Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **504,** 3221–3245 (2021) doi:10.1093/mnras/stab859
Advance Access publication 2021 March 31

New constraints on the ^{12}CO(2–1)/(1–0) line ratio across nearby disc galaxies

J. S. den Brok [1]★ D. Chatzigiannakis,[1] F. Bigiel,[1] J. Puschnig [1] A. T. Barnes [1] A. K. Leroy,[2] M. J. Jiménez-Donaire,[3,4] A. Usero,[3] E. Schinnerer,[5] E. Rosolowsky [6] C. M. Faesi,[7] K. Grasha [8] A. Hughes [9,10] J. M. D. Kruijssen [11] D. Liu,[5] L. Neumann,[1] J. Pety,[12,13] M. Querejeta,[3] T. Saito,[5] A. Schruba[14] and S. Stuber [5]

[1]*Argelander-Institut für Astronomie, Universität Bonn, Auf dem Hügel 71, D-53121 Bonn, Germany*
[2]*Department of Astronomy, The Ohio State University, 4055 McPherson Laboratory, 140 West 18th Avenue, Columbus, OH 43210, USA*
[3]*Observatorio Astronómico Nacional (IGN), C/Alfonso XII 3, E-28014 Madrid, Spain*
[4]*Centro de Desarrollos Tecnológicos, Observatorio de Yebes (IGN), E-19141 Yebes, Guadalajara, Spain*
[5]*Max Planck Institute for Astronomy, Königstuhl 17, D-69117 Heidelberg, Germany*
[6]*University of Alberta, 4-183 CCIS, Edmonton AB T6G 2E1, Alberta, Canada*
[7]*Department of Astronomy, University of Massachusetts Amherst, 710 N. Pleasant Street, Amherst, MA 01003, USA*
[8]*Research School of Astronomy and Astrophysics, Australian National University, Canberra, ACT 2611, Australia*
[9]*Université de Toulouse, UPS-OMP, F-31028 Toulouse, France*
[10]*CNRS, IRAP, Av. du Colonel Roche BP 44346, F-31028 Toulouse cedex 4, France*
[11]*Astronomisches Rechen-Institut, Zentrum für Astronomie der Universität Heidelberg, Mönchhofstraße 12-14, D-69120 Heidelberg, Germany*
[12]*IRAM, 300 rue de la Piscine, F-38406 Saint Martin d'Hères, France*
[13]*Sorbonne Université, Observatoire de Paris, Université PSL, École normale supérieure, CNRS, LERMA, F-75005 Paris, France*
[14]*Max-Planck Institut für Extraterrestrische Physik, Giessenbachstraße 1, D-85748 Garching, Germany*

Accepted 2021 March 12. Received 2021 March 9; in original form 2020 May 6

ABSTRACT

Both the CO(2–1) and CO(1–0) lines are used to trace the mass of molecular gas in galaxies. Translating the molecular gas mass estimates between studies using different lines requires a good understanding of the behaviour of the CO(2–1)-to-CO(1–0) ratio, R_{21}. We compare new, high-quality CO(1–0) data from the IRAM 30-m EMIR MultiLine Probe of the ISM Regulating Galaxy Evolution survey to the latest available CO(2–1) maps from HERA CO-Line Extragalactic Survey, Physics at High Angular resolution in Nearby Galaxies-ALMA, and a new IRAM 30-m M51 Large Program. This allows us to measure R_{21} across the full star-forming disc of nine nearby, massive, star-forming spiral galaxies at 27 arcsec (∼1–2 kpc) resolution. We find an average $R_{21} = 0.64 \pm 0.09$ when we take the luminosity-weighted mean of all individual galaxies. This result is consistent with the mean ratio for disc galaxies that we derive from single-pointing measurements in the literature, $R_{21,\mathrm{lit}} = 0.59^{+0.18}_{-0.09}$. The ratio shows weak radial variations compared to the point-to-point scatter in the data. In six out of nine targets, the central enhancement in R_{21} with respect to the galaxy-wide mean is of order of ∼10−20 per cent. We estimate an azimuthal scatter of ∼20 per cent in R_{21} at fixed galactocentric radius but this measurement is limited by our comparatively coarse resolution of 1.5 kpc. We find mild correlations between R_{21} and carbon monoxide (CO) brightness temperature, infrared (IR) intensity, 70–160 μm ratio, and IR-to-CO ratio. All correlations indicate that R_{21} increases with gas surface density, star formation rate surface density, and the interstellar radiation field.

