.172−0.157_2)[9] having λ_J/X less than 1. For the majority of TEMPO fields therefore the observed mean edge length between fragments is not consistent with thermal Jeans fragmentation and thus another non-thermal mechanism must be presented to account for the observed fragmentation.

Filamentary or cylindrical fragmentation as seen in the works of Ostriker (1964), Henshaw et al. (2016), and Lu et al. (2018) would tend to have length scales greater than those observed in the TEMPO fields. Using T_{clump} from Table 1 and equation (2) from Henshaw et al. (2016) the λ_{frag} for the TEMPO sample was calculated. As the hydrogen number density is unknown for the TEMPO sample values between 10^4 and $10^6 cm^{-3}$ were input. Comparing of the $\lambda_{frag,\,f}$ to X for each TEMPO source shows that $\lambda_{frag,\,f}$ is consistent with X for 19 TEMPO fields at a density value of 5.0×10^5 cm^{-3} and 32 TEMPO fields at value of 1.0×10^6 cm^{-3}, both densities appropriate for star-forming regions. Meaning that filamentary fragmentation could account for the fragmentation scales seen some of the TEMPO fields. However it is noted that, morphologically the TEMPO sample do not appear particularly filamentary.

It is noted that the works of Henshaw et al. (2016) and Lu et al. (2018) have observed filamentary fragmentation in mosaic images of larger regions of sky than the present work and were targeted towards known filamentary objects, whereas the TEMPO sample had no such selection criteria. It is expected that the TEMPO fields observe the whole of the local star-forming core because to the physical scale of the ALMA field of view at the distances to the TEMPO sample is being greater than the fiducial star forming core size. However, it is not possible to rule out additional sources beyond the field of view limits without additional data to create mosaics covering a region of the sky.

Additionally, turbulent fragmentation can cause a deviation away from the Jeans length, in either direction (Pineda et al. 2015) and could potentially also account for the fragmentation scales seen in the TEMPO sample in addition to some filamentary fragmentation.

[9]This field is one of the lowest SNR sources in the sample and contains only two fragments, which also may account for this result.

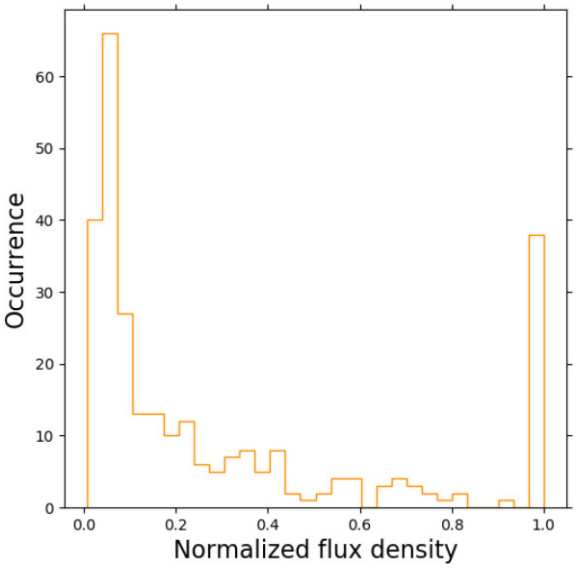

Figure 10. Histogram of normalized flux density of each fragment in the sample. We normalize the flux density per field by the highest flux density fragment in the each field, $S_{frag}/S_{brightest}$.

4.3 Emission properties

Beyond the physical structure of the observed fields, an examination of the distribution of observed flux density within each region was conducted. This analysis aimed at resolving whether the protoclusters comprise several equally bright fragments or are dominated by a single high flux density fragment. Due to relatively small numbers of fragments in each field, the combination of data across all observed fields was used to assess the general trend of flux density distribution within the sample.

Fig. 10 gives the distribution of fragments, over all target fields, as a function of normalized flux density. The normalized flux density in each field was defined as the division of each individual fragments observed flux density by that of the fragment with the highest flux density in its host field. As such the brightest fragment in each field will have a normalized flux density of 1 (and clearly seen in Fig. 10) and all other sources value <1.

It is clear from Fig. 10 that the TEMPO fields appear dominated by single (or infrequently a very small numbers) of bright fragment(s) with the remainder of the population being significantly fainter. Across the whole sample the majority (69.4 per cent) of fragments have <20 per cent of the flux density of the brightest fragment in their respective field.

To assess this, the ratio of the flux density of the brightest object to the sum of the flux density of all other fragments in a given field was calculated as, $\frac{S_{max}}{\Sigma S_{other}}$, hereafter termed S_{budget}. This value would be $\sim\leq 1$ if the 'faint' field fragments dominate the flux density budget or >1 if the brightest fragment dominates. Of the 38 TEMPO fields, 22 fields have an $S_{budget} \leq 1$ and as such the fainter fragments dominate the flux density budget, suggesting that the flux density is relatively evenly distributed amongst the fragments in these fields.

For the 16 fields with $S_{budget} > 1$, indicating the flux density distribution is dominated by one (or a small number of) fragment(s), the ratio of the brightest fragment in that field to the second brightest was calculated. This allowed assessment of whether the flux density budget is dominated by a single source. Of these 16 fields, 14 contain a bright fragment which has a flux density at least 3× greater than that of the next brightest fragment in the field and as such these fields

appear dominated by a single high flux density object. The remaining two fields (SDC18.816−0.447_1 and SDC30.172−0.157_2) contain a second fragment with between 0.83× and 0.91× the flux density of the brightest, with the remaining fragments in these fields not contributing significantly to the flux density budget. For these two fields, it is noted that both are found to contain only two fragments, and that these two fragments are separated by 0.12 pc (5.9 arcsec at a distance of 4.3 kpc for SDC18.816−0.447_1) and 0.06 pc (3.4 arcsec at a distance of 4.2 kpc for SDC30.172−0.157_2). Not accounting for projection these separation are larger than λ_J for SDC18.816−0.447_1 and smaller than λ_J for SDC30.172−0.157_2 as calculated in Section 4.2 (cf. Table 1). It is apparent from the TEMPO fields that whilst the faint fragments dominate the number counts they do not typically dominate the flux density budget in a given field.

Whilst is is possible to equate the measured flux density of a fragment to a mass for that fragment, this has not been attempted within the current work for the following reason. Given the expectation that each small scale fragment is internally heated by an evolving protostar, then to derive a meaningful masses would require knowledge of the temperatures of each fragment. This cannot be derived from the the continuum flux density alone and as such the analysis has been limited to discussion of flux density. Further investigation of the masses of the observed fragments will be conducted under a future work, when a more detailed analysis of the chemical properties of the TEMPO sample has been completed. Such an analysis should gives a reliable way to estimate temperatures and calculate meaningful masses.

4.3.1 Brightest source properties

Given the dominance, in terms of flux density, of single or small numbers of fragments within each TEMPO field, an analysis of the properties of these objects with respect to high-mass star-formation tracers, and their relative position in the TEMPO field was undertaken. Three samples were considered, in addition to the brightest fragment per field (sample size 38, one per field). Those being methanol maser associated TEMPO fragments (sample size 27, explained in next paragraph), IR object associated fragments (sample size 38), and the sample of the most central fragment in each TEMPO field (e.g. those fragments located closest to the non-intensity-weighted mean position in each TEMPO field, sample size 38).

There are 28 TEMPO fields with a known 6.7 GHz CH_3OH maser source within the ALMA primary beam (see Table 1) in each case there is only a single maser within the ALMA primary beam. A maximum offset limit between a TEMPO fragment peak position and the maser position of 2 arcsec (equivalent to a physical separation of $\sim$ 0.04 pc at the average source distance of 3.9 kpc) was applied to assign maser association with a TEMPO fragment. With this limit, the maximum offset retained is 1.4 arcsec (a physical separation of 0.03 pc at the assumed target distance). All other source-maser offsets are below this, with a minimum of 0.07 arcsec (0.8 Mpc at the source distance). This offset limit excludes the maser in field RMS-G034.8211 + 00.3519 for which the maser is offset by $\sim$ 13.8 arcsec from the nearest TEMPO source. It should be noted that the maser in this field only has a position recorded from the single dish *Parkes Radio Telescope*, rather than an interferometric position from ATCA in the MMB catalogues. Thus its positional accuracy is significantly lower.

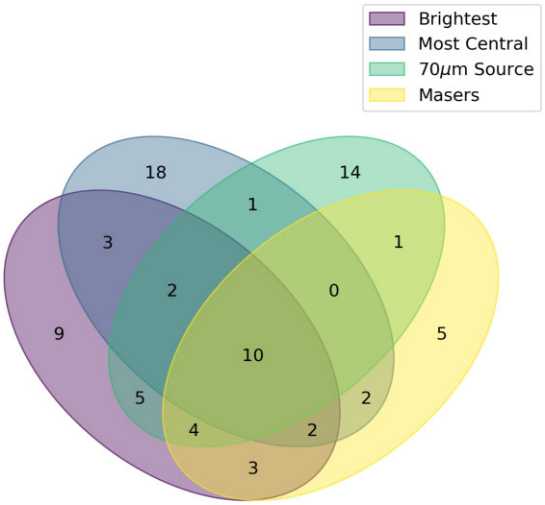

Figure 11. Venn diagram showing the overlap of samples comprising, the brightest TEMPO field fragment (purple segment, 38 fragments total), the high-mass star forming tracing CH_3OH masers associated fragments (yellow segment, 27 fragments total), IR sources associated sources (detections at 70 μm by *Herschel* (Elia et al. 2021), green segment, 38 fragments total), and the most central fragments in each TEMPO fields (arithmetic average, as the blue segment, 38 fragments total).

