



**HAL**  
open science

## ALMA resolves giant molecular clouds in a tidal dwarf galaxy

M. Querejeta, F. Lelli, E. Schinnerer, D. Colombo, U. Lisenfeld, C. Mundell, F. Bigiel, S. García-Burillo, C. Herrera, A. Hughes, et al.

► **To cite this version:**

M. Querejeta, F. Lelli, E. Schinnerer, D. Colombo, U. Lisenfeld, et al.. ALMA resolves giant molecular clouds in a tidal dwarf galaxy. *Astronomy and Astrophysics - A&A*, 2021, 645, pp.A97. 10.1051/0004-6361/202038955 . hal-03519527

**HAL Id: hal-03519527**

**<https://hal.science/hal-03519527>**

Submitted on 10 Jan 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# ALMA resolves giant molecular clouds in a tidal dwarf galaxy<sup>★,★★</sup>

M. Querejeta<sup>1</sup>, F. Lelli<sup>2</sup>, E. Schinnerer<sup>3</sup>, D. Colombo<sup>4</sup>, U. Lisenfeld<sup>5,6</sup>, C. G. Mundell<sup>7</sup>, F. Bigiel<sup>8</sup>, S. García-Burillo<sup>1</sup>, C. N. Herrera<sup>9</sup>, A. Hughes<sup>10,11</sup>, J. M. D. Kruijssen<sup>12</sup>, S. E. Meidt<sup>13</sup>, T. J. T. Moore<sup>14</sup>, J. Pety<sup>9,15</sup>, and A. J. Rigby<sup>2</sup>

<sup>1</sup> Observatorio Astronómico Nacional, Alfonso XII, 3, 28014 Madrid, Spain  
e-mail: m. querejeta@oan.es

<sup>2</sup> School of Physics and Astronomy, Cardiff University, 5 The Parade, CF24 3AA Cardiff, UK

<sup>3</sup> Max-Planck-Institut für Astronomie, Königstuhl 17, 69117 Heidelberg, Germany

<sup>4</sup> Max-Planck-Institut für Radioastronomie, Auf dem Hügel 9, 53121 Bonn, Germany

<sup>5</sup> Departamento de Física Teórica y del Cosmos, Universidad de Granada, 18071 Granada, Spain

<sup>6</sup> Instituto Carlos I de Física Teórica y Computacional, Facultad de Ciencias, 18071 Granada, Spain

<sup>7</sup> Department of Physics, University of Bath, Claverton Down, Bath BA2 7AY, UK

<sup>8</sup> Argelander-Institut für Astronomie, Universität Bonn, Auf dem Hügel 71, 53121 Bonn, Germany

<sup>9</sup> IRAM, 300 rue de la Piscine, 38406 Saint Martin, d'Hères, France

<sup>10</sup> Université de Toulouse, UPS-OMP, 31028 Toulouse, France

<sup>11</sup> CNRS, IRAP, Av. du Colonel Roche BP 44346, 31028 Toulouse Cedex 4, France

<sup>12</sup> Astronomisches Rechen-Institut, Zentrum für Astronomie der Universität Heidelberg, Mönchhofstraße 12-14, 69120 Heidelberg, Germany

<sup>13</sup> Sterrenkundig Observatorium, Universiteit Gent, Krijgslaan 281 S9, 9000 Gent, Belgium

<sup>14</sup> Astrophysics Research Institute, Liverpool John Moores University, IC2, Liverpool Science Park, 146 Brownlow Hill, Liverpool L3 5RF, UK

<sup>15</sup> Sorbonne Université, Observatoire de Paris, Université PSL, CNRS, LERMA, 75005 Paris, France

Received 16 July 2020 / Accepted 30 October 2020

## ABSTRACT

Tidal dwarf galaxies (TDGs) are gravitationally bound condensations of gas and stars that formed during galaxy interactions. Here we present multi-configuration ALMA observations of J1023+1952, a TDG in the interacting system Arp 94, where we resolved CO(2–1) emission down to giant molecular clouds (GMCs) at  $0.64'' \sim 45$  pc resolution. We find a remarkably high fraction of extended molecular emission ( $\sim 80$ – $90\%$ ), which is filtered out by the interferometer and likely traces diffuse gas. We detect 111 GMCs that give a similar mass spectrum as those in the Milky Way and other nearby galaxies (a truncated power law with a slope of  $-1.76 \pm 0.13$ ). We also study Larson's laws over the available dynamic range of GMC properties ( $\sim 2$  dex in mass and  $\sim 1$  dex in size): GMCs follow the size-mass relation of the Milky Way, but their velocity dispersion is higher such that the size-linewidth and virial relations appear super-linear, deviating from the canonical values. The global molecular-to-atomic gas ratio is very high ( $\sim 1$ ) while the CO(2–1)/CO(1–0) ratio is quite low ( $\sim 0.5$ ), and both quantities vary from north to south. Star formation predominantly takes place in the south of the TDG, where we observe projected offsets between GMCs and young stellar clusters ranging from  $\sim 50$  pc to  $\sim 200$  pc; the largest offsets correspond to the oldest knots, as seen in other galaxies. In the quiescent north, we find more molecular clouds and a higher molecular-to-atomic gas ratio ( $\sim 1.5$ ); atomic and diffuse molecular gas also have a higher velocity dispersion there. Overall, the organisation of the molecular interstellar medium in this TDG is quite different from other types of galaxies on large scales, but the properties of GMCs seem fairly similar, pointing to near universality of the star-formation process on small scales.

**Key words.** galaxies: dwarf – galaxies: interactions – galaxies: star formation – galaxies: ISM – galaxies: kinematics and dynamics

## 1. Introduction

When galaxies collide, a portion of the gas is ejected by tidal forces and may eventually collapse under self-gravity, giving rise to new low-mass galaxies along the tidal debris. These newborn systems, known as tidal dwarf galaxies (TDGs), are thought to be devoid of dark matter, but they retain the metallicity of the galaxy from which they were originally stripped (Duc et al. 2000; Duc 2012; Lelli et al. 2015). TDGs tend to have high gas fractions and they are often active star-formers (Braine et al. 2001; Lisenfeld et al. 2016); many of them also host old stel-

lar populations, but this is not always the case (Fensch et al. 2016). Such a dark matter-free setting, which is subject to tidal forces from their neighbours, affords the possibility of testing theories of star formation in an extreme dynamical environment. While TDGs have been extensively mapped through deep optical imaging, only a few studies have looked at their molecular gas content and distribution, and these studies target spatial resolutions that are too coarse to resolve individual giant molecular clouds (GMCs; e.g. Braine et al. 2000, 2001; Duc et al. 2007; Lisenfeld et al. 2016).

One of the most fundamental challenges of modern astrophysics is understanding the process by which gas transforms into stars, and how this process is orchestrated as a function of the environment. Indeed, theories of galactic-scale star formation and cosmological simulations of galaxy formation are

\* Full Table A.3 is only available at the CDS via anonymous ftp to [cdsarc.u-strasbg.fr](https://cdsarc.u-strasbg.fr) (130.79.128.5) or via <http://cdsarc.u-strasbg.fr/viz-bin/cat/J/A+A/645/A97>

\*\* Movie is available at <https://www.aanda.org>

both guided by the Kennicutt-Schmidt relation (Schmidt 1959; Kennicutt 1998), the observed correlation between the star formation rate surface density  $\Sigma_{\text{SFR}}$  and neutral gas surface density  $\Sigma_{\text{gas}}$  (e.g. Schaye & Dalla Vecchia 2008; Semenov et al. 2018). The correlation between star formation rates and the amount of molecular gas is also known to be very tight (e.g. Bigiel et al. 2008; Schruba et al. 2011; Leroy et al. 2013).

An increasing body of observational work suggests that, when molecular gas is probed down to GMC scales, its properties depend on local physical conditions (Hughes et al. 2013b; Colombo et al. 2014; Pan & Kuno 2017). Dynamical effects have been proposed to explain the differences in the state of molecular gas and its ability to form stars (e.g. Luna et al. 2006; Meidt et al. 2018, 2013; Kruijssen et al. 2014; Renaud et al. 2015). Cloud-cloud collisions have also been suggested as a trigger of star formation (e.g. Tan 2000; Fujimoto et al. 2014; Dobbs et al. 2015; Torii et al. 2015). Several other factors, such as changes in the incident radiation field, might further contribute to the observed differences in the shapes of GMC mass spectra between M51, M33, and the Large Magellanic Cloud (Hughes et al. 2013a), or NGC 6946, NGC 628, and M101 (Rebolledo et al. 2015). These environmental factors influence the ability of a GMC to form stars and, indirectly, the shape of the Kennicutt-Schmidt relation, but their relative importance remains unknown. To understand the physics of star formation, it is crucial to probe conditions down to GMC scales, the units that are associated with massive star formation in our own Galaxy.

J1023+1952 is a kinematically detached HI cloud in front of the spiral galaxy NGC 3227, in the interacting system Arp 94 (Mundell et al. 1995b). Figure 1 shows our target in context, highlighting the large size of the Arp 94 system and its complexity in terms of tidal tails as revealed by HI and deep optical emission. Specifically, the Arp 94 system is composed of a spiral galaxy hosting a Seyfert 2 nucleus (NGC 3227; Mundell et al. 1995a; Schinnerer et al. 2001; Noda et al. 2014) interacting with a ‘green valley’ elliptical (NGC 3226; Appleton et al. 2014). The TDG lies at the base of the northern tidal tail and partially overlaps in projection with the disc of NGC 3227. Since the TDG is obscuring stellar light from NGC 3227, it must be in front of the spiral galaxy (Mundell et al. 1995b, 2004). The total content of atomic gas estimated by Mundell et al. (2004) in the TDG is  $M_{\text{HI}} = 1.9 \times 10^8 M_{\odot}$  (corrected for our assumed distance), distributed over a projected area of  $7 \text{ kpc} \times 5 \text{ kpc}$ . As shown in Fig. 1, active star formation is ongoing in the southern half of the cloud, with blue knots clearly visible in  $\text{H}\alpha$  emission at velocities that closely match those of the HI cloud, ruling out the possibility that the star forming knots belong to the background galaxy, which is offset in velocity by almost  $300 \text{ km s}^{-1}$  (Mundell et al. 2004). Furthermore, the metallicity of these star-forming knots is near-solar ( $12 + \log(\text{O}/\text{H}) = 8.6$ ; Lisenfeld et al. 2008), comparable to the metallicity of the spiral galaxy NGC 3227, and much higher than the metallicity that would be expected for a dwarf galaxy of this luminosity ( $12 + \log(\text{O}/\text{H}) \approx 8$ , according to the magnitude-metallicity relation for classical dwarfs; Richer & McCall 1995; Zhao et al. 2010). This strongly supports the idea that J1023+1952 is a TDG and not a pre-existing dwarf captured by the Arp 94 system.

Using the IRAM 30 m telescope, Lisenfeld et al. (2008) demonstrated that this TDG is also rich in molecular gas ( $M_{\text{H}_2} = 1.5 \times 10^8 M_{\odot}$ , corrected for our distance). In spite of comparable atomic and molecular gas surface densities in the north and south of the cloud, star formation in J1023+1952 is predominantly happening along a 2-kpc-long ridge in the southern half

(traced by optical blue colours,  $\text{H}\alpha$ , and UV emission, as well as young stellar clusters resolved with HST; see Fig. 1).

There might be some evidence for low-level star formation in the north-west of the TDG (see Sect. 4.4), but this is very limited compared to the vigorous star formation activity in the south.

Thus, we distinguish two environments within the TDG: the quiescent north and the star-forming south, separated by a declination cut at  $\text{Dec} = 19^{\circ}51'50''$  (Fig. 1). This cut is motivated by the two main concentrations of HI, which are connected by a thin bridge on the eastern end. It might also be possible that these two gas condensations correspond to two close-by but distinct TDGs, but the existing HI data does not allow us to test this scenario by studying their internal dynamics. The absence of detectable near-infrared continuum suggests a scarcity of old stars in the TDG ( $M_{\star}^{\text{old}} < 4.7 \times 10^8 M_{\odot}$  corrected for our distance; Mundell et al. 2004), implying that this must be a young TDG which might not be in dynamical equilibrium.

TDGs constitute exciting laboratories to study how gas assembles within a newborn galaxy. In particular, it is yet not known how quickly gas condenses into compact structures in such a pristine system and whether those structures are similar to the GMCs observed in our Galaxy or in external galaxies. To answer these questions, we present ALMA observations of J1023+1952 which resolve the molecular emission down to GMC scales.

In Sect. 2 we describe the ALMA data (Sect. 2.1) and a rich set of ancillary data (Sect. 2.2). In Sect. 3, we identify and characterise the GMCs. In Sect. 4 we present our results regarding the overall molecular structure (Sect. 4.1), the molecular gas kinematics (Sect. 4.2), and the GMC properties and scaling relations (Sect. 4.3). We discuss our results in Sect. 5 with a special focus on the connection between GMC properties, local environment, and star formation. We close the paper with a summary in Sect. 6.

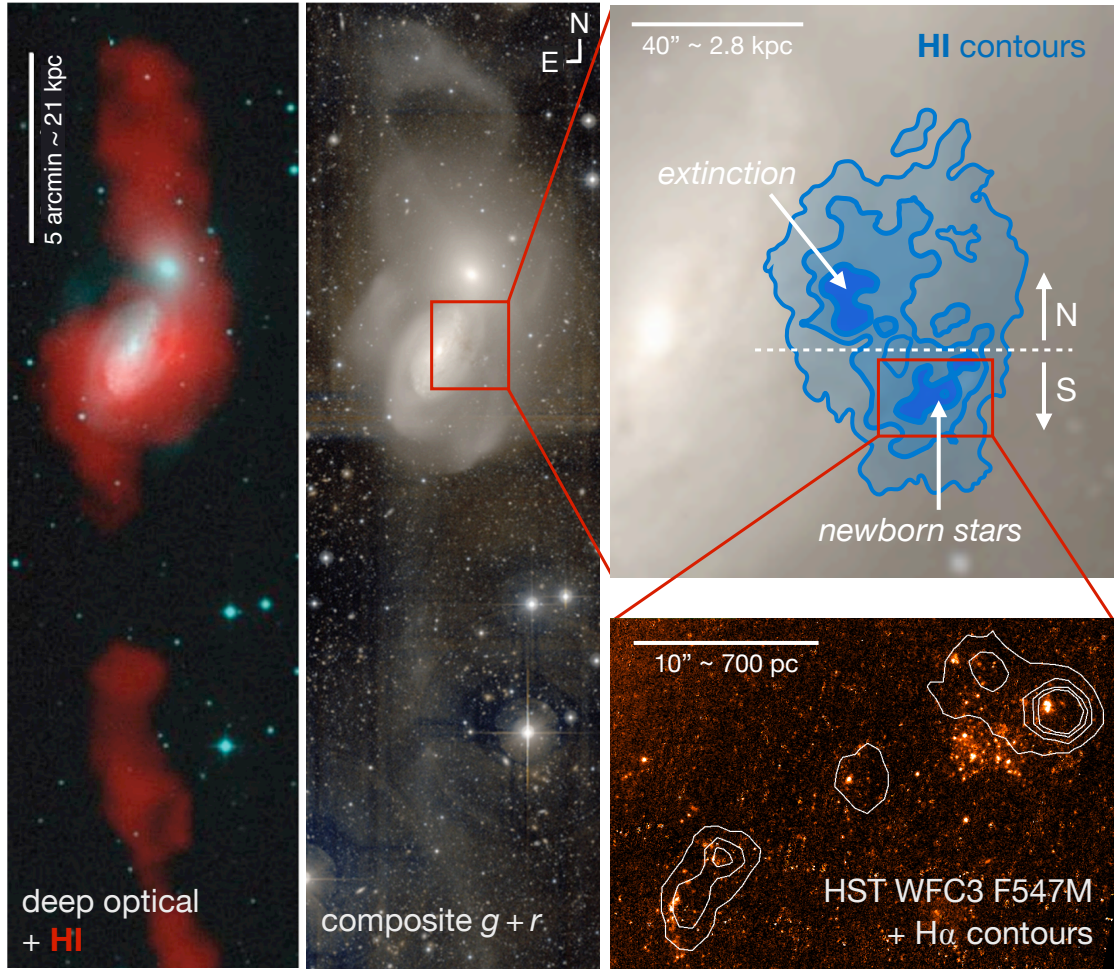
We assume a distance of  $14.5 \pm 0.6 \text{ Mpc}$  (Yoshii et al. 2014), based on a new AGN time lag method.

As a consistency check, Yoshii et al. (2014) derived an independent value of  $H_0 = (73 \pm 3) \text{ km s}^{-1} \text{ Mpc}^{-1}$  from these distances, in agreement with the literature (e.g. Schombert et al. 2020). For  $d = 14.5 \text{ Mpc}$ ,  $1''$  corresponds to  $70.3 \text{ pc}$ . This is very similar to the  $d = 15.1 \text{ Mpc}$  assumed by Mundell et al. (1995b) based on the redshift of the source, and slightly smaller than  $d = 20.4 \text{ Mpc}$  assumed by Mundell et al. (2004) and Lisenfeld et al. (2008) relying on the Tully-Fisher relation (Tully 1988).

## 2. Data

### 2.1. ALMA data: Molecular gas emission

We observed the CO(2–1) line with the ALMA Band 6 receiver (230 GHz) during Cycle 4 (project 2016.1.00648.S, PI: M. Querejeta). To recover emission from different spatial scales, we combined the 12 m array with the Atacama Compact Array (ACA, i.e. 7 m array and total power antennas). The 12 m array observations were performed between 25 and 28 November 2016, with 46 antennas in C40-4 configuration, resulting in projected baselines between 15 and 704 m which correspond to a typical angular resolution of  $0.4\text{--}0.5''$  and a largest recoverable angular scale of  $\approx 4''$ . The 7 m observations were carried out between 4 and 26 October 2016 (with 8–10 antennas), and the total power, between 3 and 15 October 2016 (with 3–4 antennas). The 7 m-array data at the CO(2–1) frequency provides a resolution of  $\approx 5.4''$  and a largest recoverable angular scale of  $\approx 29''$ , while ALMA total power has a resolution of  $\approx 28''$  and is by definition sensitive to emission from all spatial scales.



**Fig. 1.** *Left panel:* H I emission in the Arp 94 system (in red) on top of a deep optical image from [Mundell et al. \(1995b\)](#). *Central panel:* composite  $g + r$  deep optical image from MegaCam/CFHT using an “arcsinh stretch” to emphasise faint emission ([Duc et al. 2014](#)). Both the left and central panels come from [Appleton et al. \(2014\)](#). *Top-right panel:* blow-up of the  $g + r$  composite image showing the location of the TDG in H I emission as blue contours, equivalent to an extinction of 0.4, 1.4, 2.4, and  $3.4 A_B$  magnitudes, ranging from  $4 \times 10^{20}$  to  $4 \times 10^{21} \text{ cm}^{-2}$  ([Mundell et al. 2004](#)); the annotations indicate the area where a high optical extinction is seen towards the background galaxy (north-east) and the star-forming region in the south, as well as the demarcation line at  $\text{Dec} = 19^\circ 51' 50''$  that we apply to distinguish the northern from the southern part of the TDG. *Bottom-right panel:* star-forming region in the south of the TDG as revealed by young clusters visible in HST WFC3/F547M imaging and H $\alpha$  emission from [Mundell et al. \(2004\)](#) shown as contours ( $0.8, 2, 3,$  and  $4 \times 10^{33} \text{ erg s}^{-1} \text{ pc}^{-2}$ ).