Key words: ISM: molecules – galaxies: ISM – radio lines: galaxies.

1 INTRODUCTION

Carbon monoxide (CO) is the most abundant molecule in the interstellar medium (ISM) after molecular hydrogen (H_2). Unlike H_2, CO has a permanent dipole moment and its rotational transitions can be excited at low temperatures. The two lowest rotational transitions of the main CO molecule, $^{12}C^{16}O$ $J = 1 \rightarrow 0$, hereafter CO(1–0), and $^{12}C^{16}O$ $J = 2 \rightarrow 1$, hereafter CO(2–1), are among the brightest millimetre-wave spectral lines emitted by galaxies. They have critical densities of $n_{\mathrm{crit},1-0} \sim 2000$ cm^{-3} and $n_{\mathrm{crit},2-1} \sim 10\,000$ cm^{-3} for a fully molecular gas with a temperature of $T = 10$ K and optically thin transitions. Given typical optical depths for CO(1–0) of $\tau \sim$ 5–10, line trapping effects lower the effective critical density even further, to ∼100–1000 cm^{-3}. This is comparable to the mean density of molecular gas in galaxies (for more, see reviews by Bolatto, Wolfire & Leroy 2013; Heyer & Dame 2015; Shirley 2015). As a result of their brightness, low excitation requirement, and locations at favourable frequencies for observations from the ground, both transitions are often used to trace the mass of molecular gas in galaxies.

★ E-mail: jdenbrok@astro.uni-bonn.de

ALMA, NOEMA, and other mm-wave facilities now regularly map both CO(2–1) and CO(1–0) line emission across large areas and large samples of galaxies. It is increasingly important to be able to quantitatively compare results obtained using these different lines. Physically, the CO(2–1)-to-CO(1–0) line ratio, R_{21}, should depend on the temperature and density of the gas and on the optical depths of the lines (see e.g. Sakamoto et al. 1994, 1997; Peñaloza et al. 2017, 2018). Thus, understanding how R_{21} varies in response to the local environment also has the prospect to provide information regarding the physical conditions of the molecular gas.

The R_{21} ratio has been studied in both the Milky Way (e.g. Hasegawa 1997; Hasegawa et al. 1997; Sakamoto et al. 1997; Sawada et al. 2001; Yoda et al. 2010) and nearby galaxies (e.g. Eckart et al. 1990; Casoli et al. 1991a; Lundgren et al. 2004; Crosthwaite & Turner 2007; Leroy et al. 2009, 2013; Koda et al. 2012, 2020; Druard et al. 2014; Saintonge et al. 2017; Law et al. 2018; Yajima et al. 2021). Milky Way studies highlight a correlation between the R_{21} ratio and density, with R_{21} dropping with decreasing gas density from the centres to the edges of molecular clouds (e.g. Hasegawa 1997).

Studies of individual other galaxies often find higher R_{21} in the central kpc compared to the outer parts (e.g. Braine & Combes 1992; Leroy et al. 2009, 2013; Koda et al. 2020; Yajima et al. 2021). This radial behaviour could be explained if the average temperature and/or density of molecular gas drops with galactocentric radius. Independent evidence suggests that both temperature and density are often enhanced in galaxy centres (e.g. Mangum et al. 2013; Gallagher et al. 2018a; Sun et al. 2018; Jiménez-Donaire et al. 2019). Other work has focused on azimuthal variations in well-resolved galaxies with strong spiral arms, especially M51. There studies indicate enhanced excitation in the spiral arms and bar ends compared to the interarm regions (Koda et al. 2012, 2020; Vlahakis et al. 2013; Law et al. 2018).