Infrared sources were drawn from the Hi-GAL catalogues (Elia et al. 2021) at 70 μm. Given the angular resolution of these Hi-GAL data, the maximum offset limit between the TEMPO fragment peak position and the IR source was limited to 5 arcsec (equivalent to a physical separation of $\sim$ 0.09 pc at the average source distance of 3.9 kpc) following the approach used by Jones et al. (2020) for Hi-GAL – maser association. In cases where multiple TEMPO sources fell within this cutoff the source with the smallest offset was deemed the associated source. Using this limit, the maximum offset retained was 2.77 arcsec (a physical separation of 0.05 pc at the assumed target distance).

Fig. 11 is a Venn diagram of the considered samples, with the given values indicating the number of fragments in each overlapping set. From this figure it can be seen that the brightest fragments in the TEMPO fields are commonly associated with the other sample types, with 76 per cent of the sample (29/38) being a member of at least one of the other sets. Looking at two sample comparisons the brightest TEMPO fragments are, perhaps unsurprisingly, most commonly associated with 70 μm IR sources (55 per cent of fields), followed by CH_3OH masers (in 50 per cent of fields), and are also the most central source in their respective field in 45 per cent of cases.

For the 50 per cent of fields which do not show a maser-brightest TEMPO source association, 11 (29 per cent) do not have a maser detection and of the remaining eight sources, six have the second brightest source in the field associated with the maser. Viewed another way, in 70 per cent (19/27) of the TEMPO fields with a maser, the maser-associated fragments is also the brightest fragment. Such high overlap in membership of the brightest fragment and maser associated samples indicates that the brightest fragment in each field is a good proxy for the local high-mass star forming core candidate. All TEMPO fields were covered by the MMB survey at 6.7GHz meaning fields without a maser are due to a non-detection during that survey, not a lack of observational data. Thus the absence of CH_3OH masers in 11 of the TEMPO fields may be indicative of a younger evolutionary stage in those fields, making the brightest fragments within these fields good candidates for follow-up maser

observations to detect emergent masers or weak masers which were below the detection limit of the MMB survey. Alternatively, the absence of masers may simply be an inclination effect due to the beamed nature of maser emission.

Making the broad assumption that the brightest fragment in each field is also the most massive, it is interesting to note that the 55 per cent of TEMPO fields do not have the most massive fragment at their central position. High-mass Main Sequence stars are more commonly seen at the centre of stellar clusters and under the *clump-fed* model are expected to spend at least part of their evolution there. This result is suggestive of either, some TEMPO fields being in early stages of evolution prior to the migration and settling of more massive cores at the cluster centre or the TEMPO observations are limited in either sensitivity or field of view meaning the sample are missing weaker (or out of field) sources thus skewing the true central position. Of course, a more robust investigation of the masses in the TEMPO sample requires a temperature measurement of each source (not just the clump temperature stated in Table 1) which is beyond the scope of this work and will be addressed in a future paper.

4.4 Looking for signatures of evolution

A primary goal of the TEMPO survey was to look for evidence of evolution, or lack thereof, with the fields observed. Two of the properties derived from the continuum maps are worthy of note when inspected against the target clump luminosity[10] from Elia et al. (2021). These are namely the number of fragments (Section 4.1.1) and percentage bandwidth which is spectral line-free within the data (column 9 in Table 1). Beyond these two properties little indication of evolutionary trends are seen within the analysis conducted for this paper. Chemical and kinematic analysis of the TEMPO sample are to be published in future works (Asabre Frimpong et al. (in preparation) and Wang et al. (in preparation)).

4.4.1 Number of fragments

Fig. 12 gives the number of fragments extracted from the TEMPO fields as a function of clump luminosity from the work of Elia et al. (2021). Here we see no clear correlation between these two properties. This is note worthy as in a typical star-forming scenario as the source evolves the power output from the bipolar outflows will increase. Given this, one could expect greater disruption of the material in the field and thus a greater amount of fragmentation in more evolved clumps, something not seen in the TEMPO fields.

4.4.2 Spectral line-free bandwidth

A 'by-product' of the LUMBERJACK (Section 2.3) analysis conducted to find spectral line-free channels within the TEMPO data, the value of percentage bandwidth used in continuum, is also a measure of a fields line richness. The lower the available bandwidth for continuum imaging the higher the spectral line density within the target.

Fig. 13 gives the percentage of the total observed ALMA bandwidth used in generating the continuum images as a function of clump luminosity (Elia et al. 2021). Here there is evidence of a tentative correlation between $L_{\rm clump}$ and percentage line-free bandwidth (albeit with a large scatter at any given luminosity), with lower luminosity clumps having less line-free bandwidth (ergo more line rich) and higher luminosity clumps having a greater available bandwidth for continuum imaging (thus less spectral line emission). This could be explained in terms of evolution as the destruction of complex molecular species by the increasing radiation output of an evolving source as its luminosity increases.

Also plotted in Fig. 13 are the average values of two TEMPO field sub-samples, those which are 6.7GHz maser associated (hexagon marker, with average $L_{\rm clump} = 4.5 \times 10^4$ L$_\odot$ standard deviation 4.9×10^4 L$_\odot$, and percentage bandwidth = 28.6 with standard deviation 17.6) and those which are not (triangle marker, with average $L_{\rm clump} = 6.9 \times 10^4$ L$_\odot$ standard deviation 1.4×10^5 L$_\odot$, and percentage bandwidth = 35.0 with standard deviation 14.8). A small offset is seen between these two samples which suggests that

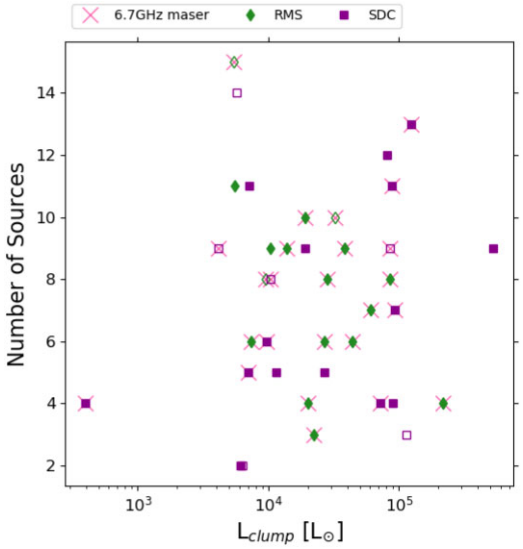

Figure 12. Number of detected fragments as a function of clump luminosity (Elia et al. 2021). Symbols are as per Fig. 1.

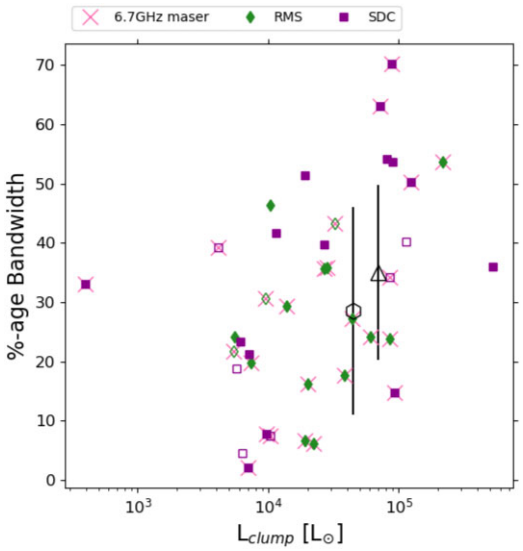

Figure 13. Percentage spectral line-free bandwidth in the TEMPO ALMA data as a function of clump luminosity (Elia et al. 2021). Symbols are as per Fig. 1, plus the black hexagon and black triangle mark the average values for the TEMPO field with and without an associated 6.7GHz CH_3OH maser, respectively. And the associated error bars give the scale of 1 standard deviation of these samples.

[10]Here luminosity acts as a proxy of age, with lower luminosity indicating younger star-forming clumps and vice versa.

the lower luminosities, thus the younger samples, are preferentially the maser associated sources. Again this aligns with expected evolutionary traits of the methanol maser, which are thought to be destroyed as protostellar luminosity increases (Breen et al. 2010; Jones et al. 2020).

4.5 Distinguishing between star-forming and non-star-forming fragments

To conclude the discussion of the detected fragments within the TEMPO fields, an initial analysis into the nature of the detected fragments was conducted with the aim of distinguishing between fragments likely to be star-forming cores and those which are not star forming, simply fragments (as they have hitherto been referred). This analysis compares the phase and amplitude properties of simulated interferometric visibility data of point sources, Gaussian profile sources and Gaussian plus point (hereafter Gaussian+Point) sources to the observed TEMPO visibility data. The three model profiles used were selected as a basis of comparison with the TEMPO fragments under the assumption that such profiles are likely to be present in actively star-forming cores. Particularly point-like, ergo unresolved, objects and point-like objects within extended envelopes. The use of a Gaussian profile as a comparison was a pragmatic choice as it provides a simple and quantifiable model of a centrally peaked extended emission. A full description of the approach used is given in Appendix C, whilst a summary is given in the following paragraphs.

A catalogue of point-like, Gaussian and Gaussian+Point source simulated data sets were created using the CASA task `simobserve`. The simulated data matched the TEMPO typical *rms*-noise, FOV, synthesized beam shape, frequency tuning, and bandwidth. The simulated data sets were created to cover a range of signal-to-noise ratios (SNR), differing source axis ratios (in the case of Gaussian & Gaussian+Point models) and differing Gaussian peak emission to Point source peak emission ratios (for Gaussian+Point models).[11]

For each SNR, axis ratio, and peak emission ratios 100 simulated data sets were generated, each with a different thermal noise spatial distribution, controlled by a random seed value within `simobserve`. From the simulated data set the amplitude and phase values were extracted at the position of the model source within them. Each simulated data set contained a single source. The simulated amplitude and phase values were then used to generate empirical relations between SNR and amplitude and phase properties (cf. Appendix C5).