The full extent of the TDG ( $\sim 64'' \times 43''$ ) was covered with an 18-pointing mosaic by the 12 m array (primary beam of full width at half maximum  $FWHM = 25.3''$ ); a 7-pointing mosaic was sufficient to cover the same field of view with the 7 m array, given the larger primary beam ( $FWHM = 43.5''$ ). In both cases, the mosaic was hexagonally packed, with approximate Nyquist sampling along rows ( $0.51 \times$  the primary beam) and a spacing  $\sqrt{3}/2$  times the Nyquist separation between rows.

We calibrated and imaged the datasets using CASA 5.4.1 (Common Astronomy Software Applications<sup>1</sup>). After standard calibration, we concatenated the visibilities corresponding to the 7 m and 12 m interferometric data and imaged them together. We used the total power observations as model within `tclean` in CASA, so that the final cube recovers all the flux. The total power data were calibrated and imaged following the strategy presented in [Herrera et al. \(2020\)](#), Appendix A. For the baseline correction, we fitted a polynomial of order 1 in a fixed velocity window ( $755\text{--}855 \text{ km s}^{-1}$  and  $1500\text{--}1600 \text{ km s}^{-1}$ ). We matched the spatial grid and spectral resolution ( $2.5 \text{ km s}^{-1}$ ) of the inter-

ferometric data. We also performed an alternative imaging of the interferometric data, combining the visibilities from the 12 m and 7 m arrays (without total power). In all cases, we cleaned using the Hogbom algorithm with natural weighting, down to a threshold of 2 mJy per beam ( $\sim 2\sigma$ ) in  $2.5 \text{ km s}^{-1}$  channels, and within the area where the primary beam response remains higher than 20% of the maximum (which roughly corresponds to a tilted rectangle of  $\sim 80'' \times 70'' \approx 5.6 \times 4.9 \text{ kpc}$ ; shown as a dashed blue line in the left panel of Fig. 3). We chose a pixel size of  $0.082''$  and an image size of  $1344 \times 1440$  pixels (centred on  $\text{RA} = 10:23:26.249$ ,  $\text{Dec} = +19:51:58.83$ ). With this imaging strategy, we obtained an average synthesised beam of  $0.69'' \times 0.60''$  ( $\text{PA} = 11^\circ$ ), which corresponds to  $\approx 45 \text{ pc}$  for our assumed distance. The rms brightness sensitivity of the resulting data cube is  $\sigma_{\text{rms}} \sim 1 \text{ mJy per beam}$  ( $\approx 56 \text{ mK}$ ), which for a representative linewidth of  $FWHM = 15 \text{ km s}^{-1}$  yields a molecular gas surface density sensitivity of  $\sim 3.9 M_\odot \text{ pc}^{-2}$  (or a  $1\sigma$  point-source sensitivity of  $\sim 9 \times 10^3 M_\odot$ ) for our assumed  $\alpha_{\text{CO}}$  conversion factor (Sect. 2.4). In order to track flux recovery and compare against ancillary data, we also produced a version of the final cube convolved to lower resolution (with a circular

<sup>1</sup> <http://casa.nrao.edu>

Gaussian kernel of  $FWHM = 1.5'', 3'',$  and  $6.3''$ , the latter to match the resolution of the HI data). We note that we did not detect any continuum emission. We applied a primary beam correction to the final cubes before performing any measurements. All velocities in this paper are heliocentric, expressed following the radio convention ( $v = c(v_0 - \nu)/\nu_0$ ).

## 2.2. Ancillary data

### 2.2.1. VLA HI data

To trace atomic gas, we use HI observations from the Very Large Array (VLA), originally published in [Mundell et al. \(2004\)](#). The data were taken by the VLA in B configuration, with a spatial resolution of  $6.3''$  and velocity resolution of  $10 \text{ km s}^{-1}$ . Previously, [Mundell et al. \(1995b\)](#) had used the VLA to map the same target at lower resolution ( $\sim 20''$ ) in C configuration. The amount of diffuse HI flux that is filtered out in B configuration is probably low, since the C array only recovered  $\sim 12\%$  more flux than the B array ([Mundell et al. 2004](#)).

### 2.2.2. IRAM 30 m data

We compare our CO(2–1) observations against previous CO(1–0) observations of the TDG with the IRAM 30 m telescope. Those single-dish observations were published by [Lisenfeld et al. \(2008\)](#), and they attained a spatial resolution of  $22''$ .

### 2.2.3. HST data

We make use of archival data from the *Hubble* Space Telescope (HST) taken with the Wide Field Camera 3 (WFC3). We obtained the data covering both NGC 3227 and the TDG from the HST archive (proposal 11661, PI: Misty Bentz), which follows on-the-fly calibration procedures. The map was obtained with the medium-band filter F547M of the WFC3 camera and has been presented in [Batiste et al. \(2017\)](#). We corrected the astrometry of the HST image using *Gaia* field stars as reference (a total of 15 stars down to 21 mag in *g*-band, *Gaia* DR2; [Gaia Collaboration 2016, 2018](#)). The offsets applied were  $\Delta RA = 1.2 \text{ px}$  ( $0.05''$ ) and  $\Delta Dec = -7.8 \text{ px}$  ( $-0.31''$ ).

### 2.2.4. H $\alpha$ data

We use H $\alpha$  observations of NGC 3227 from the 4.2 m *William Herschel* Telescope in La Palma, presented in [Mundell et al. \(2004\)](#). We corrected the astrometry of the H $\alpha$  image by applying an offset of  $\Delta RA = 0.33 \text{ px}$  ( $0.11''$ ) and  $\Delta Dec = 2.05 \text{ px}$  ( $0.67''$ ) to match *Gaia* field stars.

## 2.3. Moment maps of the TDG emission

As illustrated by Fig. 1, the TDG partially overlaps in projection with the background galaxy NGC 3227. To isolate the emission associated with the TDG from the emission arising from NGC 3227, we followed a data-driven approach. Firstly, we smoothed the HI cube to  $30''$  resolution in space and  $50 \text{ km s}^{-1}$  in velocity, boosting the signal-to-noise ratio. Then, we clipped pixels below  $3\sigma_{\text{sm}}$ , where  $\sigma_{\text{sm}} = 1 \text{ mJy beam}^{-1}$  is the rms noise per channel in the smoothed HI cube. In this clipped cube, the HI emission from the main galaxy and that from the TDG can be unambiguously identified because they cover distinct areas in

space and velocity. Thus, we constructed a Boolean mask for the main galaxy, isolating contiguous areas in the channel maps.

This mask was applied to the HI and CO cubes at full resolution to exclude emission from NGC 3227.

We obtained intensity maps for CO and HI emission by integrating the corresponding cubes, after applying the mask excluding the emission from NGC 3227. We integrated in the velocity range  $[1060, 1360] \text{ km s}^{-1}$ . We considered the velocity range  $[1360, 1450] \text{ km s}^{-1}$  as line-free and used those channels to estimate the rms noise on a pixel-by-pixel basis. When computing moment maps, we applied a dilated mask technique to minimise the impact of noise ([Rosolowsky & Leroy 2006](#)); we started from a threshold of  $4\sigma$  based on the rms map, and then dilated the mask to include any adjacent voxels above  $1.5\sigma$ . We used the same dilated mask technique to construct first- and second-order moment maps. We obtained uncertainty maps for the first- and second-order moment maps by formally propagating  $\text{rms}_{\text{channel}}$ .

## 2.4. Conversion to gas surface densities

To convert the ALMA CO(2–1) intensities into molecular gas surface densities, we applied a constant factor of  $\alpha_{\text{CO}}^{2-1} = 4.4 M_{\odot} (\text{K km s}^{-1} \text{ pc}^2)^{-1}$ , which is the average value measured by [Sandstrom et al. \(2013\)](#) directly on the CO(2–1) line for a set of resolved nearby galaxy discs (with a galaxy-to-galaxy scatter of 0.3 dex, implying fluctuations of a factor of  $\sim 2$ ). The near-solar metallicity of the TDG suggests that this is probably a reasonable assumption. The constant  $\alpha_{\text{CO}}^{2-1}$  has the advantage that our surface density maps remain proportional to the directly measured CO(2–1) intensity. This value is equivalent to  $2 \times 10^{20} \text{ cm}^{-2} (\text{K km s}^{-1})^{-1}$  when applied on the CO(2–1) line, including a factor of 1.36 to correct for the presence of helium.

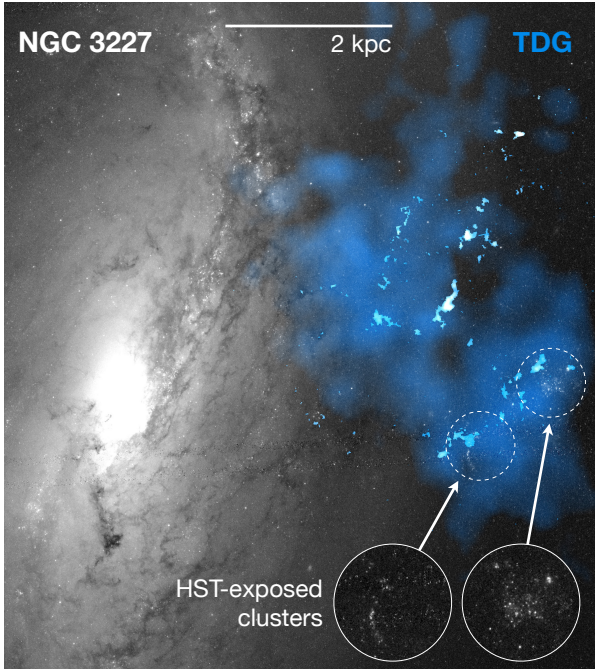
The standard Galactic value is  $2 \times 10^{20} \text{ cm}^{-2} (\text{K km s}^{-1})^{-1}$  when applied on the CO(1–0) line. Thus, considering the mean CO(2–1)/CO(1–0) line ratio of 0.5–0.6 of this TDG (see Sect. 4.1.3), the fiducial molecular gas surface densities that we derive in this paper are  $\sim 40\text{--}50\%$  lower than what would be implied by the standard Galactic conversion factor applied on CO(1–0). In principle, we could also assume a spatially varying  $\alpha_{\text{CO}}^{2-1}$ , assuming a constant  $\alpha_{\text{CO}}^{1-0}$  modulated by the observed CO(2–1)/CO(1–0) ratio map. This would result in molecular gas surface densities that are a factor  $\sim 3$  and  $\sim 1.5$  higher in the north and south, respectively. We prefer to avoid this modulation of  $\alpha_{\text{CO}}$  by  $R_{21}$  because of the coarse resolution ( $\sim 28''$ ) of the available  $R_{21}$  map, limited by the CO(1–0) single-dish data.

We transformed HI intensities into atomic gas surface densities using the standard formula for optically thin gas,  $\eta_{\text{H}} = 1.82 \times 10^{18} \times F_{\text{HI}}$ , where  $F_{\text{HI}}$  is in  $\text{K km s}^{-1}$  and  $\eta_{\text{H}}$  is in atoms per  $\text{cm}^2$  (e.g. [Condon & Ransom 2016](#)). This is equivalent to multiplying by 0.0145 to transform from  $\text{K km s}^{-1}$  to  $M_{\odot} \text{ pc}^{-2}$ .

## 3. GMC identification and characterisation

### 3.1. GMC segmentation with CPROPS

To perform the GMC identification with CPROPS ([Rosolowsky & Leroy 2006](#)), we started from a mask of significance defined as those voxels with signal above  $4\sigma$  expanded to any adjacent voxels with emission above  $1.5\sigma$ . The same mask of significance was used for SCIMES ([Colombo et al. 2015](#)), as described in Appendix A.2. Starting from the connected, discrete regions of signal within the mask (the so-called “islands”), local maxima were subsequently identified and kept only if the voxels that are closer to a given maximum than to any other maxima define



**Fig. 2.** False-colour image combining the CO(2–1) intensity from ALMA (white), optical image from HST (greyscale), and H I intensity (blue) from the VLA (Mundell et al. 2004). The circles show a blow-up where HST reveals young stellar clusters associated with the star-forming part of the TDG.

an area larger than the synthesised beam. If this area was more than one synthesised beam but less than two beams, the island was included in the catalogue but not decomposed. Additionally, the local maxima must be  $2\sigma_{\text{rms}}$  above the merge level with other clouds. For all pairs of local maxima within an island, if the computed moments based on the emission associated with each of the maxima separately differed less than 100% from the moments computed for the combined emission, then those were no longer considered as separate peaks. After this process, all the emission uniquely associated with a surviving local maximum was considered a segmentation unit (a GMC).

We considered emission as long as the primary beam response from ALMA was at least 20% of the maximum, which corresponds to an area of  $\sim 80'' \times 70''$ . We masked emission from the background galaxy NGC 3227 as described in Sect. 2.3.

### 3.2. Deriving GMC properties

We followed Rosolowsky & Leroy (2006) to extract physical properties out of the discrete units identified by CPROPS and SCIMES. The quantities extracted by the segmentation codes are the fluxes  $F$ , the sizes  $R$ , and the linewidths  $\Delta v$ .

Our ALMA cube can be characterised by a brightness temperature  $T_i$  for each volumetric pixel (voxel  $i$ ). Each voxel can also be identified by two spatial coordinates ( $x_i$ ,  $y_i$ ) and a velocity channel ( $v_i$ ); according to the segmentation, a given GMC covers a discrete range of pixels and velocity channels ( $\delta x$ ,  $\delta y$ ,  $\delta v$ ). The luminosity was obtained adding up all emission within the desired range of voxels:

$$L_{\text{CO}} = \sum_i T_i \delta x \delta y \delta v D^2, \quad (1)$$

where  $D$  is the distance to the galaxy. To obtain sizes, both CPROPS and SCIMES first rotate the  $x$  and  $y$  axes to be aligned

with the major and minor axes of the cloud, determined from principal component analysis. Subsequently, the size of the cloud was calculated as the intensity-weighted second moment ( $\sigma_{\text{maj}}$ ,  $\sigma_{\text{min}}$ ) along each spatial dimension:

$$\sigma_{\text{maj}} = \sqrt{\frac{\sum_i T_i (x_i - \bar{x})^2}{\sum_i T_i}}, \quad \sigma_{\text{min}} = \sqrt{\frac{\sum_i T_i (y_i - \bar{y})^2}{\sum_i T_i}}, \quad (2)$$

with  $\sigma_r = \sqrt{\sigma_{\text{maj}} \sigma_{\text{min}}}$ . The size of the cloud,  $R$ , was calculated as  $R = \eta \sigma_r$ , where  $\eta$  translates the spatial second moment to the radius of a spherical cloud, and it depends on the density distribution of the GMC. We adopted  $\eta = 1.91$  (Solomon et al. 1987), which is the most standard assumption in Galactic and extragalactic studies. When the cloud size provided by the segmentation codes was smaller than the synthesised beam, the GMC was considered unresolved and was excluded from the analysis.

Analogously, the velocity dispersion was obtained as:

$$\sigma_v = \sqrt{\frac{\sum_i T_i (v_i - \bar{v})^2}{\sum_i T_i}}, \quad (3)$$

with the FWHM linewidth given by  $\Delta v = \sqrt{8 \ln(2)} \sigma_v$ .

Starting from the basic properties above, we also derived a number of indirect properties, including luminous and virial masses. The luminous mass was obtained as  $M_{\text{lum}} = \alpha_{\text{CO}}^{2-1} L_{\text{CO}}$ , while the virial mass was estimated as  $M_{\text{vir}} = 1040 \sigma_v^2 R$  ( $M_{\text{vir}}$  in  $M_{\odot}$  if  $\sigma_v$  is in  $\text{km s}^{-1}$  and  $R$  in pc), which assumes clouds with density profile  $\rho \propto R^{-1}$ .

To account for the effects of finite sensitivity, which can yield slightly smaller GMC sizes than in reality, we used the extrapolation of the moments to the 0 K contour implemented in CPROPS and SCIMES (Rosolowsky & Leroy 2006). Additionally, the synthesised beam was deconvolved from the extrapolated spatial moments. To account for uncertainties, we ran CPROPS with the bootstrap option, performing 1000 iterations.

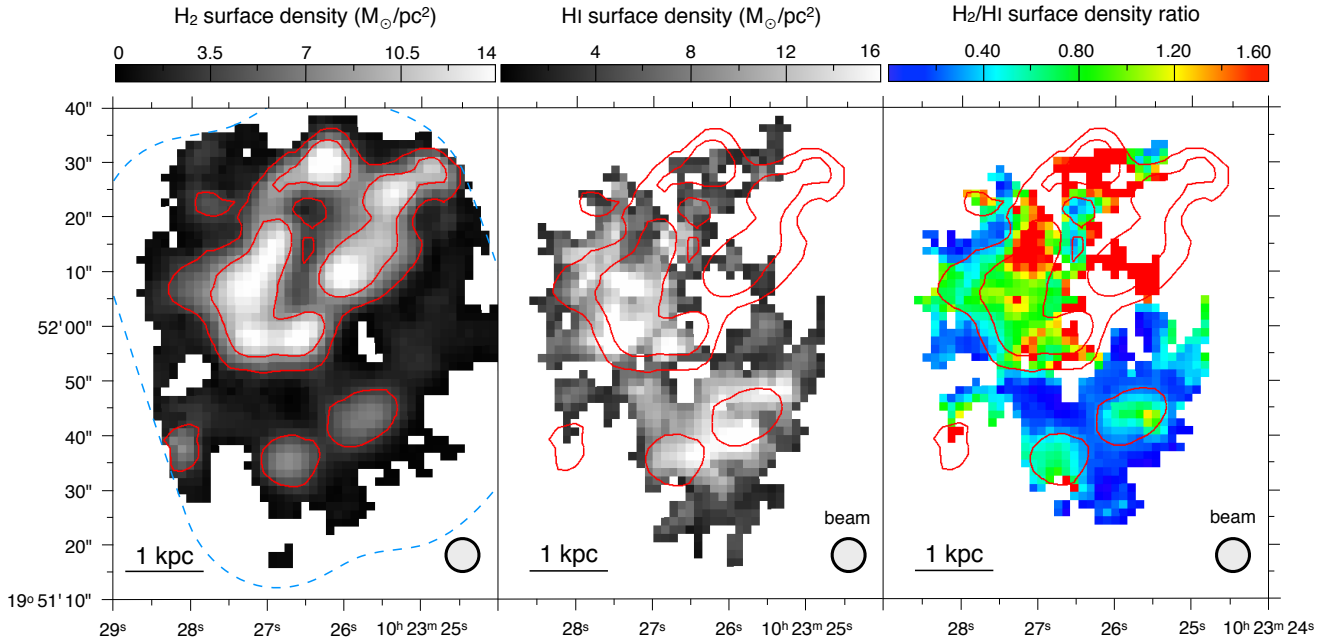
## 4. Results

Figure 2 shows the distribution of the molecular gas revealed by ALMA (white) on top of atomic gas (blue) and stellar continuum (greyscale). This composite-colour image already hints at some of the most important results of this paper. The H I observations at  $6.3''$  resolution show that atomic gas extends over a relatively large projected area ( $\sim 6 \text{ kpc} \times 4 \text{ kpc} = 24 \text{ kpc}^2$ ). The ALMA observations at high-resolution ( $0.64''$ ) reveal that the molecular emission that stands out on sub-arcsecond scales is highly clumpy and has a low covering factor. However, a very large fraction of the molecular gas is diffuse; we examine this in Sect. 4.1, together with the molecular-to-atomic gas ratio and CO(2–1)/CO(1–0) ratio. We study the GMC properties, mass spectra, and GMC scaling relations in Sect. 4.3. In Sect. 4.2 we focus on gas kinematics. The clusters visible in Fig. 2 (highlighted by the circular blow-ups) are often not aligned with the peaks of CO emission; we look at these offsets in Sect. 4.4.