However, our quantitative knowledge of how R_{21} varies across galaxies remains limited. Extensive CO(2–1) mapping has only been possible for ∼10 yr and there have been only a limited number of mapping surveys that cover both CO(1–0) and CO(2–1) in the same sample of galaxies. As a result, the magnitude of the observed variations in R_{21} remains fairly weak, with the typical range of values found in spiral galaxies spanning from 0.5 to 0.9 and often much less inside a single galaxy. This is easily within the range where even modest calibration uncertainties and heterogeneous data can obscure real astrophysical signal. Furthermore, much of the extragalactic mapping work has been confined to single-galaxy studies (e.g. Crosthwaite & Turner 2007; Koda et al. 2012, 2020; Vlahakis et al. 2013; Law et al. 2018).

In this paper, we present the full disc mapping of CO(1–0) and CO(2–1) observations. CO(1–0) is obtained with the IRAM 30-m telescope, from the EMIR MultiLine Probe of the ISM Regulating Galaxy Evolution (EMPIRE; Bigiel et al. 2016; Jiménez-Donaire et al. 2019) and for CO(2–1) we use the latest available data, selecting from the HERA CO-Line Extragalactic Survey (HERACLES; Leroy et al. 2009), the Physics at High Angular resolution in Nearby Galaxies (PHANGS) survey (Leroy et al. 2021b), or the IRAM 30-m M51 Large Program (den Brok et al., in preparation). Thus, each line is covered by a homogeneous, deep, wide-area mapping survey. Together they probe R_{21} across a sample of nine nearby spiral galaxies. Our main goals are to derive robust galaxy-wide mean values of the R_{21} ratio and to investigate how R_{21} varies systematically across the discs of these galaxies.

In Section 2, we present the data and define the physical quantities we use. Our analysis of the R_{21} ratio is presented in Section 3, where we examine the distribution of the ratio, its radial and azimuthal variations, and the possible correlations between R_{21} and physical properties such as CO brightness temperature and infrared (IR) emission. We discuss our results and compare them to results from previous observations in Section 4. We summarize our findings in Section 5.

Table 1. Galaxy sample.

Name	RA (J2000)	Dec. (J2000)	D (Mpc)	i (deg)	PA (deg)
NGC 0628	01:36:41.8	15:47:00	9.0	7	20
NGC 2903	09:32:10.1	21:30:03	8.5	65	204
NGC 3184	10:18:17.0	41:25:28	13.0	16	179
NGC 3627	11:20:15.0	12:59:30	9.4	62	173
NGC 4254	12:18:50.0	14:24:59	16.8	32	55
NGC 4321	12:22:55.0	15:49:19	15.2	30	153
NGC 5055	13:15:49.2	42:01:45	8.9	59	102
NGC 5194	13:29:52.7	47:11:43	8.4	20	172
NGC 6946	20:34:52.2	60:09:14	7.0	33	243

Note. Adopted from Jiménez-Donaire et al. (2019).

2 OBSERVATIONS

2.1 Galaxy sample

Our sample consists of the nine nearby star-forming disc galaxies targeted by the EMPIRE survey (Bigiel et al. 2016; Jiménez-Donaire et al. 2019). We list their names, orientations, and adopted distances in Table 1. For a more detailed description of the properties of our sample, we refer to Jiménez-Donaire et al. (2019). Summarizing, our targets are all massive, star-forming disc galaxies, with stellar masses of $10 < \log_{10}(M_\star/\mathrm{M}_\odot) < 10.6$, metallicities from half-solar to solar, and star formation rate (SFR) surface densities in the range $2.8-21 \times 10^{-3}\,\mathrm{M}_\odot\,\mathrm{yr}^{-1}\,\mathrm{kpc}^{-2}$.

2.2 EMPIRE CO(1–0) data

EMPIRE mapped the entire optical discs of these galaxies in several 3 mm emission lines using the EMIR receiver. One main goal of EMPIRE is to understand how the dense gas fraction depends on the environment within and among galaxies. To achieve this goal, EMPIRE obtained deep, extended maps of high critical density lines that trace dense gas, such as HCN (1–0), HCO^+ (1–0), and HNC(1–0). In order to measure the dense gas fraction, EMPIRE also required a high-quality tracer of the total molecular gas. This was accomplished by mapping the ^{12}CO(1–0) and ^{13}CO(1–0) lines.