The same amplitude and phase properties were then extracted for each detected fragment in the real TEMPO data and compared to the relations generated from the simulated data. Based on the TEMPO fragment properties at its recorded SNR a decision tree (see Fig. C4) was followed to categorize each fragment into being either a point-like source (given a score of 1), Gaussian profile source (score of 2), Gaussian+Point source (score of 3), or other morphology (score of 0). TEMPO fragments with scores of $\geq$1 are considered active star-formation candidates (ASCs) with this score hereafter referred to as the ASC_{score}. Fig. 14 plots the breakdown of ASC_{score} for the 287 sources detected in TEMPO and the ASC_{score} of each source are given in column 13 of Table 4.

[11]The simulated data were generated with SNR values of 5, 10, 15, 20, 30, 40, 50, 60, 80, 100, 120, 150, 200, 300, 500, 750, and 1000. The SNR was defined as the peak pixel emission to off-source noise ratio. For Gaussian models axis ratios of 1:1, 2:1, and 3:1 were used. For Gaussian+Point models, Gaussian peak emission to point peak emission ratios of 1:1, 1:0.5, and 1:0.1 were used.

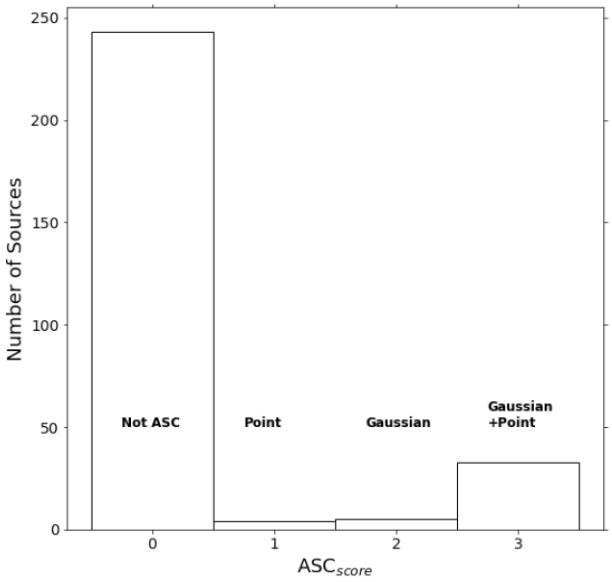

Figure 14. Distribution of ASC_{score} assigned to the TEMPO detected fragment sample. Individual scores per source are given in Table 4.

Within the TEMPO sample, 42 fragments are found with an $ASC_{score} \geq 1$. Hereafter, these 42 fragments (14.6 per cent of the sample) are discussed together and labelled as active star-formation candidates (ASC) sources. The remainder of fragments are considered not to be currently star forming, as we do not see the required characteristics within our current data. This does not exclude the possibility that they are prestellar and in the future may coalesce further to go on to form protostars nor that they currently are star forming but the recovered visibility data do not allow confirmation within these data. Alternately, fragments with $ASC_{score} < 1$ are possibly clumps of material created by, for example, the disruptive effects of outflows from the protostellar sources (e.g. Arce et al. 2007; Rosen & Krumholz 2020). A full investigation of the gas kinematics and outflow properties of the sample will appear in a future works from the TEMPO project. Across the TEMPO sample 31 field have at least 1 ASC source, with only seven fields having no detected ASC source. Specifically the fields without ASC are RMS-G017.6380+00.1566, SDC24.381-0.21_3, SDC28.147-0.006_1, SDC30.172-0.157_2, RMS-G034.8211+00.3519, SDC45.787−0.335_1, and RMS-G332.9868−00.4871.

Fig. 15 shows the distribution of SNR values for ASC sources overlaid on the SNR characteristics for the whole TEMPO sample. There is a fixed lower limit of SNR equal to 30 for ASC sources as specified in Appendix C. It is clear that ASC sources are drawn from across the SNR parameter space and a large fraction is in the low SNR regime. This is to be expected for two reasons. First, the low flux density (thus low SNR) fragments dominate the fragment counts in the TEMPO sample (cf. Section 4.3), and secondly, the bounding conditions for ASC acceptance are broader at lower SNR (cf. equations C4 through C12). This latter point may also account for some of the high SNR fragments not being included in the ASC sample in that the stricter bounds at high SNR may exclude sources which are close to, but not within, those bounds. Though of course high flux density in the mm-wavelength regime does not automatically indicate a star-forming source.

A third factor which maybe account for the exclusion of some high SNR fragments from the ASC sample is seen in the inspection

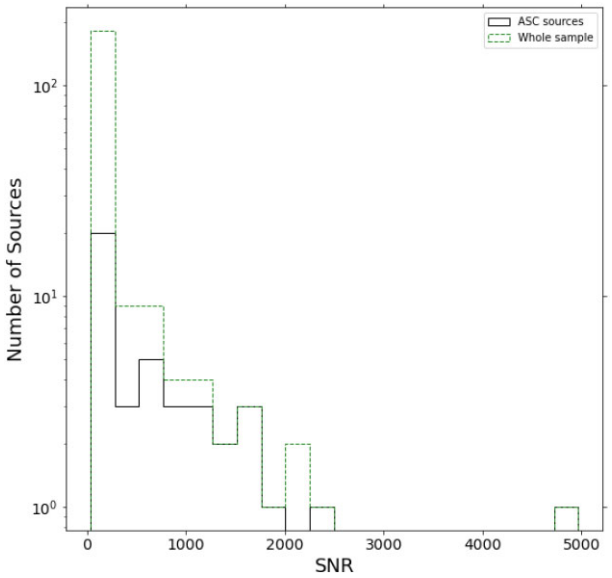

Figure 15. Distribution of SNR for ASC sources ($ASC_{score} > 0$). The whole sample are given in green-dashed and ASC sources as the black solid.

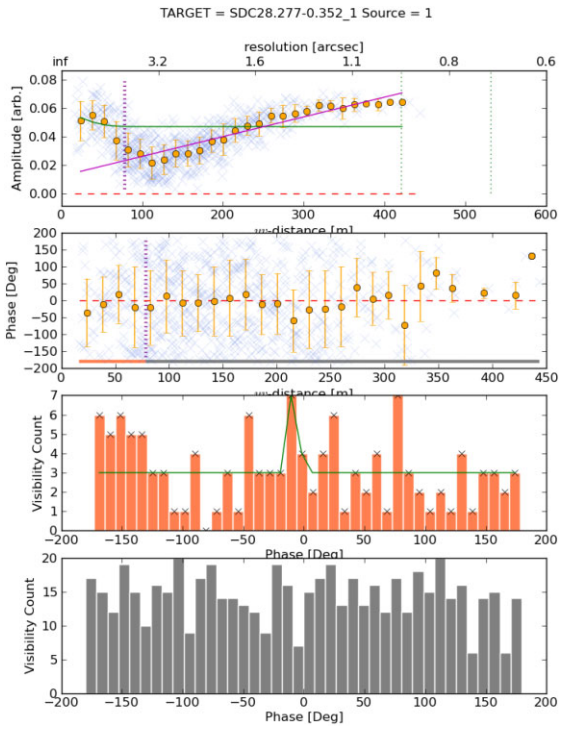

Figure 16. Visibility analysis plot for SDC28.277−0.352_1 fragment 1 cf. Fig. 3(a). Showing signs of poor visibility subtraction using a single Gaussian component at the brightest field source position.

of the post ASC analysis visibilities. The method used to extract the amplitude and phase data, uses a model subtraction of all other sources in the field to reduce their impact on the target sources visibility characteristics. However, inspecting the post-subtraction plots and images, some bright single sources do not appear well fitted by a single simple Gaussian. Fig. 16, presenting the visibility analysis plot for SDC28.277−0.352_1 source 1 is a good example of such a problem, in which some residual Gaussian-like profile in amplitude

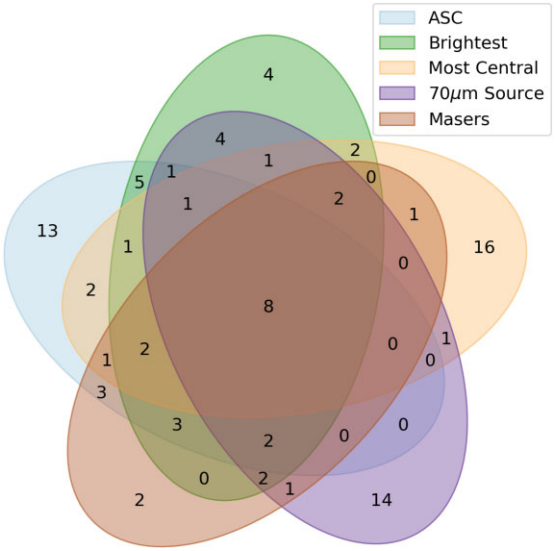

Figure 17. Venn diagram showing the overlap of samples comprising, the active star-formation candidates (ASC; pale blue segment, sample size of 42), the CH_3OH masers associated fragment sample (red segment, sample of 27 fragments), infrared object associated fragments (detections at 70 μm by *Herschel* (Elia et al. 2021), purple segment, sample of 38 fragments), the brightest fragments in each TEMPO fields (green segment, sample of 38 fragments), and the most central fragments in each TEMPO fields (arithmetic average, as the orange segment, sample of 38 fragments).

and a large scatter in phase after source subtraction can be seen. A more robust modelling of the fragments, at the sub-resolution scales would be required to account for these kinds of source. Achieving this for the TEMPO sample size is beyond the scope of this current investigatory analysis.