### 4.1. Molecular structure of the TDG

#### 4.1.1. Molecular-to-atomic gas ratio

The left and central panels of Fig. 3 show the CO and H I intensity maps at matched  $6.3''$  resolution, overlaid with CO contours. There is plenty of CO emission where H I emission is very limited (below our detection threshold of  $4\sigma$ ), particularly towards



**Fig. 3.** *Left panel:* H<sub>2</sub> surface density from ALMA CO(2–1) observations (12 m + 7 m + TP) convolved to 6.3'' resolution, with contours highlighting levels of (2, 6, 10)  $M_{\odot} \text{pc}^{-2}$ . The dashed blue line is the ALMA field of view (where the primary beam response exceeds 20%). *Central panel:* HI surface density from the VLA at 6.3'' resolution (Mundell et al. 2004), with CO contours from the left panel in red. *Right panel:* the ratio of molecular-to-atomic gas surface densities derived from our CO(2–1) and HI observations (left and central panel) at 6.3'' resolution; CO contours are in the other two panels. In all maps north is up and east is left, with J2000.0 equatorial coordinates.

**Table 1.** Molecular and atomic gas properties at matched 6.3'' resolution.

	Whole TDG	North <sup>(a)</sup>	South <sup>(b)</sup>
Total H <sub>2</sub> mass ( $M_{\odot}$ )	$8.6 \times 10^7$	$7.3 \times 10^7$	$1.3 \times 10^7$
Total HI mass ( $M_{\odot}$ )	$8.4 \times 10^7$	$4.9 \times 10^7$	$3.5 \times 10^7$
Mean $\Sigma_{\text{H}_2}$ ( $M_{\odot} \text{pc}^{-2}$ )	4.8	5.7	2.5
Mean $\Sigma_{\text{HI}}$ ( $M_{\odot} \text{pc}^{-2}$ )	8.1	7.8	8.4
Total H <sub>2</sub> mass/total HI mass	1.02	1.47	0.37
Mean $\Sigma_{\text{H}_2}/\Sigma_{\text{HI}}$	0.59	0.73	0.30
Mean $R_{21} = \text{CO}(2-1)/\text{CO}(1-0)$	0.52	0.56	0.38

**Notes.** <sup>(a)</sup>North is where Dec > 19°51'50''. <sup>(b)</sup>South is where Dec < 19°51'50''. Measurements at matched spatial resolution (6.3'') applying a mask of significance to both cubes (4 $\sigma$  dilated mask); thus, the total H<sub>2</sub> mass quoted here is lower than implied by ALMA at native resolution without any masking ( $1.6 \times 10^8 M_{\odot}$ ). The mean  $\Sigma_{\text{H}_2}$  and  $\Sigma_{\text{HI}}$  are calculated for pixels with significant detections of H<sub>2</sub> and HI, respectively; this means that they are averaged over different areas.

the north-west of the TDG. The map also suggests that CO emission is somewhat clumpier than HI emission. Indeed, the standard deviation in the flux distribution in the CO map (0.45 on a logarithmic scale) is almost twice higher than HI (0.25), confirming that HI is more homogeneous even at matched resolution. These effects contribute to the spatial variation of the molecular-to-atomic gas ratio evident in the right panel of Fig. 3: there are important changes in the H<sub>2</sub>/HI ratio both locally (around CO peaks) and globally (between the north and south of the TDG).

Table 1 lists our measurements of the molecular-to-atomic gas ratio for the two environments in the TDG. Similar to other TDGs (Braine et al. 2001), this system is highly molecular with a total H<sub>2</sub>/HI mass ratio of  $\sim 1$ . However, the molecular ratio varies significantly across the TDG with local variations up to a fac-

tor of 2–3. In addition to local variations, globally the quiescent north has a substantially higher molecular-to-atomic gas ratio (1.47) than the star-forming south (0.37). This becomes extreme in the north-west corner, where the  $\Sigma_{\text{H}_2}/\Sigma_{\text{HI}}$  ratio is at least  $\sim 4$ –6 considering 4 $\sigma$  upper limits for HI. Averaging the local  $\Sigma_{\text{H}_2}$  to  $\Sigma_{\text{HI}}$  ratio for pixels with simultaneous CO and HI detections results in a lower ratio than dividing the total H<sub>2</sub> and HI masses. Independently from the indicator used, the molecular-to-atomic gas ratio is surprisingly higher in the quiescent north, whereas star formation is enhanced in the more atomic gas-dominated south. The average H<sub>2</sub> surface density is also higher in the north, while the average HI surface density is similar among both environments. The molecular-to-atomic ratio remains higher in the north even if we use a spatially variable  $\alpha_{\text{CO}}^{2-1}$  (Sect. 2.4), but the difference between north and south becomes smaller.

#### 4.1.2. Diffuse molecular gas

Table 2 provides the CO flux recovered by our observations using different ALMA arrays. Specifically, the 12 m array (configuration C40-4) and the compact 7 m array are sensitive to scales up to 4''  $\sim 280$  pc and 29''  $\sim 2$  kpc, respectively, while the combination of 12 m + 7 m + total power should capture emission from all scales. Given the very low covering factor of emission in the 12 m + 7 m cube, we applied a mask to ensure that we extract only meaningful flux (using a 4 $\sigma$  dilated mask as explained in Sect. 2.3). This mask encapsulates the compact structures that the 12 m + 7 m interferometer is sensitive to. We measured the flux inside the same mask both for the 12 m + 7 m and 12 m + 7 m + TP cubes: the 12 m + 7 m cube recovers almost the whole flux (94%) in these compact structures. However, if we compare this compact emission against the total CO flux in the 12 m + 7 m + TP cube, almost 90% of the total CO flux is outside the mask.

**Table 2.** Fluxes (in  $\text{K km s}^{-1} \text{ kpc}^2$ ) retrieved with or without short spacings and in GMCs at the native ALMA resolution ( $0.64''$ ).

Environment	Area ( $\text{kpc}^2$ )	Flux in 12 m + 7 m + TP		Flux in 12 m + 7 m Inside mask <sup>(d)</sup>	Flux in GMCs <sup>(e)</sup>
		Total	Inside mask <sup>(d)</sup>		
Whole TDG <sup>(a)</sup>	25.6	37.1	4.86 (13.1%)	4.58 (94.2%)	6.76 (18.2%)
North <sup>(b)</sup>	15.8	28.5	3.93 (13.8%)	3.70 (94.1%)	5.44 (19.1%)
South <sup>(c)</sup>	9.9	8.5	0.93 (10.9%)	0.88 (94.6%)	1.32 (15.5%)

**Notes.** <sup>(a)</sup>Whole TDG is the entire ALMA field-of-view, where the primary beam response exceeds 20% (dashed blue line in Fig. 3;  $\sim 80'' \times 70''$ ). <sup>(b)</sup>North is where  $\text{Dec} > 19^\circ 51' 50''$ . <sup>(c)</sup>South is where  $\text{Dec} < 19^\circ 51' 50''$ . <sup>(d)</sup>Flux inside  $4\sigma$  dilated mask (the mask is identical in both cases and is based on the 12 m + 7 m cube). The fourth and fifth columns indicate the percentage of the total 12 m + 7 m + TP flux that is inside the dilated mask and the percentage of 12 m + 7 m + TP flux recovered with 12 m + 7 m inside the same mask, respectively. <sup>(e)</sup>Total flux in all GMCs identified with CPROPS, after extrapolation to 0 K and deconvolution of the beam; without extrapolation and deconvolution, the total flux in GMCs is considerably lower,  $2.46 \text{ K km s}^{-1} \text{ kpc}^2$ . The percentage of the total 12 m + 7 m + TP flux is listed. The total flux in GMCs identified with SCIMES (Appendix A.2) is 8.18 and  $2.94 \text{ K km s}^{-1} \text{ kpc}^2$  for the extrapolated and non-extrapolated version, respectively.

Extended emission can also be quantified as the fraction of the total CO flux that is contained in the GMCs identified by CPROPS (see Sect. 3.1). Depending on whether we include the extrapolation to infinite sensitivity or not, the flux in GMCs is 18% or 7% of the total, respectively. This roughly agrees with the amount of flux in compact structures that is captured by the interferometer (12 m + 7 m data). Thus, most molecular emission is not arising from the type of compact structures that are responsible for massive star formation in galaxies. The extended emission is likely associated with diffuse molecular gas, and it accounts for as much as  $\sim 80\text{--}90\%$  of the molecular emission in the TDG. This is substantially higher than the values found in other nearby galaxies (typically 10–60%; Pety et al. 2013; Caldú-Primo et al. 2015), only comparable to the 74–91% extended emission found in the bar of NGC 1300 (Maeda et al. 2020b). There is no strong dependence of the diffuse fraction on environment: it is slightly higher in the north but quite similar in both cases (see Table 2).

Figure 4 shows the spatial distribution of compact CO emission at  $6.3''$  resolution (ALMA 12 m + 7 m data, without total power) against the diffuse component (the result of subtracting the intensity from ALMA 12 m + 7 m + TP minus the intensity from 12 m + 7 m only, channel by channel). We performed this analysis at  $6.3''$  resolution to avoid being dominated by local spikes due to noise, and because it is the resolution of the HI map. As expected, the diffuse component is smoother than the compact emission even at matched  $6.3''$  resolution. Yet, the diffuse emission is not perfectly homogeneous either, and seems to be more intense towards the north.

The diffuse component shows a north-south velocity gradient and, as one might expect, the velocity in diffuse CO varies more smoothly than in compact CO structures. The CO linewidth is high ( $\sigma \sim 50\text{--}60 \text{ km s}^{-1}$ ), and correlates with intensity. This is in stark contrast with the compact component, which shows large local variations in velocity and linewidth, as well as much more limited velocity dispersion (mostly below  $\sim 20\text{--}30 \text{ km s}^{-1}$ ). The kinematics of the compact component does not follow in detail the velocity field and dispersion map of the diffuse component.

#### 4.1.3. CO(2–1)/CO(1–0) ratio

We explore spatial variations in  $R_{21}$  by combining our ALMA total power CO(2–1) observations ( $28''$  resolution) with the IRAM 30 m CO(1–0) single-dish observations ( $22''$  resolution) from Lisenfeld et al. (2008). We first matched resolutions

by convolving the IRAM 30 m cube to  $28''$  resolution. Then, we computed the relevant moment-zero maps (Sect. 2.3) and took their ratio, CO(2–1)/CO(1–0), in flux units of  $\text{K km s}^{-1}$ . The average value (masking NGC 3227) is 0.52, in good agreement with the global  $R_{21} = 0.54 \pm 0.10$  found by Lisenfeld et al. (2008).

Figure 5 shows that there is a significant gradient in  $R_{21}$  along the TDG, from a maximum of  $\sim 0.7$  in the north down to a minimum of  $\sim 0.3$  in the south. This might indicate changes in the physical properties of molecular gas between the quiescent and star-forming part of the TDG. We also computed  $R_{21}$  for the part of the background galaxy NGC 3227 simultaneously covered by the ALMA and IRAM 30 m observations. In NGC 3227, we find an average value of 0.53, similar to the global ratio for the TDG. In Sect. 5.2 we discuss how the mean  $R_{21}$  ratio in the TDG is compatible with the lowest values observed in nearby galaxies.

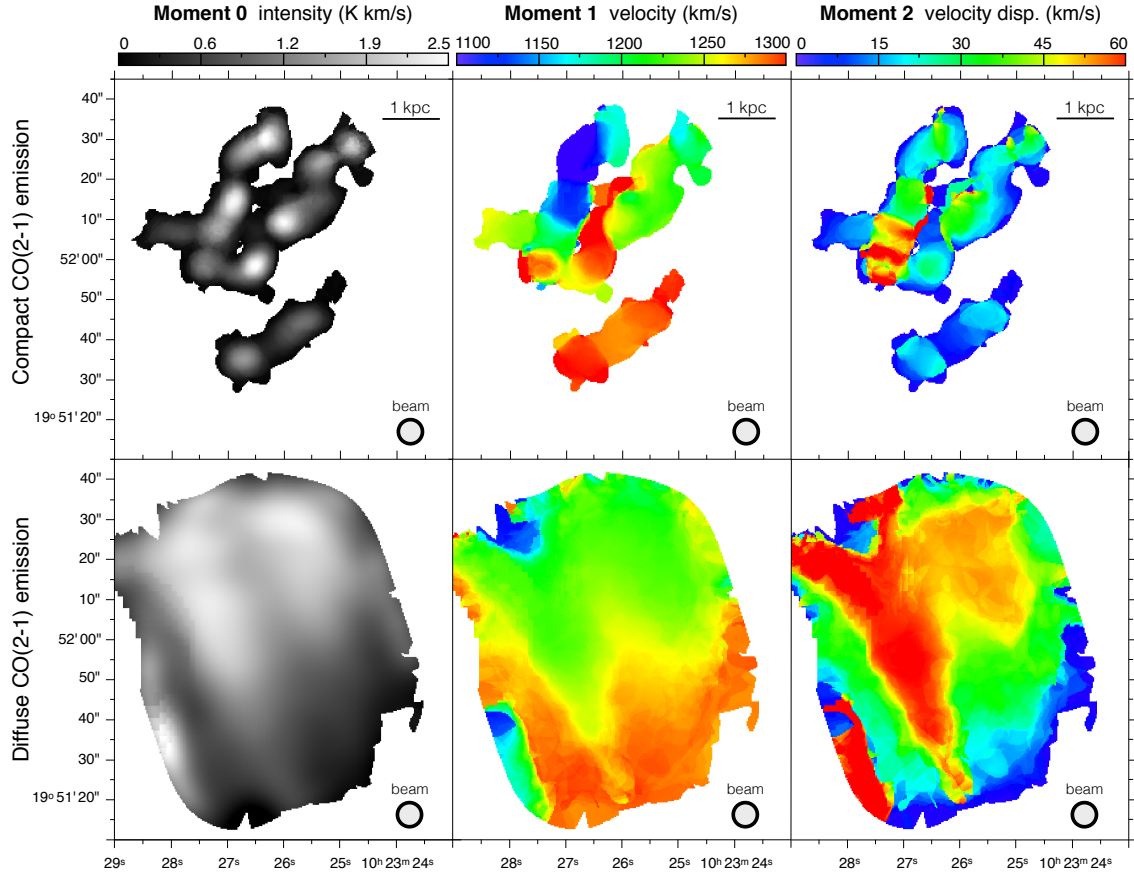
#### 4.2. Kinematics

The HI emission shows a clear velocity gradient along the north-south direction which was interpreted by Mundell et al. (2004) as possibly due to rotation. The left panel of Fig. 6 shows the difference between the moment-1 maps for HI and CO at matched  $6.3''$  resolution. There are significant velocity offsets towards some regions in the north (up to  $\sim 100 \text{ km s}^{-1}$ ). Conversely, there is reasonably good agreement between CO and HI velocities towards the south (mostly within  $\pm 20 \text{ km s}^{-1}$ ).

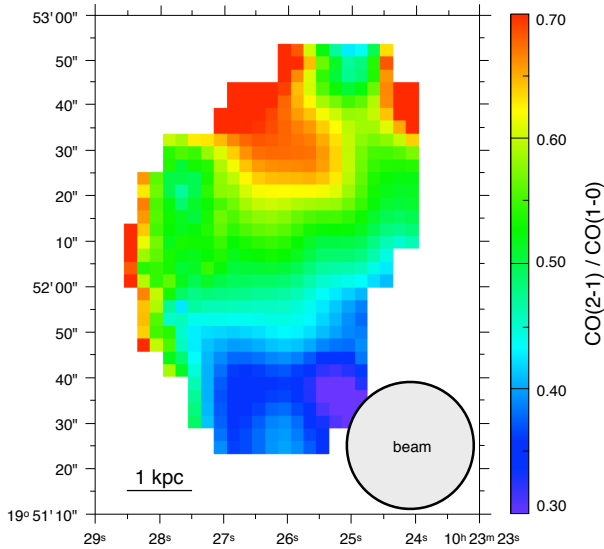
The middle and left panel of Fig. 6 focus on the highest resolution available from ALMA. These high-resolution moment maps show differences with HI since the CO peaks seem to be organised into several kinematically coherent filaments. Rather than a smooth gradient, there are quite abrupt changes in the velocities of these filaments. This could be indicating that the high-density interstellar medium (ISM) is made up of relatively independent sub-structures, embedded in a more diffuse molecular medium.

Figure 7 shows position-velocity diagrams of HI along the three slits indicated in Fig. 6. Simultaneously, we track as red contours the diffuse CO emission at a matched resolution of  $6.3''$  (i.e. the emission that is filtered out by the interferometer, like the bottom panels of Fig. 4). To first-order approximation, there is very good agreement between the atomic gas and diffuse CO, and they both reflect the north-south velocity gradient seen at lower resolution. There are two noteworthy differences, though: (1) in the NE ridge, there is plenty of diffuse CO at low velocities ( $\sim 1050\text{--}1150 \text{ km s}^{-1}$ ) where the VLA cube did not register





**Fig. 4.** *Top panels:* compact emission traced by the ALMA 12 m + 7 m arrays without TP. *Bottom panels:* diffuse molecular emission obtained by subtracting the compact emission channel by channel (*top panels*) from the total molecular emission measured by ALMA 12 m + 7 m + TP (i.e. including short spacings); this is the emission that is filtered out by the interferometer. *Left panels:* zeroth-order moment maps, measuring the integrated intensity (in  $\text{K km s}^{-1}$ ). *Middle panels:* first-order moment maps, representing the velocity field (in  $\text{km s}^{-1}$ ). *Right panels:* second-order moment maps, tracing CO linewidths ( $\sigma$  in  $\text{km s}^{-1}$ ). In all panels, the grey circle in the bottom right represents the FWHM of the synthesised beam ( $6.3''$ ).