We employ the ^{12}CO(1–0) data from EMPIRE for eight galaxies (PI Jiménez-Donaire, projects 061-51 and 059-16, Jiménez-Donaire et al. 2019; PI Cormier, project D15-12 for NGC 5055, Cormier et al. 2018). For NGC 5194, we use the ^{12}CO(1–0) data from the PdBI Arcsecond Whirlpool Survey (PAWS; Pety et al. 2013; Schinnerer et al. 2013). This was also obtained by the IRAM 30-m using an almost identical strategy to the EMPIRE project.

These CO(1–0) maps cover the full disc of each galaxy, with an angular resolution of 27 arcsec (∼1–2 kpc).[1] They have rms noise

[1]When we quote the beam of single-dish maps, we refer to the effective beam size, which combines the primary beam of the telescope and that of the gridding kernel. All observations used short dump times that critically sampled the beam along the scan direction. See Mangum, Emerson & Greisen (2007) for general information on gridding kernels and see Leroy et al. (2009), Jiménez-Donaire et al. (2019), and Herrera et al. (2020) for information on the HERACLES, EMPIRE, and PHANGS-ALMA maps.

between 13 and 24 mK in each 4 km s^{-1} channel. For full details regarding the observing strategy, reduction, and data products, see Jiménez-Donaire et al. (2019).

We estimate the overall calibration uncertainty of EMPIRE to be ∼5 per cent and we expect this to mostly be a multiplicative factor that scales the whole map. Most EMPIRE maps were made by combining many observing sessions that each covered the whole galaxy. Therefore, we expect the maps to be well-calibrated internally. We regularly observed line calibrators as part of the EMPIRE observing strategy. Cormier et al. (2018) and Jiménez-Donaire et al. (2019) showed that the absolute flux calibration of the EMPIRE data showed rms variation of ∼4−8 per cent from session to session.

Given the overall brightness of the CO(1–0) line, this calibration term represents the dominant source of uncertainty over the inner region of many galaxies.[2]

2.3 CO(2–1) data

We compare the EMPIRE CO(1–0) data to CO(2–1) maps from the IRAM 30-m and ALMA. In each case, we picked the highest quality available CO(2–1) map. All of the CO(2–1) data have higher native resolution than the CO(1–0) maps. We convolved them to the resolution of EMPIRE using a Gaussian kernel with width determined by subtracting the current beam from the target beam in quadrature. We then aligned the CO(2–1) data to the EMPIRE astrometric grid and rebinned to 4 km s^{-1} channels.

For NGC 5194 (M51), we use observations from a new IRAM 30-m Large Program (PI: Toshiki Saito, project 055-17; den Brok et al., in preparation). The goal of this programme is to obtain sensitive observations of 1-mm and 3-mm CO isotopologue transitions and thereby to improve constraints of ISM physical quantities. These observations were carried out using the EMIR instrument at the IRAM 30-m telescope with a total of 172 h. The programme included new CO(2–1) observations, which we use here. At 27 arcsec resolution and 4 km s^{-1} channel width, this new CO(2–1) cube has rms noise 5.6 mK.

The observing strategy and instrument for the M51 EMIR maps closely resemble that of EMPIRE. Therefore, we expect that the uncertainty in the amplitude calibration of the M51 EMIR CO(2–1) map is also similar to that of the EMPIRE maps and consider ∼5 per cent a good estimate.

ALMA observed CO(2–1) emission from NGC 0628, NGC 2903, NGC 3627, NGC 4254, and NGC 4321 part of the PHANGS-ALMA survey (Leroy et al. 2021b). PHANGS-ALMA is using ALMA's 12-m, 7-m and total power antennas to observe CO(2–1) emission from a large sample of nearby. We begin with the cubes made from combining the 12-m, 7-m, and total power observations. Because we convolve the data to 27 arcsec for our analysis and the ALMA single dishes have a ∼27 arcsec beam, the total power data contribute almost all of the information for our analysis. As a result, the details of interferometric imaging are secondary. The PHANGS-ALMA total power pipeline is described by Herrera et al. (2020). After convolving to 27 arcsec, the PHANGS-ALMA cubes have on average an rms noise of 1–2 mK in each 4 km s^{-1} channel.

[2]For CO(1–0), considering all individual lines of sight, we find 839/5416 points have S/N > 20, compared to 3528/5416 points with S/N > 3. For CO(2–1) on the other hand, we have 1130 points with S/N > 20 compared to 4134 with S/N > 3.