4.5.1 ASC characteristics

The ASC source sample was compared with the same three source samples as in Section 4.3.1 to inspect for common characteristics within the ASC sample. The source samples used in the comparison were, 6.7GHz CH_3OH maser associated sources, brightest TEMPO field sources, most central TEMPO field sources (using the arithmetic mean of field source positions). The latter two sample have a size of 38 (one per TEMPO field).

Fig. 17 is a Venn diagram of the overlapping membership of the five samples. There is some observed overlap between members of the ASC sample and the brightest field fragment (22/38, 58 per cent), maser associated fragments (18/27, 67 per cent), 70 μm IR source (16/38, 37 per cent), and most central field fragment (14/38, 37 per cent).

Such correspondence between the ASC candidates and star-forming core indicators (maser and brightest field source particularly), within the initial implementation of the described visibility analysis is a good indicator of the validity of the method. It provides additional constraints on those fragments within the TEMPO sample which are likely truly star forming.

However, it is noted that in cases where the brightest fragment, CH_3OH maser hosting fragment and/or IR counterpart fails to meet our ASC_{score} criterion, visual inspection of the target field, and post-analysis visibilities reveals indications that these sources are more complex than simple, single point-like or Gaussian sources and potentially maybe unresolved multiple systems. The technique

also suffers from a requirement to know the exact position of an ASC source very accurately to recover the extracted visibility information without position errors affecting the recovered phase (see Appendix C7 for more detail). The technique could be extended and modified to include a more general visibility parameters space analysis to determine better positions.

5 DISCUSSION

The detected fragments in the TEMPO sample fields do not display a simple radial profile and may exhibit fractal or other distributions. This finding appears to agree with those found in other clusters, both observed and modelled, from within the literature. Though quantitative comparisons based on the Q-parameter cannot be made for the TEMPO fields, owing to the discussion given in Appendix B, qualitatively we find similarities to a number of young star forming clusters.

The L1622, NGC2068/NGC2071 and NGC2023/NGC2024 star-forming regions within Orion B are all found to be mildly substructured by Parker (2018; all with Q <0.8 for source numbers of 29, 322, and 564, respectively), though they caution the use of the Q value for the limited number of sources in the case of L1622 due to its low source numbers (for the same reasons discussed in Appendix B).

Sanhueza et al. (2019) also use the Q-parameters on their IRDC derived sample finding in the majority of cases values indicative of substructure (Q<0.8). However, it is noted that the source numbers in the Sanhueza et al. (2019) sample are between 13 and 37, so the validity of using Q with these fields is unclear. Despite this, visual inspection of the reported fields, particularly when considering the published MST (their figs 5–10) shows that most fields within their sample appear to contain some substructure. The region NGC 6334 I(N) was found by Hunter et al. (2014) to be close to a Q indicative of uniform density (0.82), though again for small source numbers. From the associated MST for this region whether or not the region is substructred or has a radial profile is unclear.

Using an alternative parameter, $\delta_{\rm ADP,\,N}$ to gauge the level of substructure in the Orion Nebular Cluster (ONC) Da Rio, Tan & Jaehnig (2014; see their equation 1) find that this more evolved stellar cluster has a low level of substructure (see also Bate, Clarke & McCaughrean 1998). These authors note that the ONC appears somewhere between the substructured young Taurus molecular cloud (see e.g. Cartwright & Whitworth 2004) and the radial distributions seen in globular clusters.

Indeed, there is evidence within the literature from both observed and modelled young stellar clusters that cluster structure tends to evolve from an originally sub-structured formation toward a centrally concentrated final state (e.g. Bonnell, Bate & Vine 2003; Schmeja & Klessen 2006; Bate 2009; Maschberger et al. 2010) as sufficient time for dynamical processing of the sources within the cluster elapses. The TEMPO sample was selected to give a range of ages of high-mass embedded protostars prior to the formation of an UCH II region, as such some degree of substructure at these early times would be expected. This evolution of structure may also account for the distribution of source normalized offsets seen in Fig. 9 with some fields beginning to show a more centrally concentrated profile than others. However, the TEMPO sample lacks sufficient source counts within individual fields to test this quantitatively as a function of e.g. IR colour. To further this analysis higher sensitivity and resolution images of the TEMPO fields is required to detected any fainter sources present and to resolve closely paired objects, which may currently appear as a single source within the TEMPO data.

The majority of the TEMPO fields show fragmentation on scales which are inconsistent with (with 87 per cent of fields having a mean edge length $X_{\rm mean}$ ≥1.5 × up to 9.1× the $\lambda_{\rm J}$) the thermal Jeans length when using the clump radii, mass and temperatures from Elia et al. (2021) within the calculation. This is suggestive of a non-thermal fragmentation being present within the TEMPO fields. Similar results have been seen within other works. Traficante et al. (2023) found in the SQUALO sample found a range of values of source separation to thermal Jeans length ratio (their $\lambda_{\rm J,\,3D}$) of 1.06 <$\lambda_{\rm J,\,3D}$ < 7.04, suggestive of some non-thermal fragmentation. SQUALO had similar observing characteristics to the TEMPO sample. Observations made over larger spatial scales, using mosaic rather than single pointing observations, also tend to find fragmentation scales which are better explained by turbulent or cylindrical fragmentation (Henshaw et al. 2016; Lu et al. 2018).

Conversely, the results seen by Svoboda et al. (2019) targeted toward high-mass starless clump candidates, the ASHES sample (Sanhueza et al. 2019) toward 70 μm dark high-mass clumps and in the CORE survey (Beuther et al. 2018) toward known high-mass star-forming regions found fragmentation scales consistent with the thermal Jeans length scale. It is interesting to note that the calculation of the thermal Jeans length (equation 3) is particularly sensitive to the value of $R_{\rm clump}$ used, as it scales with $R_{\rm clump}^{3/2}$. Using different $R_{\rm clump}$ values for the TEMPO fields, for example those calculated by Traficante et al. (2015) for the SDC fields and Urquhart et al. (2014) for the RMS fields, brings the TEMPO field $X_{\rm mean}$ values in to a more comparable range with thermal Jeans (in the range 0.5–1.5× $\lambda_{\rm J}$). Such sensitivity to changes in input is important to consider when it has such an impact on the findings. The use of Elia et al. (2021) values has been retained within this work to allow use of a single consistently derived set of parameters from the literature.

The emission fraction values within the TEMPO sample appears to be largely consistent with other high-mass star-forming regions which have been independently studied. For example, in the study of NGC6334 I(N) Hunter et al. (2014) find ∼83 per cent of their sources have a flux density < 20 per cent of the bright source (derived from values in their Table 2). The Hunter et al. (2014) data has a slightly lower sensitivity to TEMPO[12] with ΔS = 2.2 mJy per beam. Both studies were conducted at 1.3 mm.

The work on G28.34 + 0.06 P1 also at 1.3 mm by Zhang et al. (2015), however, finds lower values of 47 per cent of sources with <20 per cent of the highest flux density. The Zhang et al. (2015) data is slightly higher sensitivity to the mean value of the TEMPO data at 0.075 mJy per beam. Whilst not as high percentage as those reported by Hunter et al. (2014) and the TEMPO result, Zhang et al. (2015) note in their paper there is an under abundance of low mass cores in their target field, which would drive the low mass percentage down in this source. The authors suggest this may be caused by lower mass stars forming later and the trend seen across these various studies may be indicative of the relative evolutionary stages across the studied sources.

6 CONCLUSION

The TEMPO survey conducted a high resolution (0.8 arcsec), high sensitivity (mean *rms*-noise ∼0.26 mJy equivalent to ∼1.0–2.5 $M_\odot$ for T = 30/15K, respectively) ALMA survey of 38 colour-luminosity selected high-mass star-forming regions. The continuum emission from fragments within the survey sample fields has been imaged

[12]cf. TEMPO ΔS = 0.26/0.23/0.69/0.09 mJy (mean/median/max/min).

and the clustering, fragmentation, and distribution of emission has been assessed. Additionally we have undertaken analysis to gauge whether the observed sources are matter over-densities or centrally condensed (and therefore likely currently star forming).

Our key findings are given in the following bullet point list:

(i) Each field has between 2 and 15 detected fragments (average 7.6).

(ii) The observed clusters in our target fields do not show distributions consistent with a simple radial profile ($r^{-\alpha}$) for $\alpha = 0,1,2$ but it is possible to exclude higher α values.

(iii) The Cartwright & Whitworth (2004) Q-parameter does not work to distinguish fractal from radial cluster distributions for small number ($N < 15$) clusters. See Appendix B.

(iv) The fragmentation scale, calculated as the mean edge length for the MST in each field, is not consistent with thermal Jeans fragmentation for the majority of fields in the TEMPO sample. With 33 (87 per cent of the sample) having a mean edge length, X, greater than or equal to 1.5× the thermal Jeans fragmentation scale suggesting that some other mode of fragmentation may be in effect in these fields. The remaining five fields have fragmentation scales comparable with (or in one case) smaller than the thermal Jeans length.

(v) Across the whole sample the majority (∼69 per cent) of detected fragments have a low flux density compared to the brightest source in that field, where low is defined as <20 per cent of the flux of the brightest fragment in their respective fields. The flux budget within the TEMPO fields is divided approximately evenly, 47 per cent:53 per cent between fields where the sum of low flux density fragments is greater than that of the brightest field fragment, and where the highest flux density fragment dominates. For the latter fields, predominantly the brightest fragment has greater than 3× the flux density of the next brightest object, indicating that these fields are truly dominated by a single high flux density object.