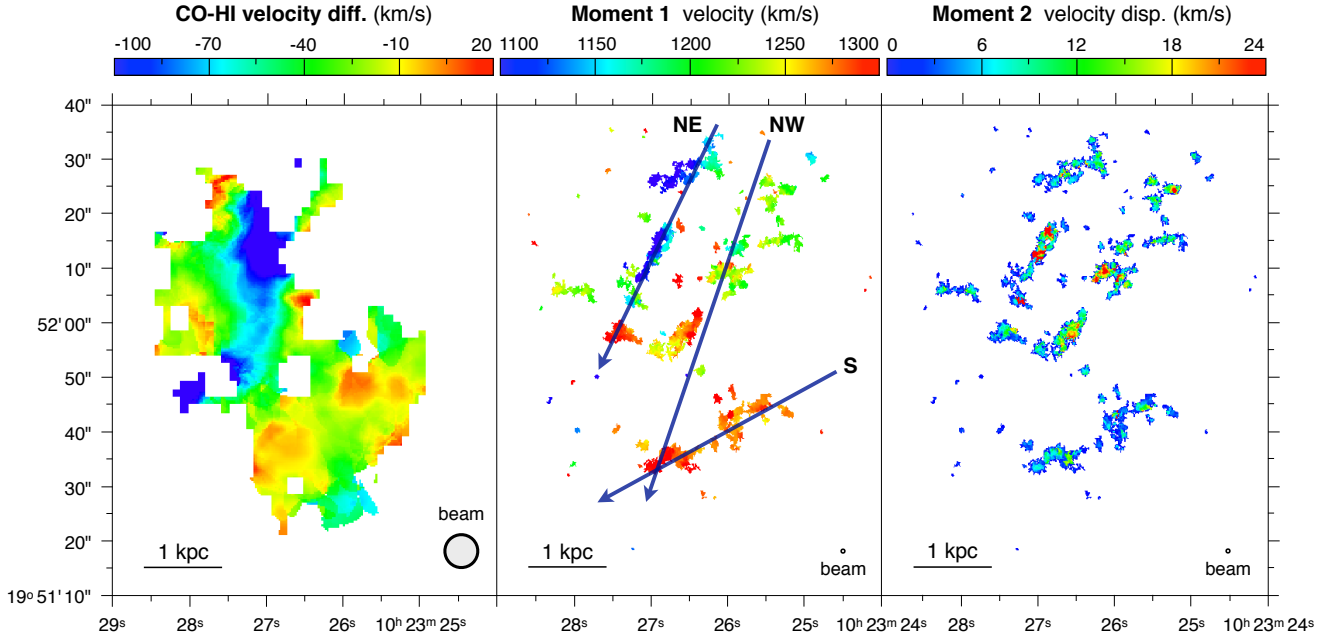
**Fig. 5.** CO(2–1)/CO(1–0) ratio at  $28''$  resolution.

significant HI emission; (2) there is an offset of  $\sim 30 \text{ km s}^{-1}$  between HI and diffuse CO in the star-forming ridge in the south.

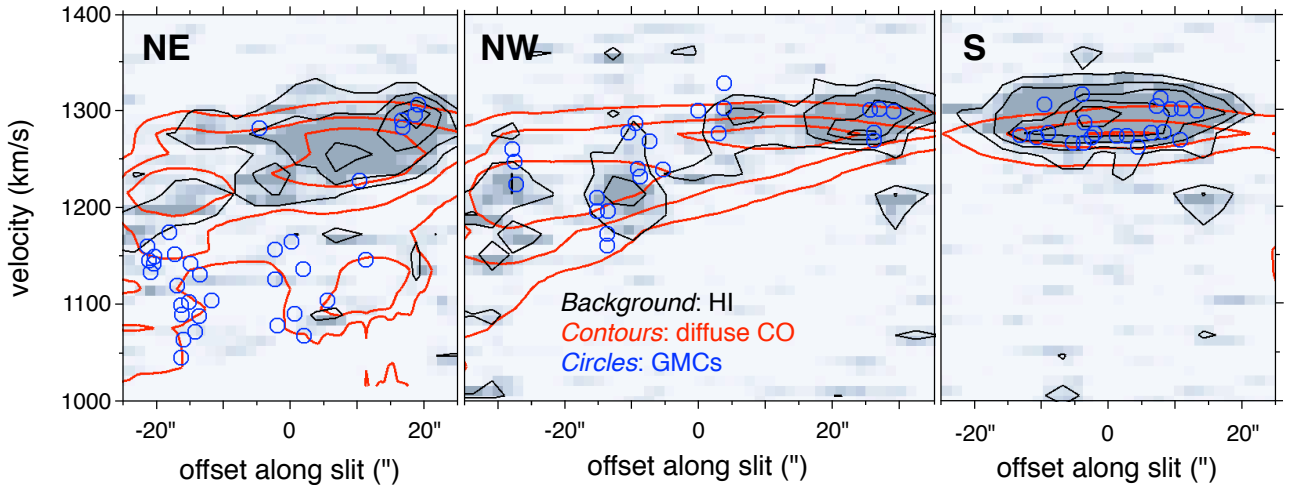
We can also examine the distribution of GMCs (Sect. 3) on the position-velocity diagrams of Fig. 7. GMCs tend to cluster

around regions of enhanced HI emission, except in the NE ridge, where most GMCs are offset from HI by  $\sim 50\text{--}150 \text{ km s}^{-1}$  towards lower velocities. There is diffuse CO emission coinciding with or near those offset GMCs, but almost no significant HI visible (some diffuse, faint HI emission is however visible in a deeper cube from the C-array of the VLA; Mundell et al. 1995b). This offset towards lower velocities is suggestive of an overlap of gaseous structures along the line of sight: the distribution of diffuse emission and GMCs is rather continuous, starting from the main HI emission around  $\sim 1200\text{--}1300 \text{ km s}^{-1}$  and extending towards velocities closer to the disc of NGC 3227 ( $\sim 1000 \text{ km s}^{-1}$ ). This may indicate that a gaseous bridge connects the TDG with the background spiral galaxy. The clustering of clouds at lower velocities can be well visualised through a rotating position-position-velocity representation of the GMCs which is available [online](#).

Another important issue from the kinematic point of view is the higher CO velocity dispersion found in the north of the TDG compared to the south. Lisenfeld et al. (2008) suggested that this might reflect an increase in large-scale turbulence in the north that might, in turn, prevent gas from forming stars in that part of the TDG. Taking the velocity centroid from each GMC (Sect. 4.3), we computed the statistical dispersion of those velocities for the ensemble of clouds in the north and in the south. Within GMCs, the intrinsic velocity dispersions are slightly higher in the north (median  $\sigma_{\text{GMC}} \sim 6.5 \text{ km s}^{-1}$ ) than in



**Fig. 6.** *Left panel:* difference between moment-1 maps for CO (ALMA 12 m + 7 m + TP) and H I at matched 6.3'' resolution. *Middle panel:* CO velocity field at 0.64''  $\approx$  45 pc resolution measured as the moment-1 map from ALMA (12 m + 7 m + TP), using a 4 $\sigma$  dilated mask. The blue arrows are the slits used for the position-velocity diagrams from Fig. 7. *Right panel:* CO velocity dispersion measured as the moment-2 map from ALMA (12 m + 7 m + TP), using the same 4 $\sigma$  dilated mask.



**Fig. 7.** Position-velocity diagrams of H I (at 6.3'' resolution) in greyscale and black contours along the three slits indicated in Fig. 6. Red contours show the distribution of diffuse CO emission from ALMA (12 m + 7 m + TP minus 12 m + 7 m only), corresponding to 6, 9, 12 mK. Blue circles indicate the location of GMCs identified with CPROPS.

the south ( $\sigma_{\text{GMC}} \sim 4.5 \text{ km s}^{-1}$ ), and the relative cloud-to-cloud velocity dispersion is also higher in the north ( $\sigma_{\text{cloud-cloud}} \sim 73 \text{ km s}^{-1}$ ) than in the south ( $\sigma_{\text{cloud-cloud}} \sim 46 \text{ km s}^{-1}$  and as low as  $16 \text{ km s}^{-1}$  if restricted to the star-forming area).

#### 4.3. GMC properties and scaling relations

The exquisite resolution of ALMA (0.64''  $\approx$  45 pc) allowed us to resolve large and intermediate-size GMCs in this system, with molecular gas masses ranging from  $\sim 10^4 M_{\odot}$  to  $\sim 10^6 M_{\odot}$ . We estimated our 5 $\sigma$  completeness limit as  $5 \times 10^4 M_{\odot}$  based on the point-source sensitivity derived in Sect. 2.1.

We employed the CPROPS algorithm to identify GMCs (Rosolowsky & Leroy 2006), as described in Sect. 3.1. In Appendix A.2 we also present an alternative approach using

SCIMES (Colombo et al. 2015), a dendrogram-based method (Rosolowsky et al. 2008) which, on top of segmenting the molecular ISM, informs on how it is hierarchically structured. In both cases, we follow the same strategy to derive physical properties (Sect. 3.2). The results for the cloud ensemble derived using CPROPS and SCIMES are very similar, and that is why we focus on CPROPS next, while we point the interested reader to Appendix A.2 for the details on SCIMES. The final catalogue of GMC properties from CPROPS is presented in Table A.3.

##### 4.3.1. GMC properties

We find a total of 111 GMCs with CPROPS (out of which 81 have extrapolated and deconvolved values), and Fig. 8 shows their distribution, sizes, and orientation. Far from clustering

mostly towards the star-forming south (19 clouds), the number of GMCs is larger in the north (62 clouds), even though the projected area in the north is only  $\sim 50\%$  larger. As expected, the distribution of GMCs follows the areas of strongest CO emission. Some GMCs are rounder (axis ratio  $\sim 1$ ) while others are more elongated ( $\sim 3$ ), and reassuringly we find no correlation between the orientation of the GMCs and the PA of the ALMA synthesised beam.

Table 3 summarises the typical properties of the GMCs. We find GMCs with radii of a few tens of parsec (up to  $\sim 100$  pc), luminous masses ranging  $\sim 10^4$  to  $\sim 10^6 M_\odot$ , and velocity dispersions of a few  $\text{km s}^{-1}$ . These values are comparable to massive GMCs in the Milky Way and other nearby galaxies (e.g. Bolatto et al. 2008; Heyer et al. 2009; Colombo et al. 2014; Faesi et al. 2018), in spite of the starkly different nature of the TDG. Virial masses are larger than luminous masses, as we discuss later.

We find no significant differences between global properties of GMCs in the north and south of the TDG other than  $\sim 50\%$  higher  $\sigma_v$  and  $M_{\text{vir}}$  in the north (Table 3). We confirmed that these results do not change qualitatively if we rely on the segmentation from SCIMES instead of CPROPS (Table 3).

#### 4.3.2. GMC mass spectra

The GMC mass spectrum is an indicator of GMC formation and dispersal timescales, modulated by mechanisms such as feedback and cloud-cloud collisions, and it is expected to depend on environment (Inutsuka et al. 2015; Dobbs et al. 2015; Torii et al. 2015; Kobayashi et al. 2017). Shallower GMC mass spectra imply a higher proportion of massive clouds within a given population; we denote the slope of the mass spectrum as  $\gamma$ .

Our completeness limit of  $5 \times 10^4 M_\odot$  is deeper than the PAWS survey in M51 ( $3.6 \times 10^5 M_\odot$ ; Colombo et al. 2014), but not as deep as the study of Faesi et al. (2018) in NGC 300 ( $8 \times 10^3 M_\odot$ ). This means that the smallest clouds may be unresolved and blending is likely for some.

To compare with GMC mass spectra in other galaxies, here we focus on CPROPS, which is more widely used than SCIMES (shown in Appendix A). Unlike for GMC scaling relations, here we consider all the clouds identified by CPROPS (82 in the north and 29 in the south), as they all have extrapolated luminosities that permit to derive total masses ( $M_{\text{GMC}} = \alpha_{\text{CO}}^{2-1} L_{\text{CO}}$ ). Rather than a binned histogram (differential form), we fitted a cumulative mass distribution, which has been argued to be more robust in the case of small samples (Rosolowsky 2005) and its use is more extended. When the cumulative GMC mass spectrum is well described by a power law, it can be expressed by the following formula:

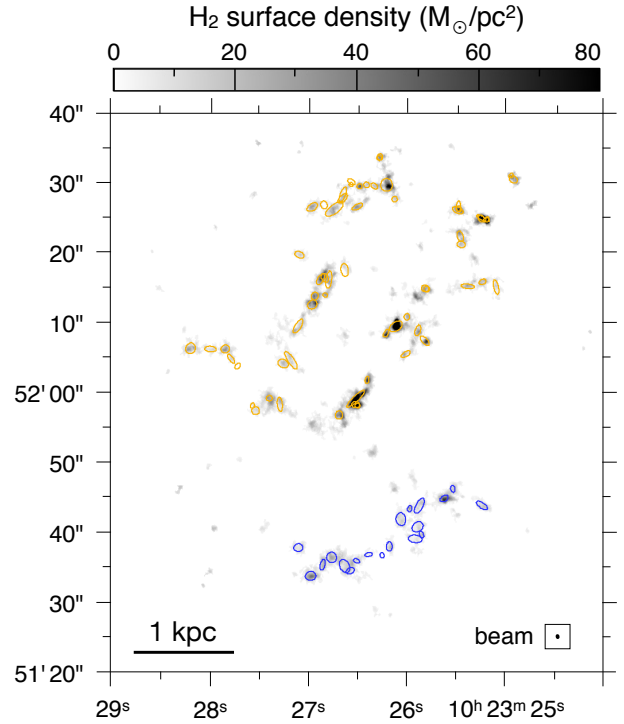
$$N(M_{\text{GMC}} > M) = \left( \frac{M}{M_0} \right)^{\gamma+1}, \quad (4)$$

where  $M_0$  is some reference mass (normalisation factor) and  $\gamma$  is the slope of the power law.

Generally, the GMC mass spectrum can be better described as a truncated power law:

$$N(M_{\text{GMC}} > M) = N_0 \left[ \left( \frac{M}{M_0} \right)^{\gamma+1} - 1 \right], \quad (5)$$

where  $N_0$  is the number of clouds more massive than  $2^{1/(\gamma+1)} M_0$ , the truncation mass where the distribution deviates from a power law.



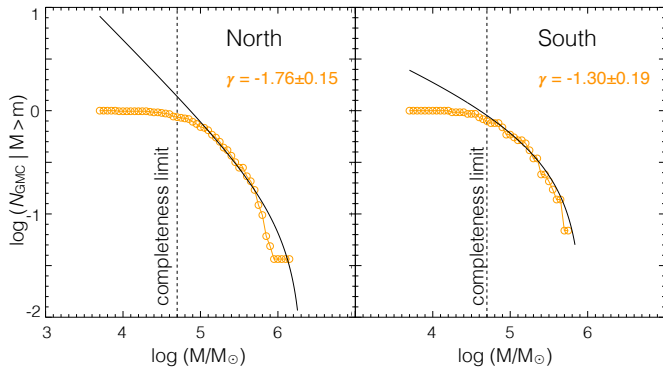
**Fig. 8.** GMC positions and orientations indicated on the ALMA CO(2–1) integrated intensity map. GMCs are shown as ellipses with the CPROPS extrapolated and deconvolved major and minor axes (2nd moment of emission). Orange and blue ellipses represent clouds in the north and south of the TDG, respectively. The bottom-right corner shows the ALMA synthesised beam ( $0.69'' \times 0.60''$  with  $\text{PA} = 11^\circ$ ).

We fitted both a truncated and non-truncated power law and found that the truncated version applies better to our data. Considering all GMCs in the TDG simultaneously (111 clouds), we find a slope of  $\gamma = -1.76 \pm 0.13$ . This value is compatible with the slope of the GMC mass spectrum in the inner Milky Way, where Rice et al. (2016) found  $\gamma = -1.59 \pm 0.11$ ; for the first Galactic quadrant, Colombo et al. (2019) also found a fully compatible value,  $\gamma = -1.76 \pm 0.01$ . Similar values have been found in M31 ( $\gamma = -1.63 \pm 0.2$ ; Rosolowsky 2007); in M33 ( $\gamma \sim -1.6 \pm 0.2$ ; Gratier et al. 2012; Braine et al. 2018); in the molecular ring and density-wave spiral arms in M51 ( $\gamma = -1.63 \pm 0.17$  to  $\gamma = -1.79 \pm 0.09$ ; Colombo et al. 2014); or in NGC 300 ( $\gamma = -1.76 \pm 0.07$ ; Faesi et al. 2018). Higher values have been reported in some cases: the outer Milky Way ( $\gamma = -2.1 \pm 0.2$ ; Rosolowsky 2005); the LMC ( $\gamma = -2.33 \pm 0.16$ ; Wong et al. 2011); in the outer M33 ( $\gamma = -2.3 \pm 0.2$ ; Gratier et al. 2012); in the material arm and the interarm region in M51 ( $\gamma = -2.44 \pm 0.40$  to  $\gamma = -2.55 \pm 0.23$ ; Colombo et al. 2014); in the lenticular galaxy NGC 4526 ( $\gamma = -2.39 \pm 0.03$ ; Utomo et al. 2015); and in the strongly barred spiral NGC 1300 ( $\gamma = -2.20 \pm 0.04$ ; Maeda et al. 2020a). Tosaki et al. (2017) found a somewhat lower slope for GMCs in NGC 1068,  $\gamma = -1.25 \pm 0.07$ , similar to the value derived by Colombo et al. (2014) along the nuclear bar of M51 ( $\gamma = -1.33 \pm 0.21$ ). All of these studies used the cumulative version of the mass spectrum, which should make the results comparable, but we note that Gratier et al. (2012) used a different algorithm to fit the cumulative mass distribution and, in general, differences in the detailed cloud segmentation strategies can affect the derived slopes.