The PHANGS-ALMA data are calibrated using observations of solar system objects or Galactic star forming regions that are pinned to the ALMA interferometric calibration scheme. The overall uncertainty in the flux calibration should be about 5 per cent at 1 mm. The flux of total power observations targeting the same PHANGS-ALMA galaxy on different days scatter from one another by ±3 per cent (Leroy et al. 2021a), in good agreement with the aforementioned 5 per cent (Bonato et al. 2018).

For the remaining three sources, NGC 3184, NGC 5055, and NGC 6946, we take observations from HERACLES (Leroy et al. 2009). HERACLES surveyed CO(2–1) emission from 48 nearby galaxies. These maps have appeared previously in Schruba et al. (2011, 2012), Leroy et al. (2013), and Sandstrom et al. (2013). After gridding, the HERACLES maps have a native resolution of 13.3 arcsec. After matching to the EMPIRE 27 arcsec beam and velocity grid, the HERACLES cubes have rms noise between 5 and 11 mK per 4 km s^{-1} channel.

The overall flux scale of HERACLES is uncertain at the ∼6−20 per cent level (see Leroy et al. 2009, and Appendix C). Also, the HERACLES maps combine information from multiple receiver pixels that can have gain uncertainties relative to one another. We investigate the internal gain variations of the HERA pixels and compare HERACLES, ALMA, and EMIR data for galaxies with multiple maps in Appendix C. This analysis yields the gain uncertainty mentioned above and also leads us to prefer ALMA or EMIR maps when available because their calibration should be more stable (i.e. their calibration uncertainties are likely to be well described by a single gain factor).

As with the CO(1–0) data, the high signal-to-noise ratio of the CO(2–1) data means that calibration often represents the dominant source of uncertainty. Anywhere that the HERACLES data exceed S/N = 5–10, and anywhere that the ALMA and EMIR data exceed S/N = 20, calibration will dominate our uncertainty on the brightness temperature. Below the S/N threshold, the uncertainty is dominated by the random noise.

We need the best intensity accuracy possible for this study. Comparing ALMA/EMIR to HERACLES, ALMA/EMIR has both lower absolute and relative flux uncertainty across the map. Although all the galaxies have been also observed with HERA, we therefore choose the ALMA/EMIR over HERACLES where possible in our analysis. In summary, the associated uncertainties for the CO(2–1) emission line from ALMA are 5 per cent, from EMIR are 5–10 per cent, and from HERA are 6–20 per cent (see Appendix C for more details).

2.4 Far-infrared data

We compare the R_{21} ratio to IR maps at wavelengths of 70, 160, and 250 μm from the *Herschel* space telescope. These were compiled and processed to match the EMPIRE beam and astrometric grid by Jiménez-Donaire et al. (2019). For seven of our targets, the data come from the KINGFISH survey (Kennicutt et al. 2011). For NGC 5194, the data come from the Very Nearby Galaxies Survey (Parkin et al. 2013). NGC 2903 lacks *Herschel* data. As a result we cannot determine the 70–160 μm ratio or the total IR (TIR) luminosity surface brightness in Section 3.4 for this galaxy.

2.5 Measured quantities

We follow a similar analysis path to the ^{13}CO-focused study of Cormier et al. (2018) and the HCN-focused study of Jiménez-Donaire et al. (2019). We measure the R_{21} ratio as a function of galactocentric

radius, the 70–160 μm ratio, CO brightness temperature, TIR surface brightness, and the TIR-to-CO ratio.

CO(2–1)/(1–0) ratio, R_{21}: We define R_{21} as the line-integrated CO(2–1) surface brightness divided by the line-integrated CO(1–0) surface brightness. For both lines, the line-integrated surface brightness has units of K km s^{-1}.

Note that our brightness temperature-based definition of R_{21} differs from the flux density based values often quoted in the high-redshift literature (e.g. Aravena et al. 2010, 2014, 2016; Daddi et al. 2010; Bothwell et al. 2013). Using the velocity-integrated flux density definition, one would expect thermalized lines to show a ratio of about four. Using the brightness temperature scale, the line ratio for a thermalized line will be about unity (see e.g. Solomon & Vanden Bout 2005) or slightly lower due to deviations from the Rayleigh–Jeans approximation.