(vi) The brightest fragment in each TEMPO field is commonly associated with high-mass star formation activity as traced by class II 6.7 GHz CH_3OH maser (70 per cent of fields with a maser present) and with the local 70 μm source (55 per cent of fields). This suggests a good correlation between the brightest 1.3-mm TEMPO fragment and the high-mass star forming core candidate in each field.

(vii) Two noteworthy trends are seen when comparing derived properties from the TEMPO continuum maps to clump luminosity. First, the number of fragments detected shows no correlation with increasing luminosity. Given outflow power from the evolving protostar(s) in each field could be expected to increase with age, the disruption of nearby material and thus the number of fragments could be expected to increase. This is not seen. Secondly, the amount of spectral line-free bandwidth for each source shows a weak positive correlation with increasing luminosity, suggesting the younger (lower luminosity) fields are more line rich than their more evolved (higher luminosity) counterparts. Splitting the TEMPO sample between those with an associated 6.7 GHz CH_3OH maser and those without, there is some indication that the maser associated sub-sample tends toward the younger, lower luminosity fields. Something which is expected from maser lifetime and pumping mechanism literature.

(viii) The interferometric visibilities properties of the TEMPO fragments were investigated and compared to those of point-like, Gaussian and Gaussian+Point profiles, to provide an indication of the centrally condensed nature of the fragments and thus whether they are actively star forming or not. This implemented technique recovered 42 fragments (∼ 15 per cent of the sample) which match the empirically derived criteria to be considered actively star forming at their respective SNR. These ASC show a high correspondence with the class II CH_3OH masers sample (67 per cent) and 70 μm IR sources (37 per cent of sample). It is noted (cf. Appendix C) that the visibility analysis applied suffers some limitation for complex and potentially unresolved objects requiring further analysis beyond the scope of this paper. However, it may be instructive to apply this technique to a wider sample of star-forming regions observed by ALMA to further establish an 'active star forming core' criteria in the ALMA-era, over reliance on classic clump scale tracers.

ACKNOWLEDGEMENTS

AA acknowledges support from STFC grants ST/T001488/1 and ST/P000827/1.

GAF acknowledges support from the Collaborative Research Centre 956, funded by the Deutsche Forschungsgemeinschaft (DFG) project ID 184018867. GAF also acknowledges financial support from the State Agency for Research of the Spanish MCIU through the AYA 2017-84390-C2-1-R grant (co-funded by FEDER) from grant PID2020-114461GB-I00, funded by MCIN/AEI/10.13039/5011000110 and through the 'Center of Excellence Severo Ochoa' award for the Instituto de Astrofísica de Andalucia (SEV-2017-0709).

Z.N. acknowledges funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement no. 716155 (SACCRED), from the ESA PRODEX contract no. 4000132054, and from the János Bolyai Research Scholarship of the Hungarian Academy of Sciences.

JEP was supported by the Max-Planck Society.

This paper makes use of the following ALMA data: ADS/JAO.ALMA#2015.1.01312.S. ALMA is a partnership of ESO (representing its member states), NSF (USA), and NINS (Japan), together with NRC (Canada) and NSC and ASIAA (Taiwan), in cooperation with the Republic of Chile. This research made use of ASTRODENDRO, a PYTHON package to compute dendrograms of Astronomical data (http://www.dendrograms.org/) and ASTROPY, a community-developed core PYTHON package for Astronomy (Astropy Collaboration 2013).

DATA AVAILABILITY

The observational ALMA data is publicly available via the ALMA Archive under project code 2015.1.01312.S. The data underlying the analysis within this article are available in the article and in its online supplementary material and will be shared on reasonable request to the corresponding author.

REFERENCES

Arce H. G., Shepherd D., Gueth F., Lee C. F., Bachiller R., Rosen A., Beuther H., 2007, in Reipurth B., Jewitt D., Keil K., eds, Protostars and Planets V, University of Arizona Press, Tucson, P. 245
Astropy Collaboration, 2013, A&A, 558, A33
Bate M. R., 2009, MNRAS, 392, 1363
Bate M. R., Clarke C. J., McCaughrean M. J., 1998, MNRAS, 297, 1163
Beuther H., Ragan S. E., Johnston K., Henning T., Hacar A., Kainulainen J. T., 2015, A&A, 584, A67
Beuther H. et al., 2018, A&A, 617, A100
Bonnell I. A., Bate M. R., 2006, MNRAS, 370, 488
Bonnell I. A., Bate M. R., Vine S. G., 2003, MNRAS, 343, 413
Breen S. L., Ellingsen S. P., Caswell J. L., Lewis B. E., 2010, MNRAS, 401, 2219

Breen S. L., Ellingsen S. P., Contreras Y., Green J. A., Caswell J. L., Stevens J. B., Dawson J. R., Voronkov M. A., 2013, MNRAS, 435, 524
Breen S. L. et al., 2015, MNRAS, 450, 4109
Cartwright A., Whitworth A. P., 2004, MNRAS, 348, 589
Caswell J. L. et al., 2010, MNRAS, 404, 1029
Caswell J. L. et al., 2011, MNRAS, 417, 1964
Churchwell E. et al., 2009, PASP, 121, 213
Da Rio N., Tan J. C., Jaehnig K., 2014, ApJ, 795, 55
Egan M. P., Price S. D., Kraemer K. E., 2003, in American Astronomical Society Meeting Abstracts. AAS Publishing, Washington DC, USA, p. 1301
Elia D. et al., 2021, MNRAS, 504, 2742
Fontani F. et al., 2016, A&A, 593, L14
Green J. A. et al., 2009, MNRAS, 392, 783
Green J. A. et al., 2012, MNRAS, 420, 3108
Henshaw J. D. et al., 2016, MNRAS, 463, 146
Hughes V. A., MacLeod G. C., 1989, AJ, 97, 786
Hunter T. R., Brogan C. L., Cyganowski C. J., Young K. H., 2014, ApJ, 788, 187
Jones B. M. et al., 2020, MNRAS, 493, 2015
Kennicutt R. C., Evans N. J., 2012, ARA&A, 50, 531
Lu X. et al., 2018, ApJ, 855, 9
Lumsden S. L., Hoare M. G., Urquhart J. S., Oudmaijer R. D., Davies B., Mottram J. C., Cooper H. D. B., Moore T. J. T., 2013, ApJS, 208, 11
McKee C. F., Tan J. C., 2003, ApJ, 585, 850
McMullin J. P., Waters B., Schiebel D., Young W., Golap K., 2007, in Shaw R. A., Hill F., Bell D. J., eds, ASP Conf. Ser. Vol. 376, Astronomical Data Analysis Software and Systems XVI. Astron. Soc. Pac., San Francisco, 127. p.
Maschberger T., Clarke C. J., Bonnell I. A., Kroupa P., 2010, MNRAS, 404, 1061
Massardi M. et al., 2021, PASP, 133, 85001
Minier V., Ellingsen S. P., Norris R. P., Booth R. S., 2003, A&A, 403, 1095
Molinari S., Pezzuto S., Cesaroni R., Brand J., Faustini F., Testi L., 2008, A&A, 481, 345
Molinari S. et al., 2010, PASP, 122, 314
Molinari S. et al., 2019, MNRAS, 486, 4508
Mottram J. C. et al., 2011, ApJ, 730, L33
Nony T. et al., 2018, A&A, 618, L5
Ossenkopf V., Henning T., 1994, A&A, 291, 943
Ostriker J., 1964, ApJ, 140, 1529
Parker R. J., 2018, MNRAS, 476, 617
Peretto N., Fuller G. A., 2009, A&A, 505, 405
Peretto N. et al., 2013, A&A, 555, A112
Pillai T., Kauffmann J., Wyrowski F., Hatchell J., Gibb A. G., Thompson M. A., 2011, A&A, 530, A118
Pineda J. E. et al., 2015, Nature, 518, 213
Remijan A. et al., 2019, ALMA Technical Handbook,ALMA Doc. 7.3, ver. 1.1, ALMA Technical Handbook
Rosen A. L., Krumholz M. R., 2020, AJ, 160, 78
Rosolowsky E. W., Pineda J. E., Kauffmann J., Goodman A. A., 2008, ApJ, 679, 1338
Sanhueza P. et al., 2019, ApJ, 886, 102
Schmeja S., Klessen R. S., 2006, A&A, 449, 151
Sokolov V. et al., 2018, A&A, 611, L3
Svoboda B. E. et al., 2019, ApJ, 886, 36
Tan J. C., Beltrán M. T., Caselli P., Fontani F., Fuente A., Krumholz M. R., McKee C. F., Stolte A., 2014, Protostars and Planets VI, University of Arizona Press, Tucson, p. 149
Traficante A., Fuller G. A., Peretto N., Pineda J. E., Molinari S., 2015, MNRAS, 451, 3089
Traficante A. et al., 2023, MNRAS, 520, 2306
Urquhart J. S. et al., 2014, MNRAS, 443, 1555
van der Tak F. F. S., van Dishoeck E. F., Evans Neal J. I., Bakker E. J., Blake G. A., 1999, ApJ, 522, 991
Vázquez-Semadeni E., Palau A., Ballesteros-Paredes J., Gómez G. C., Zamora-Avilés M., 2019, MNRAS, 490, 3061
Wang P., Li Z.-Y., Abel T., Nakamura F., 2010, ApJ, 709, 27
Wang K. et al., 2014, MNRAS, 439, 3275
Williams G. M., Peretto N., Avison A., Duarte-Cabral A., Fuller G. A., 2018, A&A, 613, A11
Xu Y., Li J. J., Hachisuka K., Pandian J. D., Menten K. M., Henkel C., 2008, A&A, 485, 729
Zhang Q., Wang Y., Pillai T., Rathborne J., 2009, ApJ, 696, 268
Zhang Q., Wang K., Lu X., Jiménez-Serra I., 2015, ApJ, 804, 141
Zinnecker H., Yorke H. W., 2007, ARA&A, 45, 481

SUPPORTING INFORMATION

Supplementary data are available at *MNRAS* online.