We also checked the difference between the GMC mass spectrum in the quiescent north and in the star-forming south of the

**Table 3.** GMC properties as derived by CPROPS.

Property	Unit	TDG			north			south		
		Min	Median	Max	Min	Median	Max	Min	Median	Max
$T_{\max}$	K	0.3	0.4	1.0	0.3	0.4	0.8	0.3	0.5	1.0
$L_{\text{CO}}$	$\text{K km s}^{-1} \text{ pc}^2$	$1.8 \times 10^3$	$3.9 \times 10^4$	$4.5 \times 10^5$	$1.8 \times 10^3$	$4.1 \times 10^4$	$4.5 \times 10^5$	$3.7 \times 10^3$	$3.5 \times 10^4$	$1.6 \times 10^5$
$R$	pc	12.3	58.0	123.4	12.3	59.7	123.4	15.8	58.0	102.5
$\sigma_v$	$\text{km s}^{-1}$	0.9	5.8	19.7	0.9	6.6	19.7	0.9	4.5	9.1
$M_{\text{lum}}$	$M_{\odot}$	$8.1 \times 10^3$	$1.7 \times 10^5$	$2.0 \times 10^6$	$8.1 \times 10^3$	$1.8 \times 10^5$	$2.0 \times 10^6$	$1.6 \times 10^4$	$1.5 \times 10^5$	$7.0 \times 10^5$
$M_{\text{vir}}$	$M_{\odot}$	$2.3 \times 10^4$	$2.7 \times 10^6$	$4.0 \times 10^7$	$2.8 \times 10^4$	$3.4 \times 10^6$	$4.0 \times 10^7$	$2.3 \times 10^4$	$2.3 \times 10^6$	$7.8 \times 10^6$


**Fig. 9.** Cumulative GMC mass distribution for the two environments in the TDG: north (left) and south (right). The vertical dashed lines indicate our completeness limit of  $5 \times 10^4 M_{\odot}$ .

TDG. Interestingly, we found that the south has a shallower mass spectrum,  $\gamma = -1.30 \pm 0.19$  instead of  $\gamma = -1.76 \pm 0.15$ . The difference is marginally significant given the large error bars (due to the relatively low number of clouds in each environment), but it points at an interesting idea: the higher proportion of more massive clouds in the south might be connected with the star formation activity in that region. This would be consistent with similar observations in star-forming rings or spiral arms in nearby galaxies, where shallower GMC mass spectra correlate with more intense star formation (Colombo et al. 2014; Tosaki et al. 2017).

Another relevant quantity is the truncation mass of the GMC mass spectrum. Fitting all GMCs in the TDG, we found a value of  $M_0 = 1.9 \times 10^6 M_{\odot}$ . While M51 shows a significantly higher truncation mass ( $\sim 2 \times 10^7 M_{\odot}$ ; Colombo et al. 2014),  $M_0$  in the TDG is of the same order as the LMC, NGC 4526, NGC 300, or some Galactic results ( $M_0 \sim 1\text{--}4 \times 10^6 M_{\odot}$ ; Wong et al. 2011; Utomo et al. 2015; Rice et al. 2016; Faesi et al. 2018; Colombo et al. 2019). These values agree with theoretical models for maximum cloud masses by Reina-Campos & Kruijssen (2017).

#### 4.3.3. GMC scaling relations

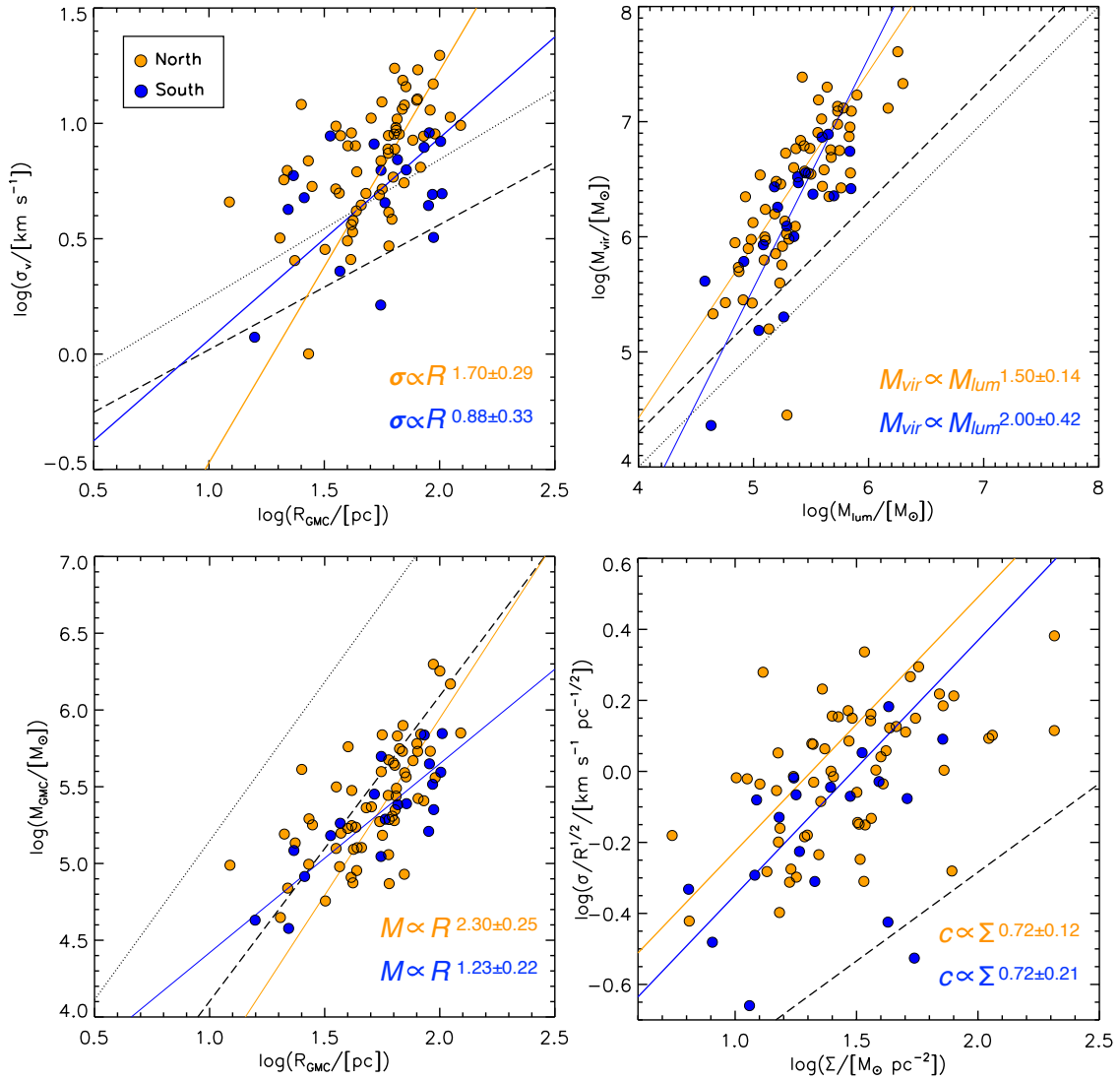
Here we focus on a widespread GMC diagnostics tool: the Larson scaling relations, shown in Fig. 10. Larson (1981) and Solomon et al. (1987) found a tight correlation between the size and linewidth of GMCs, between their virial and luminous mass, and between their size and mass. Those seminal papers examined Galactic GMCs, but more recent work has expanded the analysis to extragalactic clouds as well. We focus on CPROPS to minimise discrepancies with studies from the literature due to different segmentation codes. The best fits and Spearman rank

correlation coefficients are listed in Table 4. The linear regressions are obtained with two different methods. The first one is the popular orthogonal distance regression (ODR), where the data are fitted in the log-log plane using the Python implementation `scipy.odr`<sup>2</sup> that considers uncertainties in both variables simultaneously. As a complementary approach, we also fitted the observations using the publicly available Bayesian fitting code `BayesLineFit`<sup>3</sup> (Lelli et al. 2019) that additionally considers the intrinsic scatter along the relation. `BayesLineFit` provides similar results as the ODR method but larger uncertainties (from exploring the full posterior distribution of the free parameters). In both cases, these fits are based on the extrapolated and deconvolved GMC properties (see Sect. 3.2), which result in substantially higher error bars; for instance, the median uncertainty on the cloud radius is 18%, but as much as 55% when considering extrapolation and deconvolution.

*Size-linewidth relation.* A power-law relationship between size and velocity dispersion of GMCs has been systematically found in the Milky Way, with a slope of  $\sim 0.5$ , which has been interpreted as a manifestation of compressible turbulence within the molecular ISM (e.g. Solomon et al. 1987; Heyer et al. 2001; Rice et al. 2016; Kritsuk et al. 2017). Observations of the Galactic central molecular zone imply a slightly steeper size-linewidth relation with a slope of  $\sim 0.7$  (Shetty et al. 2012; Kauffmann et al. 2017). The clouds in the TDG seem to have a moderate degree of correlation in this plane (left panel of Fig. 10), but the slope  $\gamma = 2.00 \pm 0.33$  is significantly steeper than the canonical Galactic value of  $\sim 0.5$ . If we subdivide clouds into our two environments, we find a steeper slope for the north, but the difference is only marginally significant given the large error bars.

Within the extragalactic literature, there is some controversy as to whether a genuine size-linewidth relation holds. Bolatto et al. (2008) combined data from several nearby galaxies and found evidence for a size-linewidth relation ( $\gamma = 0.60 \pm 0.10$ ). On the other hand, a number of studies found that their data did not support the existence of a size-linewidth relation (e.g. Gratier et al. 2012 in M33; Colombo et al. 2014 in M51; Hughes et al. 2013a combining M33, the LMC, and M51), or only weakly (e.g. Braine et al. 2018 in M33). More recently, Faesi et al. (2018) found a size-linewidth relation in NGC 300 with a slope of  $0.48 \pm 0.05$ , with a larger dynamic range in GMC masses than the previous studies. Our data have a similar degree of correlation as found by Faesi et al. (2018) in NGC 300 (they quote a Pearson coefficient  $r_p = 0.55$ ). However, the substantially steeper slope of the size-linewidth relation in the TDG suggests

<sup>2</sup> <https://docs.scipy.org/doc/scipy/reference/odr.html>
<sup>3</sup> `BayesLineFit` is available at <http://astroweb.cwru.edu/SPARC/>



**Fig. 10.** Larson’s relations for the ensemble of GMCs identified in the TDG using CPROPS. We differentiate between the quiescent environment in the north (orange points) and the star-forming region in the south (blue points). The solid lines represent the best ODR fits to the data (orange for north, blue for south). *Top left panel:* size–linewidth relation. The dashed line is the relation for Galactic clouds from [Heyer et al. \(2001\)](#) and the dotted line represents the best fit to extragalactic clouds from [Bolatto et al. \(2008\)](#). *Top right panel:* virial mass–luminous mass relation. The dotted and dashed lines are the 1:1 and 2:1 relations, respectively. *Bottom left panel:* size–mass relation. The dashed and dotted lines are the relations for Milky Way clouds from [Lombardi et al. \(2010\)](#) for an extinction threshold of  $A_K = 0.1$  and 1.5, respectively. *Bottom right panel:* Heyer plot showing  $c = \sigma/\sqrt{R}$  as a function of the molecular gas mass surface density; the dashed line is the Galactic relation from [Heyer et al. \(2001\)](#).

**Table 4.** Fits and rank coefficients for GMC scaling relations with CPROPS.

Relation	Method	TDG		North		South		MW Slope
		Slope	$\rho$	Slope	$\rho$	Slope	$\rho$	
Size–linewidth	ODR	$2.00 \pm 0.33$	0.49 (0.00)	$1.70 \pm 0.29$	0.62 (0.00)	$0.88 \pm 0.33$	0.30 (0.21)	0.5 <sup>(a)</sup>
	Bayesian	$1.69^{+0.57}_{-0.39}$		$1.50^{+0.55}_{-0.37}$		$0.86^{+2.59}_{-0.80}$		
Virial relation	ODR	$1.66 \pm 0.13$	0.80 (0.00)	$1.50 \pm 0.14$	0.81 (0.00)	$2.00 \pm 0.42$	0.77 (0.00)	0.8 <sup>(a)</sup>
	Bayesian	$1.59^{+0.22}_{-0.18}$		$1.43^{+0.26}_{-0.20}$		$1.70^{+3.47}_{-1.19}$		
Mass–size	ODR	$2.08 \pm 0.19$	0.70 (0.00)	$2.30 \pm 0.25$	0.68 (0.00)	$1.23 \pm 0.22$	0.77 (0.00)	2.0 <sup>(a)</sup>
	Bayesian	$3.25^{+0.57}_{-0.45}$		$3.76^{+1.11}_{-0.73}$		$2.27^{+0.90}_{-0.61}$		
Heyer plot	ODR	$0.91 \pm 0.13$	0.46 (0.00)	$0.72 \pm 0.12$	0.45 (0.12)	$0.72 \pm 0.21$	0.44 (0.06)	0.5 <sup>(b)</sup>
	Bayesian	$0.41^{+0.10}_{-0.09}$		$0.32^{+0.10}_{-0.09}$		$0.43^{+0.25}_{-0.18}$		

**Notes.**  $\rho$  is the Spearman rank correlation coefficient, with the  $p$ -value indicated in parentheses. For each dataset we used two different fitting methods: orthogonal distance regression (ODR) and a Bayesian code, BayesLineFit. <sup>(a)</sup>[Solomon et al. \(1987\)](#). <sup>(b)</sup>[Heyer et al. \(2009\)](#).

that the organisation of the molecular medium into GMCs is different in this special environment.

*Virial relation.* The second Larson relation suggests that GMCs are self-gravitating, since virial and luminous masses track each other. The top-right panel of Fig. 10 shows a strong correlation between virial and luminous masses. The global slope  $1.66 \pm 0.13$  is considerably higher than the  $\sim 0.8$ – $1$  value found in the Milky Way (Solomon et al. 1987) and in several nearby galaxies (Bolatto et al. 2008; Hughes et al. 2010; Faesi et al. 2018). Colombo et al. (2014) derived values for M51 that approach the one of the TDG ( $\gamma \sim 1.3$  in the centre and spiral arms of M51,  $\gamma \sim 1.5$  in the inter-arm region). Both the north and south of the TDG have similar slopes and a comparable degree of correlation.

Apart from the slope of the scaling relation, the virial masses in the TDG are on average  $\sim 1$  dex higher than the luminous masses. Taken at face value, this result suggests that the most massive clouds are far from virial equilibrium and likely unbound, which is unusual for normal galaxies, but several factors could contribute to the discrepancy between  $M_{\text{vir}}$  and  $M_{\text{lum}}$ . Firstly, the assumed distance of 14.5 Mpc to Arp 94 might be wrong, and this would affect the virial mass linearly but the luminous mass quadratically. Distance estimates to this system range within  $\sim 10$ – $20$  Mpc, with a few outliers at higher distances; even if the real distance to Arp 94 is as much as 30 Mpc, that could explain at most a factor of two offset in the  $M_{\text{vir}} - M_{\text{lum}}$  relation, far from the factor of ten observed for the most massive clouds. Another important source of uncertainty is the  $\alpha_{\text{CO}}$  conversion factor. A typical uncertainty of a factor  $\sim 2$  is expected (Sandstrom et al. 2013); the near-solar metallicity of the TDG suggests that changes in  $\alpha_{\text{CO}}$  should not be much larger than that but, technically, we cannot rule out a variation of a factor  $\sim 10$  in  $\alpha_{\text{CO}}$  (as  $\alpha_{\text{CO}}$  has not been specifically characterised in TDGs); if this extremely high value of  $\alpha_{\text{CO}} \sim 40$ – $50 M_{\odot} (\text{K km s}^{-1} \text{ pc}^2)^{-1}$  is true, that could fully account for the observed offset in  $M_{\text{vir}} - M_{\text{lum}}$ . What it would not be able to explain is the observed non-linear relation, implying that more massive clouds depart more severely from the virial estimate. The cloud segmentation process could also introduce some systematic uncertainty, particularly for lower-mass clouds. However, the fact that the offset is especially pronounced for high-mass clouds makes segmentation issues an unlikely cause of the observed vertical offset. Finally, the overlap of GMCs along the line of sight could broaden linewidths, raising virial masses relative to the luminous masses.

In any case, it is physically plausible that clouds in the TDG are not (yet) in virial equilibrium between their kinetic energy and self-gravity, given that this is a newly formed galaxy that is in a dynamically complex context, where external pressure could play an important role (see e.g. Schrubba et al. 2019; Sun et al. 2020). The most massive (larger) clouds tend to have higher velocity dispersions, making both the size-linewidth and virial relations super-linear. We also see some environmental variation in the virial relation within the TDG. The virial parameter (approximately equivalent to the ratio  $M_{\text{vir}}/M_{\text{lum}}$ ) is higher for the clouds in the north (average  $\sim 13$ ) than for the clouds in the south (average  $\sim 8$ ), with significant overlap between both populations.

*Mass-size relation.* The third Larson relation describes an inverse correlation between size and density, implying that all GMCs have approximately constant surface density. The three Larson’s relations are not independent from each other, and the applicability of two of Larson’s relations immediately implies the third (e.g. Heyer & Dame 2015). We find a statistically significant correlation between GMC mass and size in our data,

with a global slope of  $2.08 \pm 0.19$ . A similar slope of  $\sim 2$  has been found in the Milky Way (Solomon et al. 1987), in M51 (Colombo et al. 2014), in the LMC (Hughes et al. 2010), in combined data for nearby galaxies (Bolatto et al. 2008), or in NGC 300 (Faesi et al. 2018). Our observations are also in good agreement with the fit to Galactic data by Lombardi et al. (2010) for an extinction of  $A_K = 0.1$ , corresponding to a surface density of  $\sim 40 M_{\odot} \text{ pc}^{-2}$ .

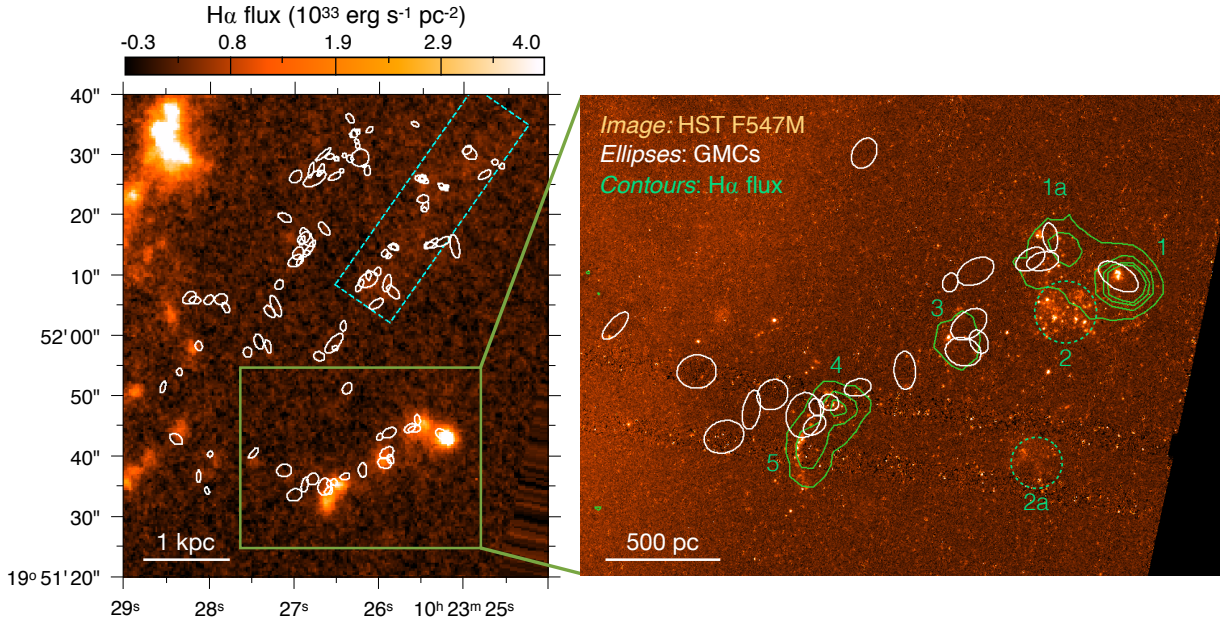
*Heyer plot.* Beyond Larson’s relations, Heyer et al. (2009) introduced a plot condensing all three relations and highlighting systematic departures. If Larson’s relations are strictly obeyed, data points should cluster around a point centred at  $c = \sigma / \sqrt{R} = (\pi G \Sigma / 5)^{1/2}$  for a constant value of  $\Sigma$ . However, similar to the findings from Heyer et al. (2009) in the Milky Way, the clouds in the TDG span an order of magnitude or more along each of the axes of the plot and show a positive correlation. For a given  $\Sigma$ , the points in the TDG are clearly displaced towards higher values of  $c$ , driven by the high velocity dispersions.