R_{21} for individual lines of sight: We calculate R_{21} for each line of sight. When doing so, we use exactly the same velocity range for the integral over both lines. To define this velocity range, we use for sightlines outside of the 0.23 r_{25} aperture the velocities covered by H I 21-cm line emission (mostly from THINGS; Walter et al. 2008) as an independent estimate for the velocity range likely to be covered by CO. For lines of sight within the 0.23 r_{25} aperture, where the ISM is mostly molecular, we use the CO(2–1) emission as a proxy for the velocity range. This way we make sure that the broad, central CO lines are fully included. We chose the CO(2–1) line instead of the CO(1–0) because our CO(2–1) maps have higher S/N than our CO(1–0) maps.

Note that because our sampling scheme oversamples the beam by a factor of 4, measurements for R_{21} from adjacent lines of sight are correlated and not independent. We take this into account in our presentations of results.

R_{21} from spectral stacking: In addition to measuring R_{21} for individual lines of sight, we employ a spectral stacking method to explore possible correlations between the R_{21} ratio and various physical quantities. In this approach, we bin the data by some other quantity, for example, TIR surface brightness. We construct an average CO(2–1) and CO(1–0) spectrum for each bin. We estimate the mean R_{21} in that bin by dividing the integrated brightness temperature calculated from each binned spectrum.

The method is described in detail by Schruba et al. (2011), Jiménez-Donaire et al. (2017), and Cormier et al. (2018). We regrid each spectrum so that the local mean velocity now corresponds to $v = 0$ km s^{-1}. For this application, we use the velocity field derived from the H I 21-cm data to estimate the local mean velocity. After regridding the spectra, we average together all spectra in each bin. Because the large-scale velocity gradient has been removed, spectra from different parts of the galaxy average coherently.

We derive the integrated brightness temperature from each stacked spectrum. We pick the velocity range for this integral by first fitting the spectrum. We use either a single-Gaussian profile or a double-horn profile, whichever fits better. The double-horn profile offers a better description of the broad, flat-topped emission lines found in some of our galaxy centres. We set the velocity range for direct integration of the spectrum to cover everywhere that the fit exceeds 1 per cent of the peak brightness temperature. Note that the fit is *only* used to set boundaries over which we integrate the spectrum.

We only present stacked measurements of R_{21} when both lines have an integrated emission above 3σ of the rms noise. In practice, our stacks almost always achieve much higher signal-to-noise ratio than this. In Table D1, we summarize the signal-to-noise ratio for our stacks as a function of galactocentric radius. Inside $r_{\rm gal} < 9$ kpc > 95 per cent of the stacked R_{21} measurements have signal to noise >11 for both emission-line brightness temperature measurements, comparable to the very high, pixel based signal-to-noise threshold values used by Koda et al. (2012, 2020).

Uncertainties on R_{21}: We compute the uncertainties on the integrated brightness temperature, $\sigma_{\rm Int}$, using the following formula:

$$\sigma_{\rm Int} = \sigma_{\rm rms} \times \Delta v_{\rm chan} \times \sqrt{n_{\rm chan}}, \quad (1)$$

where $\sigma_{\rm rms}$ is the 1σ rms value of the noise in K measured from the signal-free part of the spectrum, $\Delta v_{\rm chan}$ is the channel width in km s^{-1}, and $n_{\rm chan}$ is the number of channels that are integrated together.

When we apply equation (1) to the stacked spectra, we measure the noise, $\sigma_{\rm rms}$, from the signal-free region of the stacked spectrum itself. As a result, this approach properly accounts for the fact that our original pixels oversample the beam.

After estimating $\sigma_{\rm Int.}$ for each line, we estimate the uncertainties on R_{21} by propagating the errors of the CO(1–0) and CO(2–1) integrated brightness temperature.