Table 4. Full Catalogue Table for all fragments detected in the TEMPO sample. Sample of this table is given in the manuscript.
Appendix A. Continuum maps of all target fields.
Appendix B. Using the *Q*-PARAMETER for small source counts.

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

APPENDIX C: THE USE OF INTERFEROMETRIC VISIBILITY DATA TO ASSESS WHETHER SOURCES ARE CENTRALLY CONDENSED

Interferometric telescopes provide data of higher spatial resolution than is possible from single dish instruments. However, interfer-

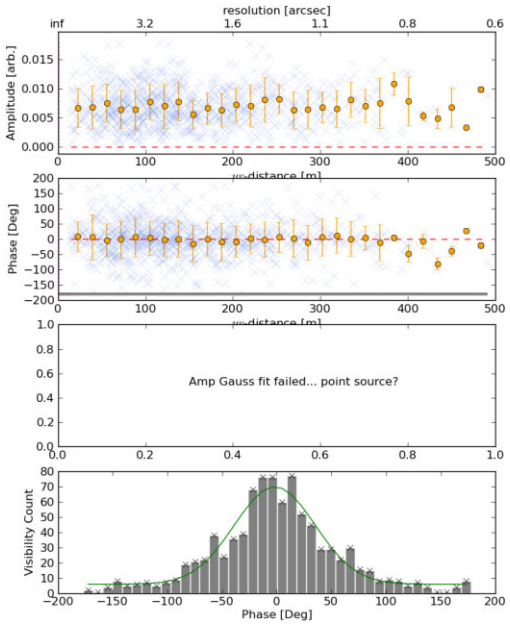

Figure C1. Simulated point source model visibility properties. The model has a SNR of 50. From top to bottom the panels give: (*Top panel*) The complex visibility amplitudes as a function of *uv*-distance, (*second panel*) the complex visibility phases as a function of *uv*-distance. In the *top* and *second* panels the blue 'x' are the visibilities values extracted from the simulated data and the orange circles averaged values in *uv*-distance bins. The errorbars associated with the binned data are ± 1 standard deviation. (*Third panel*) Blank for this model, as the fitting of the AG_FWHM parameter defined in the text was unsuccessful. (*Bottom panel*) A histogram of the unaveraged phase values. The green line describes a Gaussian fit to the data used to extract parameters the parameters for $ph2\_x0$ and $ph2\_FWHM$ as defined in the text.

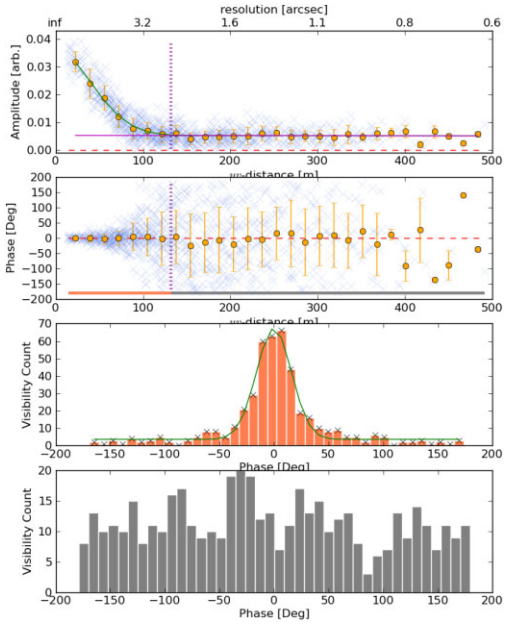

Figure C2. Simulated Gaussian source model visibility properties. The model has a SNR of 50 and major/minor axes of 2 arcsec. From top to bottom the panels give: (*Top panel*) and (*second panel*) as per Fig. C1. In the *top panel* the green line marks a Gaussian fit to the binned amplitude data. This is from the fit to measure the AG_FWHM parameter. The vertical dashed line gives the 3σ limit of the Gaussian fit. In the *second panel* the vertical dashed line gives the same 3σ limit as in the *top panel*. The coral and grey lines are used to denote colours used in plotting phase histograms in the *third* and *bottom* panels. (*Third panel*) A histogram of the unaveraged phase values in the $\leq 3\sigma$ *uv*-distance range (phase values with a *uv*-distance below the vertical dashed line in *second* panel. The green line describes a Gaussian fit to the data used to extract the $ph1\_x0$ and $ph1\_FWHM$ parameters referred to in the text. (*Bottom panel*) A histogram of the unaveraged phase values at *uv*-distances $>3\sigma$, the parameters for $ph2\_x0$ and $ph2\_FWHM$, are fit from data in this plot. In this specific case the fit is unsuccessful so now line is shown.

ometers lack sensitivity to structure on large scales. The maximum recoverable spatial scale achievable for an interferometric array is set by its minimum baseline length (distance between the two closest antennas in the array). This limitation results in the 'filtering out' of any emission from extended structures in the target field beyond this MRS. Such filtering can lead to images which appear to show several distinct clumpy regions, which are in reality only denser regions of a larger extended structure.

In addition to this imaging limitation, and of significance to this work, in star-forming regions there may exist dense clumpy regions which either do not yet contain a protostellar core or are simply transient phenomena which will never collapse to form a star.

In this appendix, we review the expected visibility properties for a set of simulated source models used to compare to the TEMPO source sample, discuss the simulated observations undertaken to define empirical relations between these simulated source model types and diagnostic visibility properties. Finally, discuss application of empirical relations of these diagnostic properties to the real TEMPO source data.

C1 Visibility theory

The cross correlation of signals for pairs of antennas in an interferometric array are used to measure complex visibilities, which are

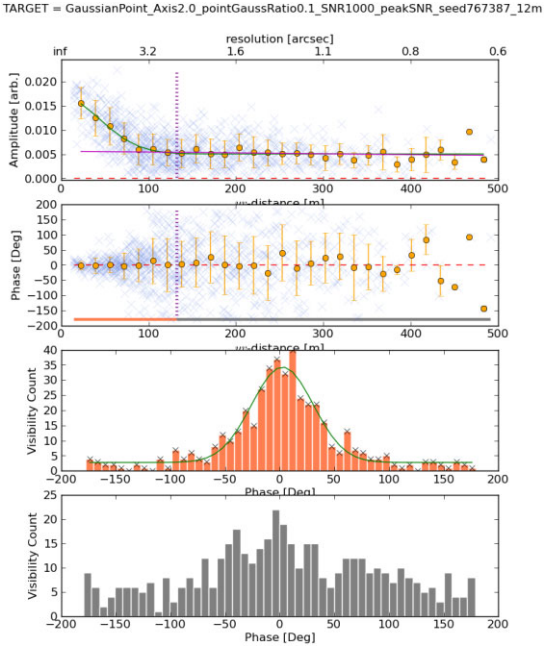

Figure C3. Simulated Gaussian+ point source model visibility properties. The model has a SNR of 1000, a major/minor axes of 2 arcsec, and a point source flux to Gaussian peak flux density of 0.1. From top to bottom the panels are as per Fig. C2.

the Fourier transform counterpart of the sky brightness distribution being observed.

Complex visibilities are of the form

$$V(u,v) = |V|e^{i\phi_V} = \int A(l,m)I(l,m)e^{-i2\pi(ul+vm)}\frac{\mathrm{d}l\mathrm{d}m}{\sqrt{1-l^2-m^2}}, \quad \text{(C1)}$$

where l and m are the direction cosines of a vector **s** from the phase centre of the observation. Interferometers measure one complex visibility per antenna pair per integration time interval. They are characterized by an amplitude,|V|, and phase, ϕ_V. The amplitude relates directly to the flux density of the sky brightness distribution on the spatial scales observable by a given antenna pair and the phase relates to the distribution of emission on the sky.

C2 Simulated source models for comparison with TEMPO data

To assess the nature of the sources in the TEMPO continuum source catalogue, an analysis of the Fourier/Visibility space properties of each object in the catalogue was undertaken. This analysis compared each detected TEMPO source to the properties of simulated source models for an unresolved point source (Point), a source with a Gaussian profile (Gaussian), and a source with a point source in a Gaussian envelope (Gaussian+Point), over a range of signal to noise ratios comparable to those seen in the TEMPO sample. The simulated source models used to compare to the TEMPO sources have the following properties in the visibility domain.