Given the variations in  $\Sigma$  and  $c$ , and the moderate but significant positive correlation, our observations suggest that the velocity dispersion of a cloud depends both on the size of the GMC and its mass surface density, in agreement with Heyer et al. (2009). Yet, at the same time, the vertical offset with respect to the Galactic observations suggests that GMCs in the TDG behave differently from those in the Milky Way, as also indicated by the departures from the first and second Larson’s relations.

#### 4.4. Comparison between GMCs and star formation sites

Star formation predominantly happens in the south of the TDG, as shown by a variety of tracers in the optical and infrared, confirmed with spectroscopy (Mundell et al. 2004; Lisenfeld et al. 2008). Figure 11 shows archival H $\alpha$  and HST WFC3/F547M observations that clearly reveal emission from young stars formed in the south of the TDG. There are also some hints of diffuse but mildly enhanced H $\alpha$  emission towards the north-west of the TDG. At lower resolution, a similar flux enhancement is present in the *Herschel* PACS  $160 \mu\text{m}$  map. When averaged inside a rectangle of  $40'' \times 11''$  with PA =  $55^\circ$  centred on RA = 10:23:25.36 and Dec = +19:52:21.7 (cyan dashed line in Fig. 11), the H $\alpha$  flux is  $\sim 1.5$ – $2\times$  higher than in the immediate surroundings (mean H $\alpha$  flux of  $2.2 \times 10^{32} \text{ erg s}^{-1} \text{ pc}^{-2}$ , as opposed to  $1.4$  and  $1.1 \times 10^{32} \text{ erg s}^{-1} \text{ pc}^{-2}$  on adjacent rectangles towards the north or south, respectively). For reference, the H $\alpha$  contrast between the southern star-forming region and its surroundings (a factor of  $\sim 8$ ) is considerably higher than in the north-west. There are some scattered compact sources visible on the HST image towards the north of the TDG which might be young massive clusters (see Fensch et al. 2019, for the efficient formation of clusters in TDGs). However, we also see similar knots towards the eastern side of NGC 3227, far from the TDG, making it likely that these compact sources coincide in position with the TDG simply due to projection, something that we cannot test further without spectroscopy. In the rest of this section, we focus on the predominant site of star formation in the south of the TDG.

Figure 11 also identifies with green numbers the star-forming knots from Lisenfeld et al. (2008). Most of these star-forming knots are overlapping with one or more GMCs, albeit there can be small projected offsets of about  $0.8$ – $1.6'' = 56$ – $112 \text{ pc}$ . Exceptions are represented by knots 2 and 5 that have no associated GMC (the nearest one is offset by  $2$ – $3'' = 140$ – $210 \text{ pc}$ ). Knot 2 was found to be the oldest in Lisenfeld et al. (2008) with ages above  $\sim 100 \text{ Myr}$ . This is consistent with the idea that young stars have had sufficient time to photodissociate or disperse the



**Fig. 11.** *Left panel:*  $H\alpha$  image showing intense star formation in the south of the TDG, with white ellipses indicating the position, size, and orientation of GMCs identified with CPROPS; the cyan dashed rectangle marks an area of mildly enhanced  $H\alpha$  emission as discussed in the text. *Right panel:* HST WFC3/F547M image zoomed on the area marked in the left panel, revealing young stellar clusters in the south of the TDG. White ellipses indicate GMCs and green contours show constant  $H\alpha$  flux ( $0.8, 2, 3,$  and  $4 \times 10^{33} \text{ erg s}^{-1} \text{ pc}^{-2}$ ). The numbering of regions follows Lisenfeld et al. (2008); the dashed green regions are visible in  $B$  band but lack detectable  $H\alpha$  emission.

surrounding molecular gas; it would also explain why we no longer see  $H\alpha$  emission. The clusters revealed by HST around knot 2 are very bright and cover a large area, reinforcing the idea that those young stars produce plenty of ionising radiation.

Finally, Lisenfeld et al. (2008) identified 2a as a region of stellar emission but without  $H\alpha$ . The HST imaging reveals weak emission throughout this area, but it is clearly less intense than the other knots. We did not detect any GMCs around this region (the closest GMC is at a projected distance of  $7'' \sim 500 \text{ pc}$ ). It can well be that 2a is similar to knot 2 and the age is higher than the rest ( $\sim 100 \text{ Myr}$ ), such that there was enough time for any surrounding leftovers of molecular gas to get photodissociated or dispersed after the episode of star formation.

## 5. Discussion

### 5.1. A very large fraction of diffuse molecular gas

In Sect. 4.1.2, we found that as much as  $\sim 80\%$  of the molecular emission in the TDG is extended, arising from scales of several kpc. This is likely tracing diffuse molecular gas, with a remarkably high contribution to the total CO budget. In fact, there is mounting evidence that nearby galaxies have a sizeable fraction of diffuse molecular emission, which is detected by single-dish telescopes but filtered out by interferometers. As opposed to compact molecular gas structures, this emission has been interpreted as evidence of thick molecular discs (e.g. Pety et al. 2013). In principle, such extended emission could also be due to an ensemble of small clouds that are homogeneously distributed and separated by less than the synthesised beam of the interferometer (e.g. Pety et al. 2013; Caldú-Primo et al. 2015; Caldú-Primo & Schruba 2016). Dame & Thaddeus (1994) also found evidence for a thick molecular disc in the Milky Way (comparable in thickness to the HI layer), and similar thick molecular discs have been reported for external galaxies, such as the edge-on spiral NGC 891 (García-Burillo et al. 1992) and

M51 (e.g. Pety et al. 2013). The fraction of diffuse gas found by these studies varies greatly from galaxy to galaxy, and is method-dependent, but typical values are  $\sim 10\text{--}60\%$  (Pety et al. 2013; Caldú-Primo et al. 2015; Chevance et al. 2020; Maeda et al. 2020b). It is noteworthy that Maeda et al. (2020b) found a particularly high fraction of diffuse molecular emission in the bar of NGC 1300: 91% in the region that they define as bar-A and 74% in bar-B, as opposed to  $\sim 30\text{--}60\%$  in the spiral arm and bar end.

In agreement with the findings in other galaxies, this diffuse CO component is associated with high velocity dispersions ( $\sigma \sim 50\text{--}60 \text{ km s}^{-1}$  as opposed to  $\sigma \lesssim 20\text{--}30 \text{ km s}^{-1}$  for the compact CO component at matched resolution; Fig. 4). Interestingly, the position-velocity diagrams from Fig. 7 generally show a very good agreement between this diffuse molecular component and atomic gas: even in regions of position-velocity space lacking compact molecular emission (GMCs), the diffuse molecular and atomic components seem to be well coupled.

It is likely that the peculiar and pristine nature of the TDG can explain this very high fraction of diffuse molecular emission. In the interacting system Arp 94 where the TDG formed, the dynamical timescales are probably long, of the order of several 100 Myr or even  $\sim \text{Gyr}$  (see Sect. 5.3 below). Therefore, atomic gas in the tidal tail may have had enough time to form molecules on its own. This should be facilitated by the lack of strong UV emission from stars, unlike typical galaxy discs where radiation from young stars can easily photo-dissociate  $\text{H}_2$ . In this framework, it seems plausible that molecular gas first appeared as a diffuse phase, and then started to locally condense into more compact structures like the GMCs.

### 5.2. Spatial variations in the CO(2–1)/CO(1–0) ratio

In Sect. 4.1.3, we showed that the  $R_{21} = \text{CO}(2-1)/\text{CO}(1-0)$  ratio varies across the TDG with a mean value of  $\sim 0.5$ . Values of  $0.5\text{--}1.0$  have typically been found in the Milky Way and at  $\sim \text{kpc}$  scales in the discs of nearby galaxies (e.g. Sakamoto 1994;

Oka et al. 1996; Yoda et al. 2010; Leroy et al. 2009; den Brok et al. 2020). Therefore, the  $R_{21}$  ratio found in this newborn TDG is not abnormal compared to nearby galaxies, even though the value of  $\sim 0.35$  in the south lies at the lower end of what has been measured so far. The low  $R_{21}$  measured might be connected to the nature of the interaction occurring in this system.

Generally, low values of  $R_{21}$  may indicate that the gas is not in thermal equilibrium due to low volumetric densities, consistent with the high fraction of diffuse molecular gas that we inferred. This agrees with the recent study by Maeda et al. (2020b) in NGC 1300, who found a very high fraction of diffuse molecular emission in its bar (74–91%) mapping to a very low  $R_{21}$  (0.17–0.34). We might be witnessing a similar effect in the TDG.

The observed gradient in  $R_{21}$  in the TDG points to different physical conditions between the quiescent north and the star-forming south. Specifically, variations in the  $R_{21}$  line ratio are expected to be driven by differences in the optical depth (closely related to density) and/or the excitation temperature of the gas (Sakamoto 1994; Leroy et al. 2009). We find lower  $R_{21}$  in the south, suggesting lower excitation temperature and/or lower density; these conditions are traditionally interpreted as associated with lower star formation efficiency. This might seem puzzling given that star formation is detected almost exclusively in the south. However, the integrated  $R_{21}$  value at 28'' resolution is dominated by emission from diffuse molecular gas that is presumably not taking part in star formation. Based on synthetic observations of a simulated GMC, Peñalosa et al. (2017) found that  $R_{21}$  follows a bimodal distribution, with  $\sim 0.7$  for cold and dense parts and  $\sim 0.3$  in warmer and more diffuse regions. This is consistent with the fact that the CO emission in the TDG is dominated by diffuse molecular gas with a low  $R_{21}$  value.

### 5.3. Dynamical state and survival of the TDG

In Sect. 4.2, we showed that atomic and diffuse molecular gas as well as most GMCs in the TDG display a north-south velocity gradient of  $\sim 100 \text{ km s}^{-1}$  (Fig. 7). To first order approximation, the diffuse CO component tracks atomic gas very well, suggesting that atomic and diffuse molecular gas are well mixed. The north-south gradient may be compatible with rotation, but given the uncertain geometry of the system it could also reflect stretching motions along the tidal tail. Moreover, there is some kinematic evidence for a gaseous “bridge” along the line of sight towards the north-east direction, which is mostly made up of compact structures (GMCs) and connects the TDG with the background galaxy NGC 3227.

Overall, the velocity of the TDG along the line of sight is positive relative to the centre of NGC 3227, similar to the southern part of NGC 3227. Since extinction arguments confirm that the TDG is currently in front of the spiral galaxy (Mundell et al. 2004), the TDG must be moving towards NGC 3227, as long as the tangential velocities (which we cannot measure) are not much larger than the radial velocity. That “backward” motion towards NGC 3227 would suggest that the TDG has been detached from its parent galaxy for a considerable amount of time; the timescales for such interactions are typically of the order of several hundred Myr or even  $\sim 1$  Gyr.

It is unclear whether the TDG will survive much longer. Simulations show that condensations of matter at the tip of tidal tails have a larger probability of surviving long and eventually becoming detached from their parent system (e.g. Bournaud & Duc 2006). However, there is no way to robustly predict if a given matter condensation will survive, as this depends strongly

on the dynamics of the system, which are not well known in detail for Arp 94. In any case, we have used J1023+1952 as a template laboratory to study what a young TDG can look like, independently from the future evolution of this specific system.

### 5.4. GMC properties and Larson's relations

In Sect. 4.3, we found that our TDG contains GMCs whose masses and sizes resemble those in spiral galaxies (Milky Way, M31, M33, M51, NGC 300) and the LMC. This TDG, however, shows distinct peculiarities compared to other galaxies. First of all, it has likely not settled into a disc and is probably not in dynamical equilibrium. Secondly, the fraction of diffuse molecular gas is extremely high. Thirdly, there are no stellar populations shining yet (neither young nor old) across most of the area covered by the TDG. This implies that the processes determining the masses and sizes of GMCs are not necessarily associated with disc dynamics or stellar feedback, since those do not play an important role in our TDG. In other words, regular disc rotation or stellar feedback do not seem to be essential for GMCs to have certain sizes, masses, and to follow a mass distribution which approaches a truncated power law with a slope of  $\gamma \sim -1.8$ .

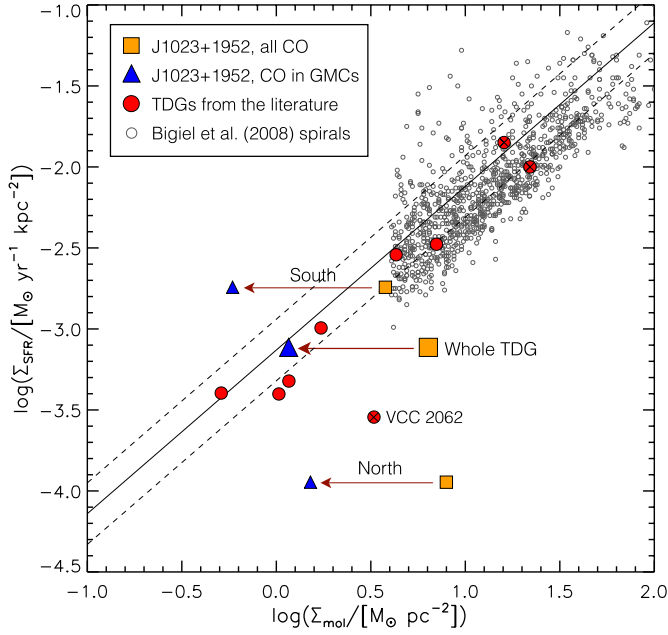
We have examined Larson's relations and found that these clouds show important departures from the scaling relations observed in the Milky Way and other galaxies. While clouds seem to obey the same size-mass relation observed in other galaxies, virial masses almost always exceed luminous masses, often by as much as an order of magnitude. This might be an observational artefact, given the large uncertainty in  $\alpha_{\text{CO}}$  in TDGs. If real, the offset could indicate that clouds are super-virial and have not settled into dynamical equilibrium yet. The higher velocity dispersion of the most massive clouds explains the super-linear trend in the size-linewidth and virial relations, steeper than observed in the Milky Way and other galaxies. This could be driven by the blending of clouds with slightly different velocities inside a beam.

### 5.5. Offsets between GMCs and young stellar clusters

In Sect. 4.4, we studied the spatial relation between young stellar clusters and GMCs in the star-forming part of the TDG. Broadly speaking, we found three different cases: (1) good spatial agreement between the position of GMCs and young clusters; (2) small but noticeable offsets between clusters and GMCs, typically of 50–100 pc; and (3) clusters not associated with any GMCs, the nearest GMC being at a projected distance of  $\geq 200$  pc (keeping in mind that the true physical separation might be larger). Interestingly, the smallest offsets coincide with the youngest star-forming knots, while the largest offsets are associated with the oldest knots, consistent with an age trend.

Our results agree with the different situations identified by Kawamura et al. (2009) in the LMC (see also Fukui et al. 1999) or the findings by Miura et al. (2012) and Gratier et al. (2012) in M33, Kreckel et al. (2018) in NGC 628, Grasha et al. (2018) in NGC 7793, and Grasha et al. (2019) in M51, with typical offsets  $\sim 40$ –100 pc if present. In particular, Grasha et al. (2018, 2019) found a clear trend between cluster age and the GMC-cluster separation, emphasising the idea that the offsets result from the time evolution of individual star-forming regions. In the framework of the uncertainty principle for star formation (Kruijssen & Longmore 2014; Kruijssen et al. 2018), Chevance et al. (2020) derived region separation lengths between molecular gas and star formation peaks ( $\lambda$ ) in the range 107–267 pc for nine nearby star-forming galaxies. Under the assumption of complete spatial





**Fig. 12.** Molecular Kennicutt-Schmidt relation for J1023+1952 compared to TDGs from the literature (compiled by Lisenfeld et al. 2016). The orange squares represent our measurements of the total molecular gas, whereas the blue triangles correspond to the flux in GMCs identified with CPROPS, associated with compact molecular emission. The small squares and triangles differentiate the north from the south of the TDG (as defined in Fig. 1). For reference, the open grey circles in the background show the kpc-scale measurements in spiral galaxies from Bigiel et al. (2008); the solid line represents their best fit to that sample and the dashed line its standard deviation. The molecular gas literature data for TDGs comes from single-dish telescopes except for the circles marked with an additional cross.

randomness, this separation length translates into an expectation value of  $\approx 0.44\lambda$  for the nearest neighbour distance (see Eq. (9) in Kruijssen et al. 2019), which results in an expected separation of  $\sim 50$ – $120$  pc between emission peaks, in good agreement with the range of observed offsets. Comparable offsets between star-forming sites and GMCs have been found using alternative tracers of star formation (e.g. infrared emission or radio continuum; Chen et al. 2010; Whitmore et al. 2014; Corbelli et al. 2017; Querejeta et al. 2019).

The observed offsets between GMCs and young clusters suggest that, once it is triggered, star formation in the TDG proceeds in a very similar fashion as in standard dwarf and spiral galaxies. Moreover, the fact that the offsets are also quantitatively similar to other galaxies points to the idea that the dispersal timescales of molecular gas around a young star-forming region is fairly universal, independent from the peculiar nature of the TDG.

### 5.6. Drivers of star formation in TDGs

To place our work within a broader context, Fig. 12 shows the molecular Kennicutt-Schmidt relation (Schmidt 1959; Kennicutt 1998) for spiral galaxies and other TDGs. When considering all the molecular gas, our TDG seems globally inefficient at forming stars, similarly to the nearby TDG VCC 2062 (Lisenfeld et al. 2016). The inefficiency is particularly acute in the north, while the south is similar to other TDGs and the extrapolation of spiral galaxies to lower surface densities.