Because of the high signal-to-noise ratio in our CO observations, the systematic uncertainty due to flux calibration often dominates the overall uncertainty in R_{21}. For example, in Table D1 we report the median signal-to-noise ratio for stacks within 9 kpc is >30. This ∼3 per cent uncertainty is lower than the systematic uncertainty due to calibration. Our EMPIRE-ALMA or EMPIRE-EMIR galaxies have R_{21} calibration uncertainties ∼7 per cent. For our three EMPIRE-HERACLES targets, this may be as high as ∼20 per cent, and at least 10−15 per cent. At least in the EMIR-ALMA targets, we expect this calibration uncertainty to act as a single multiplicative factor for the map. Thus, it will affect the mean value, but not the internal distribution in each galaxy. For the EMPIRE-HERACLES cases, we expect the primary uncertainty to be an overall scaling, but there may be second-order local variations due to the differences in the pixel gains discussed in Appendix C.

70–160 μm ratio: We compare R_{21} to the 70–160 μm ratio. This ratio traces the temperature of interstellar dust (e.g. Draine et al. 2007). Because most of the large grains in a galaxy are in thermal equilibrium with the local interstellar radiation field (ISRF; see textbook by Draine 2011), this ratio also acts as a tracer of the ISRF. Note that at the average densities and temperatures of molecular clouds traced by CO emission, we do not expect the dust and gas to collisionally couple and share the same temperature (Draine 2011), so we do expect the IR colour to directly trace the ISRF by not the gas temperature. We measure the 70–160 μm ratio after convolving the *Herschel* 70 and 160 μm maps to match the EMPIRE resolution.

TIR surface brightness: We compare R_{21} to the TIR luminosity per unit area. We use the TIR surface brightness as an observational proxy for the amount of embedded recent star formation. This tracer has the advantage compared to other SFR tracers, as it traced the embedded SFR, which means the recent SFR might affect the state of the molecular ISM. We follow the same approach as our previous work (e.g. Usero et al. 2015; Bigiel et al. 2016; Jiménez-Donaire et al. 2017; Cormier et al. 2018).

We combine *Herschel* 70, 160, and 250 μm data in order to estimate the TIR surface brightness. First, we convolve each band to our common resolution of 27 arcsec and place them on to the EMPIRE astrometric grid. Then we combine the bands, following Galametz et al. (2013),

$$S_{\rm TIR} = \sum c_i S_i, \quad (2)$$

where $S_{\rm TIR}$ refers to the TIR surface brightness, S_i to the brightness in the given *Herschel* band i, and c_i to the calibration coefficient from