C2.1 Point source at the phase centre (Point)

An unresolved point source observed with an interferometer has two identifying characteristics in visibility space. First, as the point source will be unresolved on all baselines of the interferometer the amplitude component of the complex visibilities measured will be

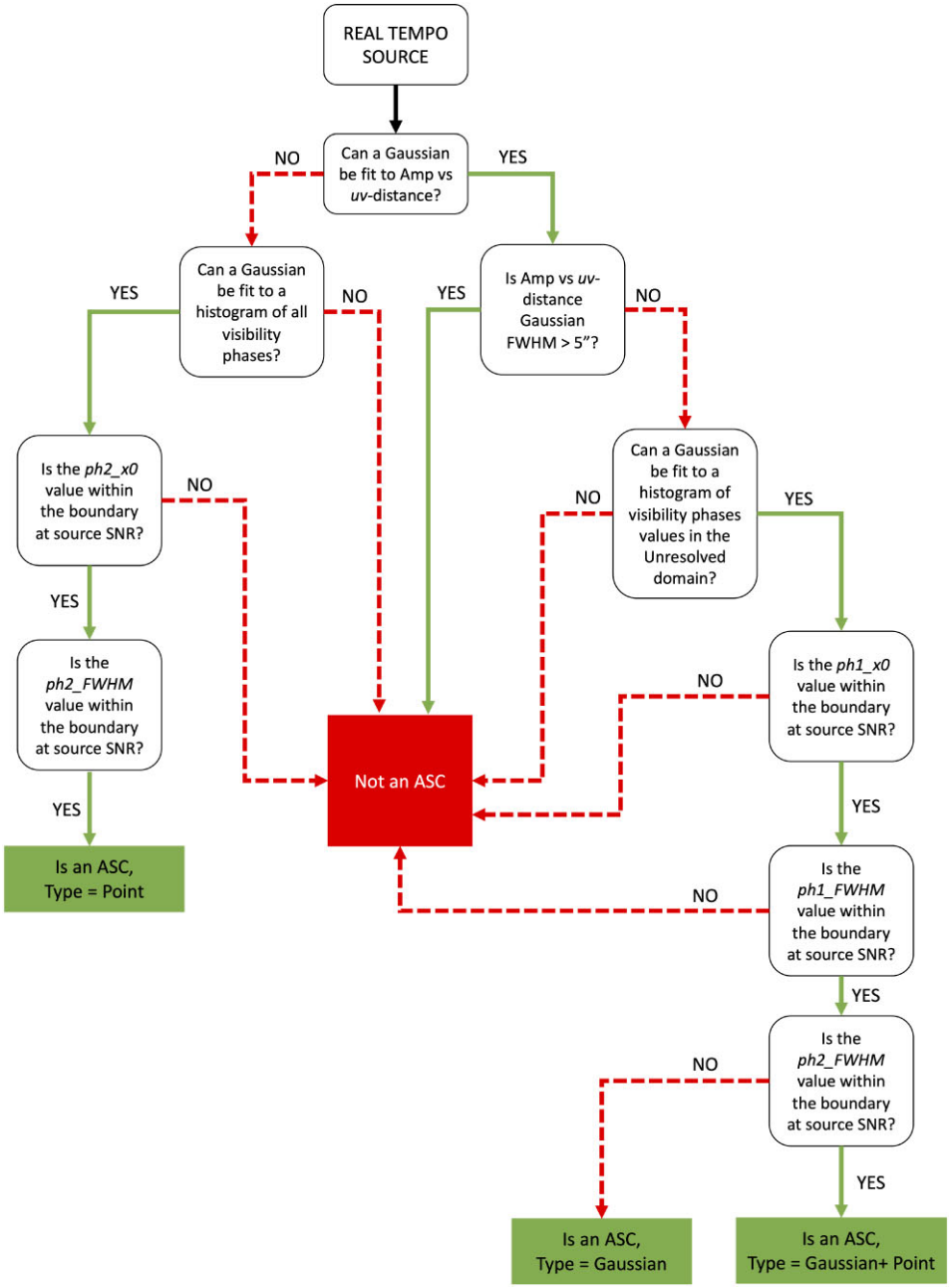

Figure C4. Decision tree used to determine if fragments within the TEMPO sample are actively star forming or not, and to specify the type if so. For each query a fragment either meets the criteria, and so follows the YES (solid green arrows) branch or fails to do so, and so follows the NO (red dashed arrows) branch until an end point is reached.

the same on all baselines of the array. Secondly, for an unresolved point source all emission is localized at a single position on the sky, thus when the point source is at the phase centre of the observation the phases of the observed complex visibilities are zero (as by definition u and v are zero at the phase centre). This is again true on all baselines.

C2.2 Gaussian source at the phase centre (Gaussian)

A Gaussian source has slightly more complex visibility characteristics. The visibility amplitudes will decrease as a function of the baseline length (and be at a maximum on the shortest baseline). Conceptually, it is perhaps easier to think about this feature in terms of the angular scales being probed. Shorter baselines are probing emission from larger spatial scales. As such, for a Gaussian source which is smaller than the angular scale measured by the shortest baseline the Gaussian appears unresolved and 100 per cent of its emission is being measured by that baseline. For increasing baseline length, the source begins to be resolved into smaller and smaller angular elements thus less emission is being recovered. For the phase properties, on shorter baselines, which provide measurements of angular scales greater than the extent of the Gaussian source, the phases appear point-like and are zero degrees. Beyond these baselines the phases become scattered away from zero.

C2.3 Point source within a Gaussian envelope at the phase centre (Gaussian + Point)

Combines the behaviour of the above two types of simulated source. The visibility amplitudes will decrease as a function of baseline length for baselines where it is possible to recover the emission of the Gaussian envelope. Beyond this at longer baselines the amplitude will be offset from zero as the point source at the centre will provide a constant amplitude to all baselines. Similarly in phase, at short baselines where the Gaussian is unresolved the phases will cluster around zero, as per a point source at those resolutions. The embedded point source will lead to a clustering of phases around zero degrees on all baselines, a signature which will distinguish it from a purely Gaussian profile source.

Real observations have a noise per visibility which contributes to the recovered amplitudes and phases and thus will cause a deviation from the idealised properties described above. Figs C1, C2, and C3 display in the upper two panels the described source properties as a function of *uv*-distance (equivalent to baseline length), with noise included. The noise in the amplitude and phase data impacts our ability to reliably compare the TEMPO sources to the idealised case, specifically for the low signal-to-noise sources. Steps to mitigate the impact of this are discussed in Section C3.

Another level of complexity to consider in real observations, is a field with more than one source of emission. Here emission from the sources which are not at the phase centre will cause deviation from the phase and amplitude properties the idealise cases described above. A step to mitigate the effects of this was used in this analysis and is described in C6.1

C3 Generating simulated source properties

The simulated observations were created using the CASA task `simobserve` and emulate closely the TEMPO observing characteristics Section 2.2. The simulated observing properties are as follows: An observing time of 300s, total BW of 1.875 GHz (equivalent to 1 SPW), a simulated percipitble water vapour (PWV) of 1.796 mm (a typical ALMA value for Band 6 observations), and the CASA ALMA array configuration file `alma.cycle6.3.cfg` was used, as it was closest to the true configuration used during TEMPO observations.

The simulated source models were generated using the CASA `componentlist` tools. Point sources were purely point like objects with a flux density set to give a desired SNR. For Gaussian models (Gaussian and Gaussian+Point) multiple size Gaussian sources models were simulated. Major axis values of 1, 2, and 3 arcsec were used in combination with major to minor axes ratios of 1:1 (radially symmetric), and 2:1 and 3:1 each with position angle set to 0 degrees in all cases. The major axis values chosen provide both marginally resolved (the 1 arcsec Gaussian) and fully resolved (2 and 3 arcsec Gaussians) in a simulated TEMPO synthesized beam.

For Gaussian+Point sources models the peak flux density is given by the addition of the point-like object and the Gaussian object, with the peak flux set to provide the required SNR. For these models a Gaussian to point flux density ratio was set to provide simulated sources where the point source object had a flux density equal to the Gaussian profile, half that of the Gaussian and a tenth of the Gaussian envelope.

In each case the model source was placed at the phase centre of the observation. The models created such that the peak emission divided by the off source noise (measured in a simulated blank sky of the same observing properties) gave a desired SNR. The SNR values used were 5, 10, 15, 20, 30, 40, 50, 60, 80, 100, 120, 150, 200, 300, 500, 750, and 1000.

At each SNR value, 100 simulations per model type were conducted (giving 62 900 simulations in total) with a unique random seed defining the 'phase screen' used to apply the noise to the data, effectively changing the distribution of the thermal noise in the recovered map.

C4 Measuring simulated source properties

With the suite of simulated observations, the CASA task `plotms` (with its graphical user interface deactivated) is used to record to text file the simulated amplitudes and phases as a function of *uv*-distance (in units metres, though the choice of *x*-axis values is arbitrary) at the phase centre. Given the data size, the visibilities were averaged up in both time and frequency channel and only the 'XX' correlation was recorded to reduce the number of data points required for analysis. After this the following values are extracted from the simulated visibility properties.

C4.1 Amplitude versus uv*-distance Gaussian profile (*AG_FWHM*)*

Here the amplitude and phase data of the recorded visibilities are first further averaged into 30 *uv*-distance bins to again reduce the data volume. An attempt is then made to fit a Gaussian profile to the amplitude as a function of *uv*-distance. If this fitting succeeds then the full width half-maximum (hereafter *AG_FWHM*) value of the Gaussian profile is recorded (in arcseconds, by conversion from metres to angular size at the observing frequency). If the fitting algorithm returns a FWHM greater than the maximum *uv*-distance or less than the minimum *uv*-distance the fit is rejected. In these cases, or if the fitting fails no *AG_FWHM* is recorded.

In cases where an *AG_FWHM* is recorded then the visibilities (no longer binned by uv-distance as above) are split between the 'Unresolved domain' which includes visibilities with *uv*-distances from the minimum *uv*-distance up to three-times the fitted Gaussian variance (σ)[13] and the 'Resolved domain', data *uv*-distances from 3σ to the max *uv*-distance. Where no *AG_FWHM* is recorded all the visibilities are considered to be in the 'Resolved domain'.