Figure 12 also highlights how much the presence of extended CO emission can affect the apparent efficiency of star formation

on the Kennicutt-Schmidt plot. For J1023+1952, there is a considerable difference depending on whether we focus on compact structures identified as GMCs or if we account for all the CO flux. For the entire system, the star formation activity seems normal when considering the compact molecular gas in GMCs, but it appears as highly inefficient at forming stars if we also include extended emission. When inspecting the north and the south separately, we find that the star formation surface density in the north is below the value expected from the Kennicutt-Schmidt relation both for the total and compact molecular gas surface density. Conversely, the star formation activity in the south follows the Kennicutt-Schmidt relation and is even above the relation when examining the compact molecular emission.

As suggested by Lisenfeld et al. (2008), increased large-scale turbulence might be preventing gas from forming stars in the north of the TDG. Our ALMA observations confirm that the increased velocity dispersion in the north arises primarily from a higher velocity dispersion in the diffuse gas and a larger relative velocity among GMCs. Cloud-cloud collisions have also been invoked to explain the triggering of star formation (e.g. Tan 2000; Fujimoto et al. 2014; Dobbs et al. 2015; Torii et al. 2015). However, if clouds are similarly extended along the line of sight across the TDG, the lower velocity dispersion among clouds in the south would suggest a lower cloud-cloud collision rate, making this factor less likely as an explanation for the enhanced star formation.

The distribution of star forming sites along a ridge in the south, with an age gradient in the stellar clusters, could be suggestive of some propagation mechanism, such as stochastic self-propagating star formation (Mueller & Arnett 1976; Gerola & Seiden 1978; Sleath & Alexander 1995). If a given mechanism initiates star formation in a certain cloud complex, feedback from star formation could subsequently trigger the collapse of other clouds in the neighbourhood, with a certain time delay. Several studies have claimed to detect the imprint of stochastic self-propagating star formation in observations of dwarf and irregular galaxies (Feitzinger et al. 1981; Dopita et al. 1985; McCray & Kafatos 1987; Nomura & Kamaya 2001).

We found that the global gas surface density (atomic + molecular) is higher in the north, with a higher molecular-to-atomic ratio, and a larger cloud-to-cloud velocity dispersion. In this context, the lack of strong star formation in the north could be explained as an increased star formation threshold due to the higher gas density and turbulence. This would qualitatively agree with the predictions from Elmegreen (2018), and with observations of a lower star formation efficiency of the (dense) molecular gas towards the centre of the Milky Way (Longmore et al. 2013; Kruijssen et al. 2014; Battersby et al. 2020) and the centres of other nearby galaxies (Usero et al. 2015; Bigiel et al. 2016; Gallagher et al. 2018; Querejeta et al. 2019). In general, wherever molecular gas surface densities are higher, it is harder for the gas to turn into stars because the “barrier” for star formation increases. This interpretation also agrees with the observation of a higher molecular-to-atomic gas ratio in the north, which would be expected if local gas density and pressure are higher. In any case, the largest limitation is that we are dealing with projected surface densities, and the observed surface densities might not necessarily track the volumetric gas densities that presumably modulate the threshold of star formation.

Alternatively (or in addition) to differences in star formation thresholds, the triggering of star formation in the south and the inhibition of star formation in the north may be the result of complex dynamical mechanisms, such as tidal torques, stretching

motions, converging flows, ram pressure, and shear (e.g. Renaud et al. 2009; Smith et al. 2013; Ploeckinger et al. 2015, 2018). At any rate, our results support a scenario where dynamical effects associated with the interaction, and not a lack of GMCs in the north, explain why star formation is limited to the south of the TDG.

## 6. Summary and conclusions

We have presented new ALMA observations of CO(2–1) emission in J1023+1952, a TDG in the system Arp 94, with  $0.64'' \sim 45$  pc resolution. To our knowledge, these are the highest resolution observations of molecular gas in a TDG so far, and the first time that individual GMCs are resolved in a TDG. Our main findings can be summarised as follows:

1. We find an extremely high fraction of extended molecular emission ( $\sim 80\text{--}90\%$ ), which is filtered out by the ALMA interferometer but recovered by the total power antennas. This emission is arising from scales larger than  $\sim 2$  kpc and is likely tracing diffuse molecular gas. High fractions of extended molecular emission have been found in nearby galaxies, but this is clearly at the upper end of what has been observed so far.
2. For our assumed distance (14.5 Mpc) and  $\alpha_{\text{CO}}$  conversion factor ( $\alpha_{\text{CO}}^{2-1} = 4.4 M_{\odot} (\text{K km s}^{-1} \text{ pc}^2)^{-1}$ ), when measured at matched  $6.3''$  resolution, the total molecular gas mass in the TDG is  $8.6 \times 10^7 M_{\odot}$ , very similar to the total amount of atomic gas ( $8.4 \times 10^7 M_{\odot}$ ). This results in a very high molecular-to-atomic gas ratio ( $\sim 1$ ), with significant variations between north ( $\sim 1.5$ ) and south ( $\sim 0.4$ ).
3. The global CO(2–1)/CO(1–0) ratio is quite low ( $R_{21} = 0.52$ ), at the lower end of measurements in nearby galaxies, and this becomes particularly extreme towards the south of the TDG ( $R_{21} = 0.38$ ). For the western part of NGC 3227 covered by our ALMA setup, the average  $R_{21} = 0.53$  is similar to the TDG, suggesting that the low  $R_{21}$  value could be related to the interaction.
4. We identified 111 GMCs in the TDG with CPROPS. These GMCs are spread throughout the entire TDG, and not preferentially concentrated in the star-forming part. They show similar sizes, masses, and linewidths as GMCs in the Milky Way and in other nearby galaxies ( $R \sim 10\text{--}100$  pc,  $M_{\text{lum}} \sim 10^4\text{--}10^6 M_{\odot}$ , and  $\sigma_v$  of a few  $\text{km s}^{-1}$ ). The total molecular gas in GMCs adds up to  $3 \times 10^7 M_{\odot}$  (extrapolated to perfect sensitivity), which is 18% of the total molecular mass ( $1.63 \times 10^8 M_{\odot}$  in the ALMA cube at native resolution, including total power, and without any masks).
5. GMCs in the TDG follow a mass spectrum that is well described by a truncated power law with a slope of  $\gamma = -1.76 \pm 0.13$  and truncation mass of  $M_0 = 1.9 \times 10^6 M_{\odot}$ . These values are comparable to the findings in the Milky Way and other nearby galaxies, in spite of the very different conditions of the TDG.
6. GMCs in the TDG follow scaling relations with similarly strong degrees of correlation as in the Milky Way and other galaxies. The slope of the size-mass relation is in agreement with previous observations. However, the size-linewidth and virial relation in the TDG are super-linear and significantly steeper than observed elsewhere. The high velocity dispersion of clouds in the TDG, particularly extreme for the most massive clouds, can explain the observed super-linear relations. Virial masses almost always exceed luminous masses, often by an order of magnitude (especially in the north). If

this is not an observational artefact (e.g. due to changes in  $\alpha_{\text{CO}}$ ), it could indicate that clouds are super-virial.

7. We find varying spatial offsets between young stellar clusters and GMCs in the star-forming part of the TDG. The smallest offsets ( $\lesssim 50$  pc) correspond to the youngest star-forming knots, while the largest offsets ( $\gtrsim 200$  pc) are associated with the oldest knots, consistent with an age trend.

In conclusion, our observations highlight the complex organisation of molecular gas in a TDG, with a large reservoir of diffuse molecular gas and many clouds that are not forming stars. Since the main difference between north and south is kinematic, we suggest that clouds in the north are stabilised against collapse by dynamical effects. GMCs in this system have a similar mass distribution and sizes as observed in other galaxies, but they show departures from the scaling relations involving velocity dispersion, and this is particularly acute in the quiescent north. On the other hand, the south of the TDG suggests that once the clouds begin to form stars, the process of star formation and subsequent feedback proceeds in a very similar fashion as in other galaxies. This process might be assisted by a mechanism such as stochastic self-propagating star formation.

*Acknowledgements.* We appreciate helpful comments from Bruce Elmegreen and Andreas Schruba. MQ and SGB acknowledge support from the research project PID2019-106027GA-C44 from the Spanish Ministerio de Ciencia e Innovación. UL acknowledges support by the research project AYA2017-84897-P from the Spanish Ministerio de Economía y Competitividad, from the European Regional Development Funds (FEDER), and the Junta de Andalucía (Spain) grants FQM108. FB acknowledges funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 726384/EMPIRE). SGB acknowledges support through Grants PGC2018-094671-B-I00 and AYA2016-76682-C3-2-P (MCIU/AEI/FEDER,UE). JMDK gratefully acknowledges funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through an Emmy Noether Research Group (grant number KR4801/1-1) and the DFG Sachbeihilfe (Grant number KR4801/2-1), as well as from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme via the ERC Starting Grant MUSTANG (Grant agreement number 714907).

## References

- Appleton, P. N., Mundell, C., Bitsakis, T., et al. 2014, *ApJ*, **797**, 117
- Batiste, M., Bentz, M. C., Manne-Nicholas, E. R., Onken, C. A., & Bershad, M. A. 2017, *ApJ*, **835**, 271
- Battersby, C., Keto, E., Walker, D., et al. 2020, *ApJS*, **249**, 35
- Bigiel, F., Leroy, A., Walter, F., et al. 2008, *AJ*, **136**, 2846
- Bigiel, F., Leroy, A. K., Jiménez-Donaire, M. J., et al. 2016, *ApJ*, **822**, L26
- Bolatto, A. D., Leroy, A. K., Rosolowsky, E., Walter, F., & Blitz, L. 2008, *ApJ*, **686**, 948
- Bournaud, F., & Duc, P. A. 2006, *A&A*, **456**, 481
- Braine, J., Lisenfeld, U., Due, P.-A., & Leon, S. 2000, *Nature*, **403**, 867
- Braine, J., Duc, P. A., Lisenfeld, U., et al. 2001, *A&A*, **378**, 51
- Braine, J., Rosolowsky, E., Gratier, P., Corbelli, E., & Schuster, K. F. 2018, *A&A*, **612**, A51
- Caldú-Primo, A., & Schruba, A. 2016, *AJ*, **151**, 34
- Caldú-Primo, A., Schruba, A., Walter, F., et al. 2015, *AJ*, **149**, 76
- Chen, C.-H. R., Indebetouw, R., Chu, Y.-H., et al. 2010, *ApJ*, **721**, 1206
- Chevanze, M., Kruijssen, J. M. D., Hygate, A. P. S., et al. 2020, *MNRAS*, **493**, 2872
- Colombo, D., Hughes, A., Schinnerer, E., et al. 2014, *ApJ*, **784**, 3
- Colombo, D., Rosolowsky, E., Ginsburg, A., Duarte-Cabral, A., & Hughes, A. 2015, *MNRAS*, **454**, 2067
- Colombo, D., Rosolowsky, E., Duarte-Cabral, A., et al. 2019, *MNRAS*, **483**, 4291
- Condon, J. J., & Ransom, S. M. 2016, *Essential Radio Astronomy* (Princeton, NJ: Princeton University Press)
- Corbelli, E., Braine, J., Bandiera, R., et al. 2017, *A&A*, **601**, A146
- Dame, T. M., & Thaddeus, P. 1994, *ApJ*, **436**, L173
- den Brok, J. S., Chatzigiannakis, D., & Bigiel, F. 2020, *MNRAS*, submitted
- Dobbs, C. L., Pringle, J. E., & Duarte-Cabral, A. 2015, *MNRAS*, **446**, 3608

- Dopita, M. A., Mathewson, D. S., & Ford, V. L. 1985, *ApJ*, 297, 599
- Duc, P.-A. 2012, *Astrophys. Space Sci. Proc.*, 28, 305
- Duc, P.-A., Brinks, E., Springel, V., et al. 2000, *AJ*, 120, 1238
- Duc, P. A., Braine, J., Lisenfeld, U., Brinks, E., & Boquien, M. 2007, *A&A*, 475, 187
- Duc, P.-A., Paudel, S., McDermid, R. M., et al. 2014, *MNRAS*, 440, 1458
- Elmegreen, B. G. 2018, *ApJ*, 854, 16
- Faesi, C. M., Lada, C. J., & Forbrich, J. 2018, *ApJ*, 857, 19
- Feitzinger, J. V., Glassgold, A. E., Gerola, H., & Seiden, P. E. 1981, *A&A*, 98, 371
- Fensch, J., Duc, P.-A., Weilbacher, P. M., Boquien, M., & Zackrisson, E. 2016, *A&A*, 585, A79
- Fensch, J., Duc, P.-A., Boquien, M., et al. 2019, *A&A*, 628, A60
- Fujimoto, Y., Tasker, E. J., & Habe, A. 2014, *MNRAS*, 445, L65
- Fukui, Y., Mizuno, N., Yamaguchi, R., et al. 1999, *PASJ*, 51, 745
- Gaia Collaboration (Prusti, T., et al.) 2016, *A&A*, 595, A1
- Gaia Collaboration (Brown, A. G. A., et al.) 2018, *A&A*, 616, A1
- Gallagher, M. J., Leroy, A. K., Bigiel, F., et al. 2018, *ApJ*, 858, 90
- García-Burillo, S., Guelin, M., Cernicharo, J., & Dahlem, M. 1992, *A&A*, 266, 21
- Gerola, H., & Seiden, P. E. 1978, *ApJ*, 223, 129
- Grasha, K., Calzetti, D., Bittler, L., et al. 2018, *MNRAS*, 481, 1016
- Grasha, K., Calzetti, D., Adamo, A., et al. 2019, *MNRAS*, 483, 4707
- Gratier, P., Braine, J., Rodríguez-Fernández, N. J., et al. 2012, *A&A*, 542, A108
- Herrera, C. N., Pety, J., Hughes, A., et al. 2020, *A&A*, 634, A121
- Heyer, M., & Dame, T. M. 2015, *ARA&A*, 53, 583
- Heyer, M. H., Carpenter, J. M., & Snell, R. L. 2001, *ApJ*, 551, 852
- Heyer, M., Krawczyk, C., Duval, J., & Jackson, J. M. 2009, *ApJ*, 699, 1092
- Hughes, A., Wong, T., Ott, J., et al. 2010, *MNRAS*, 406, 2065
- Hughes, A., Meidt, S. E., Schinnerer, E., et al. 2013a, *ApJ*, 779, 44
- Hughes, A., Meidt, S. E., Colombo, D., et al. 2013b, *ApJ*, 779, 46
- Inutsuka, S.-I., Inoue, T., Iwasaki, K., & Hosokawa, T. 2015, *A&A*, 580, A49
- Kauffmann, J., Pillai, T., Zhang, Q., et al. 2017, *A&A*, 603, A89
- Kawamura, A., Mizuno, Y., Minamidani, T., et al. 2009, *ApJS*, 184, 1
- Kennicutt, R. C., Jr 1998, *ApJ*, 498, 541
- Kobayashi, M. I. N., Inutsuka, S.-I., Kobayashi, H., & Hasegawa, K. 2017, *ApJ*, 836, 175
- Kreckel, K., Faesi, C., Kruijssen, J. M. D., et al. 2018, *ApJ*, 863, L21
- Kritsuk, A. G., Ustyugov, S. D., & Norman, M. L. 2017, *New J. Phys.*, 19, 065003
- Kruijssen, J. M. D., & Longmore, S. N. 2014, *MNRAS*, 439, 3239
- Kruijssen, J. M. D., Longmore, S. N., Elmegreen, B. G., et al. 2014, *MNRAS*, 440, 3370
- Kruijssen, J. M. D., Schrubba, A., Hygate, A. P. S., et al. 2018, *MNRAS*, 479, 1866
- Kruijssen, J. M. D., Schrubba, A., Chevance, M., et al. 2019, *Nature*, 569, 519
- Larson, R. B. 1981, *MNRAS*, 194, 809
- Lelli, F., Duc, P.-A., Brinks, E., et al. 2015, *A&A*, 584, A113
- Lelli, F., McGaugh, S. S., Schombert, J. M., Desmond, H., & Katz, H. 2019, *MNRAS*, 484, 3267
- Leroy, A. K., Walter, F., Bigiel, F., et al. 2009, *AJ*, 137, 4670
- Leroy, A. K., Walter, F., Sandstrom, K., et al. 2013, *AJ*, 146, 19
- Lisenfeld, U., Mundell, C. G., Schinnerer, E., Appleton, P. N., & Allsopp, J. 2008, *ApJ*, 685, 181
- Lisenfeld, U., Braine, J., Duc, P. A., et al. 2016, *A&A*, 590, A92
- Lombardi, M., Alves, J., & Lada, C. J. 2010, *A&A*, 519, L7
- Longmore, S. N., Bally, J., Testi, L., et al. 2013, *MNRAS*, 429, 987
- Luna, A., Bronfman, L., Carrasco, L., & May, J. 2006, *ApJ*, 641, 938
- Maeda, F., Ohta, K., Fujimoto, Y., Habe, A., & Ushio, K. 2020a, *MNRAS*, 495, 3840
- Maeda, F., Ohta, K., Fujimoto, Y., & Habe, A. 2020b, *MNRAS*, 493, 5045
- McCray, R., & Kafatos, M. 1987, *ApJ*, 317, 190
- Meidt, S. E., Schinnerer, E., García-Burillo, S., et al. 2013, *ApJ*, 779, 45
- Meidt, S. E., Leroy, A. K., Rosolowsky, E., et al. 2018, *ApJ*, 854, 100
- Miura, R. E., Kohno, K., Tosaki, T., et al. 2012, *ApJ*, 761, 37
- Mueller, M. W., & Arnett, W. D. 1976, *ApJ*, 210, 670
- Mundell, C. G., Pedlar, A., Axon, D. J., Meaburn, J., & Unger, S. W. 1995a, *MNRAS*, 277, 641
- Mundell, C. G., Holloway, A. J., Pedlar, A., et al. 1995b, *MNRAS*, 275, 67
- Mundell, C. G., James, P. A., Loiseau, N., Schinnerer, E., & Forbes, D. A. 2004, *ApJ*, 614, 648
- Noda, H., Makishima, K., Yamada, S., et al. 2014, *ApJ*, 794, 2
- Nomura, H., & Kamaya, H. 2001, *AJ*, 121, 1024
- Oka, T., Hasegawa, T., Handa, T., Hayashi, M., & Sakamoto, S. 1996, *ApJ*, 460, 334
- Pan, H.-A., & Kuno, N. 2017, *ApJ*, 839, 133
- Peñaloza, C. H., Clark, P. C., Glover, S. C. O., Shetty, R., & Klessen, R. S. 2017, *MNRAS*, 465, 2277
- Pety, J., Schinnerer, E., Leroy, A. K., et al. 2013, *ApJ*, 779, 43
- Ploeckinger, S., Recchi, S., Hensler, G., & Kroupa, P. 2015, *MNRAS*, 447, 2512
- Ploeckinger, S., Sharma, K., Schaye, J., et al. 2018, *MNRAS*, 474, 580
- Querejeta, M., Schinnerer, E., Schrubba, A., et al. 2019, *A&A*, 625, A19
- Rebolledo, D., Wong, T., Xue, R., et al. 2015, *ApJ*, 808, 99
- Reina-Campos, M., & Kruijssen, J. M. D. 2017, *MNRAS*, 469, 1282
- Renaud, F., Boily, C. M., Naab, T., & Theis, C. 2009, *ApJ*, 706, 67
- Renaud, F., Bournaud, F., Emsellem, E., et al. 2015, *MNRAS*, 454, 3299
- Rice, T. S., Goodman, A. A., Bergin, E. A., Beaumont, C., & Dame, T. M. 2016, *ApJ*, 822, 52
- Richer, M. G., & McCall, M. L. 1995, *ApJ*, 445, 642
- Rosolowsky, E. 2005, *PASP*, 117, 1403
- Rosolowsky, E. 2007, *ApJ*, 654, 240
- Rosolowsky, E., & Leroy, A. 2006, *PASP*, 118, 590
- Rosolowsky, E. W., Pineda, J. E., Kauffmann, J., & Goodman, A. A. 2008, *ApJ*, 679, 1338
- Sakamoto, S. 1994, *PASP*, 106, 1112
- Sandstrom, K. M., Leroy, A. K., Walter, F., et al. 2013, *ApJ*, 777, 5
- Schaye, J., & Dalla Vecchia, C. 2008, *MNRAS*, 383, 1210
- Schinnerer, E., Eckart, A., & Tacconi, L. J. 2001, *ApJ*, 549, 254
- Schmidt, M. 1959, *ApJ*, 129, 243
- Schombert, J., McGaugh, S., & Lelli, F. 2020, *AJ*, 160, 71
- Schruba, A., Leroy, A. K., Walter, F., et al. 2011, *AJ*, 142, 37
- Schruba, A., Kruijssen, J. M. D., & Leroy, A. K. 2019, *ApJ*, 883, 2
- Semenov, V. A., Kravtsov, A. V., & Gnedin, N. Y. 2018, *ApJ*, 861, 4
- Shetty, R., Beaumont, C. N., Burton, M. G., Kelly, B. C., & Klessen, R. S. 2012, *MNRAS*, 425, 720
- Sleath, J. P., & Alexander, P. 1995, *MNRAS*, 275, 507
- Smith, R., Duc, P. A., Candlish, G. N., et al. 2013, *MNRAS*, 436, 839
- Solomon, P. M., Rivolo, A. R., Barrett, J., & Yahil, A. 1987, *ApJ*, 319, 730
- Sun, J., Leroy, A. K., Ostriker, E. C., et al. 2020, *ApJ*, 892, 148
- Tan, J. C. 2000, *ApJ*, 536, 173
- Torii, K., Hasegawa, K., Hattori, Y., et al. 2015, *ApJ*, 806, 7
- Tosaki, T., Kohno, K., Harada, N., et al. 2017, *PASJ*, 69, 18
- Tully, R. B. 1988, *Nearby Galaxies Catalog* (Cambridge: University Press)
- Usero, A., Leroy, A. K., Walter, F., et al. 2015, *AJ*, 150, 115
- Utomo, D., Blitz, L., Davis, T., et al. 2015, *ApJ*, 803, 16
- Whitmore, B. C., Brogan, C., Chandar, R., et al. 2014, *ApJ*, 795, 156
- Wong, T., Hughes, A., Ott, J., et al. 2011, *ApJS*, 197, 16
- Yoda, T., Handa, T., Kohno, K., et al. 2010, *PASJ*, 62, 1277
- Yoshii, Y., Kobayashi, Y., Minezaki, T., Koshida, S., & Peterson, B. A. 2014, *ApJ*, 784, L11
- Zhao, Y., Gao, Y., & Gu, Q. 2010, *ApJ*, 710, 663

## Appendix A: Alternative identification of GMCs with SCIMES

We applied SCIMES (Colombo et al. 2015) to our data as an alternative approach to identify GMCs. SCIMES is based on dendrograms, tree diagrams that reflect the hierarchical structure of a dataset Rosolowsky et al. (2008). A dendrogram tree is made up of trunks (top hierarchical level), and substructures such as branches (intermediate level) and leaves (lowest level): a branch can split into multiple leaves. The goal of applying SCIMES to our data is two-fold: we want to check if our results are sensitive to the specific choice of code and examine the degree of hierarchy within the molecular medium.

### A.1. GMC segmentation with SCIMES

SCIMES relies on a graph-based clustering technique (spectral clustering) to optimally segment the structure tree based exclusively on the input data. The method is tuned to the segmentation of the molecular ISM by employing the luminosity and volume of the isosurfaces to establish the similarity matrices that are key to the spectral clustering.

SCIMES has been used to isolate molecular clouds and cloud substructures in Galactic data (e.g. filaments or cores). For extragalactic data, however, given the limited spatial resolution, leaves typically map to entire GMCs while branches can correspond to larger structures, such as groups of GMCs, that are physically connected.

### A.2. Results with SCIMES

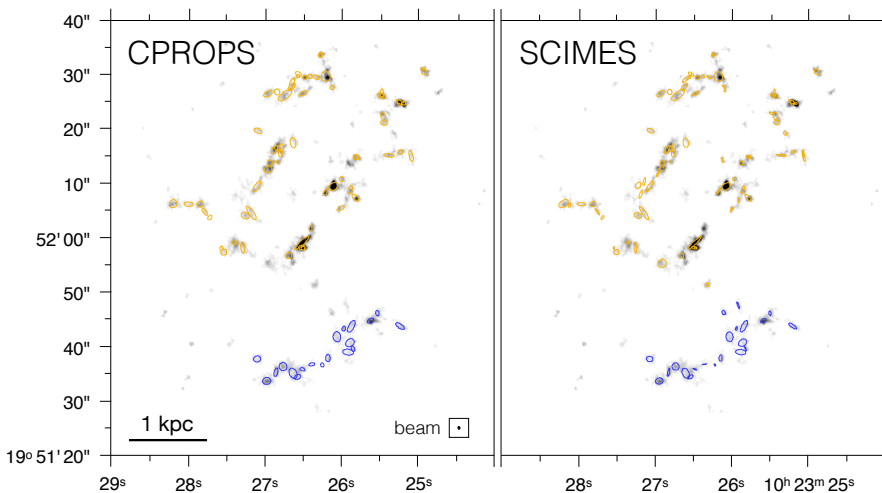
The results from this section confirm that GMC properties do not look very different when identified with CPROPS or SCIMES. Fig. A.1 shows that, even though the segmentation is not identical on a cloud-by-cloud basis, the overall distribution of GMCs is similar, and clouds also show similar sizes and axis ratios. This is further emphasised by Table A.1, where the properties of

the “leaves” can be contrasted against the typical properties of GMCs from CPROPS as listed in Table 3.

As an ensemble CPROPS and SCIMES yield cloud populations with similar properties, with a small trend for SCIMES to recover slightly more luminous and massive clouds. There are also small differences in the GMC mass spectrum: the slope of the global truncated power-law fit is slightly higher with SCIMES,  $\gamma = -1.84 \pm 0.15$  (instead of  $\gamma = -1.76 \pm 0.13$ ), but they are well compatible within the uncertainties. The mass spectrum in the north also remains steeper with SCIMES ( $\gamma = -1.74 \pm 0.16$  vs  $\gamma = -1.57 \pm 0.37$ ), but the difference between north and south becomes even less significant than with CPROPS.

The distribution of molecular mass in GMCs between north and south is also comparable to what we found with CPROPS:  $2.9 \times 10^7 M_{\odot}$  in the north and  $6.8 \times 10^6 M_{\odot}$  in the south (as opposed to  $2.4 \times 10^7 M_{\odot}$  and  $5.8 \times 10^6 M_{\odot}$  with CPROPS). The GMCs identified with SCIMES also behave very similarly in terms of the scaling relations that we examined in Sect. 4.3.3. Table A.2 shows the Spearman rank correlation coefficients and the best power-law fits for the GMC scaling relations with SCIMES. The correlation coefficients are generally similar to the ones implied by CPROPS, and the fitted relations also tend to agree with the results from CPROPS within the error bars; the strongest departures take place in cases where the uncertainty of the fits was already large (e.g. size-linewidth relation for the south). The super-linear nature of the virial relation persists with SCIMES.

The application of SCIMES to our data has revealed some hierarchical clustering in the molecular gas of the TDG. SCIMES identified a total of 127 leaves, 49 branches, and 26 trunks in the TDG. On average, each trunk has 3.9 leaves pending directly from it and 1.9 branches (but some of these branches are indirect, in the sense that a branch can split into other branches). Each branch, in turn, hosts 3 leaves on average. There is a total of 26 leaves that do not have a direct parent.



**Fig. A.1.** GMC positions and orientations for CPROPS (left panel) compared to SCIMES (right panel) indicated on the ALMA CO(2–1) integrated intensity map as in Fig. 8. GMCs are shown as ellipses with the extrapolated and deconvolved major and minor axes (2nd moment of emission). Orange and blue ellipses represent clouds in the north and south of the TDG, respectively. The bottom-right corner of the left panel shows the ALMA synthesised beam ( $0.69'' \times 0.60''$  with  $PA = 11^\circ$ ).

In the Milky Way, where much higher spatial resolutions can be achieved, GMCs have been proven to consist of hierarchical substructures (filaments, clumps, etc.), but we cannot resolve such intra-GMC substructures here. It can happen, conversely, that the GMCs are not fully isolated from each other and are actually embedded in larger macrostructures. SCIMES suggests

that this is the case to some extent, but not to the extreme of all GMCs in the TDG being connected up to a single “tree”. The application of SCIMES to data in external galaxies is so far limited and the interpretation of these hierarchical macrostructures is not straight-forward when the synthesised beam is comparable to an entire GMCs.

**Table A.1.** Properties of structures derived by SCIMES.

Property	Unit	TDG			North			South		
		Min	Median	Max	Min	Median	Max	Min	Median	Max
<i>Leaves</i>										
$T_{\max}$	K	0.3	0.4	1.1	0.3	0.4	0.9	0.3	0.5	1.1
$L_{\text{CO}}$	$\text{K km s}^{-1} \text{ pc}^2$	$5.2 \times 10^3$	$4.1 \times 10^4$	$5.0 \times 10^5$	$5.2 \times 10^3$	$4.2 \times 10^4$	$5.0 \times 10^5$	$6.5 \times 10^3$	$4.1 \times 10^4$	$1.8 \times 10^5$
$R$	pc	14.6	51.4	122.8	14.6	51.3	122.8	18.0	55.8	102.5
$\sigma_v$	$\text{km s}^{-1}$	1.0	6.1	19.7	1.0	6.5	19.7	1.2	4.7	9.8
$M_{\text{lum}}$	$M_{\odot}$	$2.3 \times 10^4$	$1.8 \times 10^5$	$2.2 \times 10^6$	$2.3 \times 10^4$	$1.9 \times 10^5$	$2.2 \times 10^6$	$2.9 \times 10^4$	$1.8 \times 10^5$	$7.8 \times 10^5$
$M_{\text{vir}}$	$M_{\odot}$	$2.8 \times 10^4$	$2.4 \times 10^6$	$4.1 \times 10^7$	$5.9 \times 10^4$	$2.7 \times 10^6$	$4.1 \times 10^7$	$2.8 \times 10^4$	$2.2 \times 10^6$	$7.6 \times 10^6$
<i>Branches</i>										
$T_{\max}$	K	0.4	0.6	1.1	0.4	0.5	0.8	0.5	0.6	1.1
$L_{\text{CO}}$	$\text{K km s}^{-1} \text{ pc}^2$	$2.5 \times 10^4$	$2.2 \times 10^5$	$7.7 \times 10^5$	$2.5 \times 10^4$	$1.9 \times 10^5$	$7.7 \times 10^5$	$1.1 \times 10^5$	$2.9 \times 10^5$	$6.1 \times 10^5$
$R$	pc	14.7	97.9	310.2	14.7	95.7	181.4	81.7	153.2	310.2
$\sigma_v$	$\text{km s}^{-1}$	4.3	12.6	26.2	4.3	12.6	26.2	4.3	10.6	13.5
$M_{\text{lum}}$	$M_{\odot}$	$1.1 \times 10^5$	$9.7 \times 10^5$	$3.4 \times 10^6$	$1.1 \times 10^5$	$8.5 \times 10^5$	$3.4 \times 10^6$	$4.9 \times 10^5$	$1.3 \times 10^6$	$2.7 \times 10^6$
$M_{\text{vir}}$	$M_{\odot}$	$3.2 \times 10^5$	$1.6 \times 10^7$	$1.3 \times 10^8$	$3.2 \times 10^5$	$1.6 \times 10^7$	$1.3 \times 10^8$	$2.5 \times 10^6$	$1.3 \times 10^7$	$5.3 \times 10^7$
<i>Trunks</i>										
$T_{\max}$	K	0.3	0.6	1.1	0.3	0.5	0.9	0.5	0.7	1.1
$L_{\text{CO}}$	$\text{K km s}^{-1} \text{ pc}^2$	$4.3 \times 10^4$	$2.0 \times 10^5$	$1.1 \times 10^6$	$4.3 \times 10^4$	$2.0 \times 10^5$	$1.1 \times 10^6$	$5.7 \times 10^4$	$5.9 \times 10^5$	$6.2 \times 10^5$
$R$	pc	36.3	108.1	307.9	36.3	108.0	208.1	76.6	211.6	307.9
$\sigma_v$	$\text{km s}^{-1}$	4.3	12.8	29.4	7.7	13.4	29.4	4.3	12.8	14.3
$M_{\text{lum}}$	$M_{\odot}$	$1.9 \times 10^5$	$8.8 \times 10^5$	$4.9 \times 10^6$	$1.9 \times 10^5$	$8.8 \times 10^5$	$4.9 \times 10^6$	$2.5 \times 10^5$	$2.6 \times 10^6$	$2.7 \times 10^6$
$M_{\text{vir}}$	$M_{\odot}$	$1.5 \times 10^6$	$2.1 \times 10^7$	$1.9 \times 10^8$	$3.8 \times 10^6$	$2.1 \times 10^7$	$1.9 \times 10^8$	$1.5 \times 10^6$	$4.5 \times 10^7$	$5.2 \times 10^7$

**Table A.2.** Fits and rank coefficients for GMC scaling relations with SCIMES.

Relation	Method	TDG		North		South		MW Slope
		Slope	$\rho$	Slope	$\rho$	Slope	$\rho$	
Size-linewidth	ODR	$1.40 \pm 0.20$	0.46 (0.00)	$1.41 \pm 0.20$	0.57 (0.00)	$0.03 \pm 0.14$	0.07 (0.77)	0.5 <sup>(a)</sup>
	Bayesian	$1.27^{+0.53}_{-0.36}$		$1.33^{+0.45}_{-0.33}$		$0.01^{+0.27}_{-0.32}$		
Virial relation	ODR	$1.51 \pm 0.11$	0.80 (0.00)	$1.42 \pm 0.11$	0.83 (0.00)	$1.41 \pm 0.35$	0.66 (0.00)	0.8 <sup>(a)</sup>
	Bayesian	$1.42^{+0.19}_{-0.16}$		$1.35^{+0.21}_{-0.17}$		$1.20^{+3.39}_{-1.49}$		
Mass-size	ODR	$2.02 \pm 0.16$	0.71 (0.00)	$2.22 \pm 0.21$	0.70 (0.00)	$1.21 \pm 0.20$	0.76 (0.00)	2.0 <sup>(a)</sup>
	Bayesian	$3.08^{+0.48}_{-0.37}$		$3.54^{+0.81}_{-0.57}$		$2.22^{+0.80}_{-0.58}$		
Heyer plot	ODR	$0.72 \pm 0.08$	0.54 (0.00)	$0.59 \pm 0.08$	0.52 (0.00)	$0.72 \pm 0.16$	0.58 (0.01)	0.5 <sup>(b)</sup>
	Bayesian	$0.37^{+0.08}_{-0.08}$		$0.30^{+0.08}_{-0.08}$		$0.45^{+0.25}_{-0.18}$		

**Notes.**  $\rho$  is the Spearman rank correlation coefficient, with the  $p$ -value indicated in parentheses. For each dataset we used two different fitting methods: orthogonal distance regression (ODR) and a Bayesian code, BayesLineFit. <sup>(a)</sup>Solomon et al. (1987). <sup>(b)</sup>Heyer et al. (2009).

**Table A.3.** GMC catalogue from CPROPS.

Cloud ID	RA <sub>J2000</sub> (deg)	Dec <sub>J2000</sub> (deg)	(a/b)	PA (deg)	$v_{\text{cen}}$ ( $\text{km s}^{-1}$ )	$R$ (pc)	$\sigma_v$ ( $\text{km s}^{-1}$ )	$T_{\text{peak}}$ (K)	$L_{\text{CO}}$ ( $\text{K km s}^{-1} \text{ pc}^2$ )	$M_{\text{lum}}$ ( $10^3 M_{\odot}$ )	$M_{\text{vir}}$ ( $10^3 M_{\odot}$ )	Environment
1	155.8609	19.8751	3.58	33.0	1039.9		5.6 ± 3.8	0.394	12710 ± 9534	56 ± 42		N
2	155.8606	19.8750	1.77	146.5	1044.6	45.7 ± 23.3	4.4 ± 2.1	0.431	28853 ± 10756	127 ± 47	931 ± 987	N
3	155.8610	19.8751	3.68	87.8	1048.8		1.7 ± 2.5	0.394	9822 ± 8829	43 ± 39		N
4	155.8606	19.8749	1.09	69.3	1063.0	12.3 ± 20.3	4.6 ± 1.3	0.396	22143 ± 6205	97 ± 27	266 ± 478	N
5	155.8625	19.8702	1.90	52.0	1066.6		9.1 ± 3.6	0.394	31495 ± 11951	139 ± 53		N
6	155.8619	19.8712	2.05	58.0	1077.5	69.2 ± 15.2	15.4 ± 2.2	0.547	180161 ± 25269	793 ± 111	17012 ± 6101	N
7	155.8622	19.8705	1.13	60.9	1089.5	60.0 ± 16.4	8.9 ± 1.9	0.582	107533 ± 15746	473 ± 69	4892 ± 2607	N
8	155.8623	19.8702	1.22	39.6	1135.3	79.6 ± 13.5	12.6 ± 1.5	0.455	137245 ± 12763	604 ± 56	13120 ± 4050	N
9	155.8612	19.8749	1.73	29.6	1067.2		1.6 ± 2.4	0.375	18177 ± 9285	80 ± 41		N
...	...	...	...	...	...	...	...	...	...	...	...	...

**Notes.** The first ten rows are shown for guidance on the format. The full table is available at the CDS.