The upper panel of Fig. C2 shows a successful fit for the AG_FWHM parameter.

*C4.2 'Unresolved domain' visibility histogram centre and FWHM (*ph1_x0 *and* ph1_FWHM*)*

When an *AG_FWHM* is recorded this indicates that a Gaussian source is likely present in the model. On short baselines the (assumed) Gaussian will be marginally to completely unresolved and behave like a point source on these baselines, with the phases tightly clustered around zero. To measure this a Gaussian profile is fit to a histogram of the recorded phase values, in 50 bins. The centre of the Gaussian profile peak *ph1_x0* and its FWHM *ph1_FWHM* are recorded. In the case of a failure to fit a dummy value is recorded and ignored in further analysis.

The third panel from the top of Fig. C2 shows a successful fit for the *ph1_x0* and *ph1_FWHM* parameters.

[13] Related to *AG_FWHM* by $AG\_FWHM = 2\sqrt{2ln2}\sigma$.

*C4.3 'Resolved domain' visibility histogram centre and FWHM (*ph2_x0 *and* ph2_FWHM*)*

For visibilities in the 'Resolved domain' a Gaussian profile is fit to a histogram of the recorded phase values, in 50 bins. The centre of the Gaussian profile peak *ph2_x0* and its FWHM *ph2_FWHM* are recorded. In the case of a failure to fit a dummy value is recorded and ignored in further analysis.

The bottom panel of Fig. C1 shows a successful fit for the *ph2_x0* and *ph2_FWHM* parameters.

C5 Defining empirical relations of SNR and recorded simulated source properties

The *x*0 and *FWHM* values recovered from simulations were used to generate empirical bounding relations for these properties as a function of SNR. To do this, the average and standard deviation of both *x*0 and *FWHM*, for each model type [Point sources and Gaussian (including pure Gaussian and Gaussian + Point)], were calculated. Upper and lower limit values were then set as the average value ± 3×-the standard deviation at each SNR (across each model type). For *x*0 values an inverse relation, and for *FWHM* a power law relation between the data points and SNR were found to provide the best fits to the resultant profiles. For the lower *FWHM* boundary the value at the maximum SNR in the simulated model suite was used as a fixed limit across all FWHMs as any value between this and the upper bound at any given SNR provides a realistic FWHM value.

C5.1 Point model parameter boundaries

$$ph2\_x0_{\rm upper} = \frac{319.21}{\rm SNR} + -0.17, \tag{C2}$$

$$ph2\_x0_{\rm lower} = \frac{-301.91}{\rm SNR} + 0.06, \tag{C3}$$

$$ph2\_FWHM_{\rm upper} = 5067.0{\rm SNR}^{-0.96}, \tag{C4}$$

$$ph2\_FWHM_{\rm lower} = 4169.8{\rm SNR}^{-0.98}. \tag{C5}$$

C5.2 Gaussian model parameter boundaries

$$ph1\_x0_{\rm upper} = \frac{113.23}{\rm SNR} + 3.8, \tag{C6}$$

$$ph1\_x0_{\rm lower} = \frac{-100.12}{\rm SNR} + -4.1, \tag{C7}$$

$$ph1\_FWHM_{\rm upper} = 534.6{\rm SNR}^{-0.34}, \tag{C8}$$

$$ph1\_FWHM_{\rm lower} = -572.3{\rm SNR}^{-0.75}, \tag{C9}$$

$$ph2\_x0_{\rm upper} = \frac{36.83}{\rm SNR} + 20.04, \tag{C10}$$

$$ph2\_x0_{\rm lower} = \frac{-52.24}{\rm SNR} + -18.8, \tag{C11}$$

$$ph2\_FWHM_{\rm upper} = 722.9{\rm SNR}^{-0.3}, \tag{C12}$$

$$ph2\_FWHM_{\rm lower} = -396.5{\rm SNR}^{-0.27}, \tag{C13}$$

C6 Application to the TEMPO data

With the simulated model boundaries in place a comparison to the real TEMPO data is then possible. First, the visibility data for each TEMPO field was prepared to mitigate the effects of multiple sources in the same field.

C6.1 Preparing the data

To extract the visibility data for a specific TEMPO source in its host field the following steps were taken:

(i) Using the position, major and minor axis, position angle, and measured flux density for all sources in the current TEMPO field, *excluding* the source under investigation, a CASA component list of Gaussian sources was generated.

(ii) This component list was subtracted from the visibility data using the CASA task `uvsub`. This removes, or minimizes, the effect of multiple sources in a given field adding extra 'noise' to the expected source properties.

(iii) The phase centre of the field visibilities is shifted to the source position and the amplitude and phase data extracted by the same method as used for the simulated models.

In theory, the properties of the amplitudes and phases should then match those of the models in the case that the target has centrally condensed properties. In practice, there are some additional, unavoidable issues which much be considered. These are discussed in Section C7.

C6.2 Assigning a star forming classification

The same method used to measure the simulated model properties was then applied to the real TEMPO sample giving, for each source, values (or null results) for the parameters *AG_FWHM*, *ph1_FWHM*, *ph1_x0*, *ph2_x0*, and *ph2_FWHM*. With these values and the data boundaries from the empirical relations given in Section C5, each TEMPO source was assessed at the recorded SNR in the combined continuum images following a series of descision steps, used to assign if a source was a Point, Gaussian, Gaussian+Point, or None of the Above type source. For Point, Gaussian, and Gaussian+Point types these are considered ASC. Fig. C4, gives the decision tree used to determine if a TEMPO source is considered actively star forming or not.

Given the spread in the empirically derived bounds used a lower SNR cut off was used so that only sources with SNR $\geq$ 30 were considered int the classification analysis.

C7 Limitations to this analysis

Two limitations to this method when applied to real data exist. The first concerns the subtraction of field sources from the visibilities. The source properties (major and minor axes and position angle) used to generate the component list which is subtracted from the visibility data are based on those reported by the dendrogram analysis as discussed in the main text. In cases that assuming a Gaussian

profile is a poor fit to the true source shape, for example if the source is structured or extended then subtracting a Gaussian will lead to residual emission structure in the remaining visibilities. When imaged such residual emission structure could exhibit features like negative holes with positive emission halos or arcs around them. Similarly, subtracting a Gaussian profile for a source which is a combination of unresolved components in the TEMPO data will leave residual structure in the visibility data.

Both effects will limit our ability to assign a star-forming status to some sources within the TEMPO sample. This effect is hard to mitigate, as correctly modelling the emission properties of discreet >200 sources in complex fields containing extended structures is both time and computationally expensive and beyond the scope of the work conducted here.

The second artefact which can present itself in this method is setting an incorrect position when shifting the phase centre of the visibilities. For a point source small positional offsets from the true source position results in the phase data showing an 'arrow' or '<'-like profile. This indicates a delay-like behaviour caused by the offset between the phase centre and the true source position. The slope of the < can be used to indicate how large this offset is, as the *uv*-distance in which the phase slope would take to trace a full 360 degrees gives you a baseline length. Convert that baseline length to an angular scale (λ/b) gives the magnitude of the offset. Unfortunately, to probe all positions at this magnitude offset is again beyond the scope of this analysis.

[1]*Jodrell Bank Centre for Astrophysics, Department of Physics and Astronomy, School Of Natural Science, The University of Manchester, Manchester M13 9PL, UK*
[2]*UK ALMA Regional Centre Node, M13 9PL, Manchester, UK*
[3]*SKA Observatory, Jodrell Bank, Lower Withington, Macclesfield SK11 9FT, UK*
[4]*Instituto de Astrofísica de Andalucia (CSIC), Glorieta de al Astronomia s/n, E-18008 Granada, Spain*
[5]*I. Physikalisches Institut, University of Cologne, Zülpicher Str. 77, D-50937 Köln, Germany*
[6]*Ghana Space Science and Technology Institute, Radio Astronomy and Astrophysics Centre, Atomic-Haatso Road, Kwabenya, Accra, Greater Accra, Ghana*
[7]*School of Physics and Astronomy, University of Leeds, Leeds LS2 9JT, UK*
[8]*School of Physics and Astronomy, Cardiff University, Queens Buildings, The Parade, Cardiff CF24 3AA, UK*
[9]*IAPS-INAF, Via Fosso del Cavaliere, 100, I-00133 Rome, Italy*
[10]*SRON Netherlands Institute for Space Research, Landleven 12, 9747 AD Groningen, Netherlands*
[11]*Kapteyn Astronomical Institute, University of Groningen, Groningen 9747 AD, Netherlands*
[12]*Max-Planck-Institut für extraterrestrische Physik, Giessenbachstrasse 1, D-85748 Garching, Germany*
[13]*INAF-Osservatorio Astrofisico di Arcetri, Largo E. Fermi 5, I-50125 Firenze, Italy*
[14]*Max-Planck-Institut für Radioastronomie, Auf dem Hügel 69, D-53121 Bonn, Germany*
[15]*Centre for Astrophysics Research, Science and Technology Research Institute, University of Hertfordshire, College Lane, Hatfield, Hertfordshire AL10 9AB, UK*
[16]*Konkoly Observatory, Research Centre for Astronomy and Earth Sciences, Eötvös Loránd Research Network (ELKH), Konkoly Thege Miklós út 15-17, 1121 Budapest, Hungary*
[17]*CSFK, MTA Centre of Excellence, Konkoly Thege Miklós út 15-17, 1121 Budapest, Hungary*
[18]*Western Sydney University, Kingswood, NSW 2747, Australia*
[19]*Leiden Observatory, Leiden University, PO Box 9513, NL-2300 RA, Leiden, the Netherlands*
[20]*Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK*

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