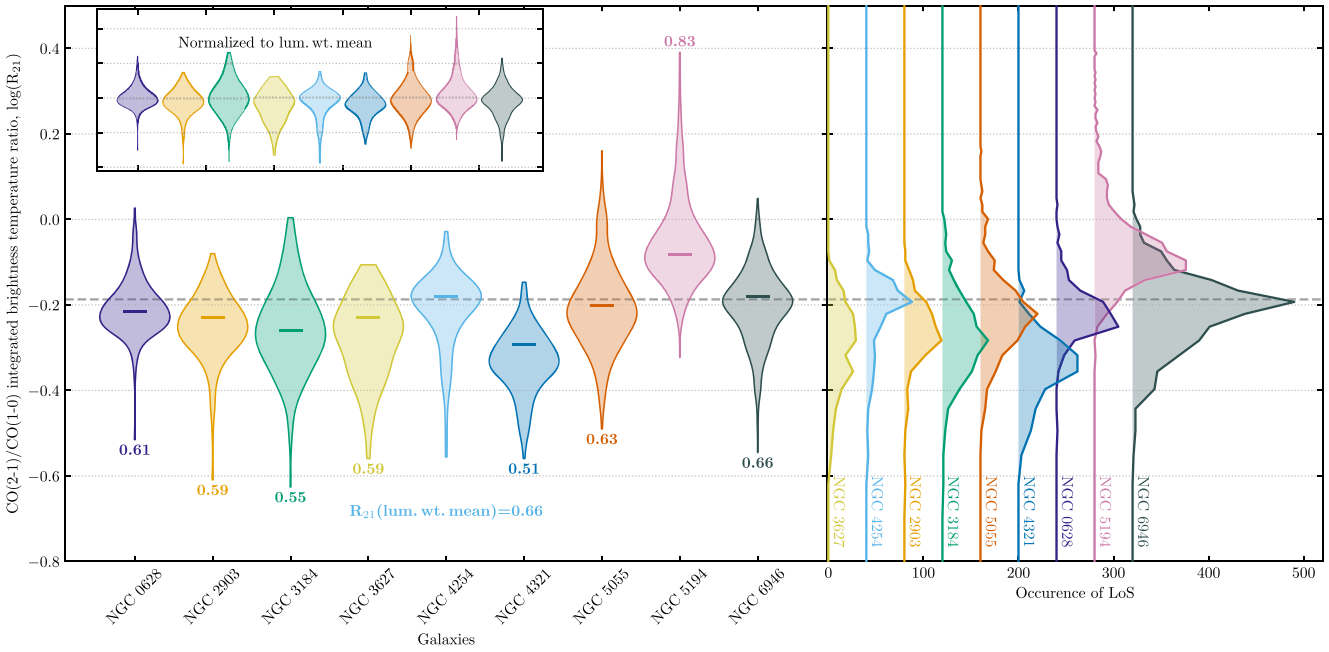

Figure 1. Histograms of the CO(2–1)/CO(1–0) integrated brightness temperature ratio, $\log(R_{21})$, for nine nearby star-forming spiral galaxies. The *left-hand panel* shows distributions of $\log(R_{21})$ as violin histograms. These histograms only show results for the positions that have S/N > 3 integrated brightness temperature detections in both the CO(1–0) and CO(2–1) maps. These histograms treat each line of sight equally. The *luminosity-weighted mean* R_{21} value for each galaxy appears as a coloured horizontal line inside each histogram, with the value reported above or below each violin histogram. All data, regardless of S/N, are included within the calculation of the luminosity-weighted mean, as shown in Table 2. We note that the uncertainty incorporates physical galaxy-to-galaxy scatter as well as uncertainties in flux scale calibration. The *inset* within the left-hand panel shows the R_{21} violin histograms for each galaxy normalized to their luminosity-weighted mean R_{21} values. The *right-hand panel* shows the histogram of the combined R_{21} distribution for all targets. The dashed grey line extending across all panels shows $\langle R_{21} \rangle = 0.65$, the non-weighted mean of R_{21} if including all lines of sight with S/N > 3 (see last column of Table 2).

combined brightness. We use the specific calibration coefficients provided for each galaxy, with the exception of NGC 5194 where we use the generic calibration provided by Galametz et al. (2013), since this galaxy was not explicitly studied.

We focus on TIR surface brightness because it represents a simple, reproducible quantity that is closely related to the local surface density of recent star formation. We do not implement any specific conversions or consider second-order effects like IR cirrus. Our analysis also does not hinge on any numerical conversion of TIR surface brightness to an SFR. For a detailed discussion of the use of TIR as an SFR proxy and a quantitative comparison to other star formation tracers (e.g. Cormier et al. 2018; Gallagher et al. 2018b; Jiménez-Donaire et al. 2019).

3 RESULTS

3.1 Overall distribution of R_{21}

We estimate the line ratio, R_{21}, for each line of sight that has a measured brightness temperature for both CO(2–1) and CO(1–0) lines. In total, this yields 5416 measurements across nine galaxies at 27 arcsec (∼1–2 kpc) resolution.

Fig. 1 shows the R_{21} distribution of all individual lines of sight for each galaxy, as well as a histogram of the combined R_{21} distribution for the entire sample. These histograms visualize results only for lines of sight with S/N > 3 in both lines. We do not find many cases where only one of the two lines is detected, highlighting that the two lines follow similar distributions and the data sets are well matched. The individual sightlines are arranged in a hexagonal grid (see Fig. 2), where the points have a half-beam separation distance. Consequently, the beam size is oversampled by four hexagonal grid pixels.

Table 2 reports the luminosity-weighted mean value for each galaxy, as well as the 16th/84th and 5th/95th percentiles. Here, 'luminosity weighted' means averaging over the individual R_{21} values weighted by the corresponding CO (1–0) brightness temperature. We prefer to use these intensity-weighted values for our quantitative results because they map straightforwardly to the results expected from galaxy-integrated measurements.

For individual galaxies, we find luminosity-weighted mean R_{21} ratios ranging between ∼0.51 and 0.87. In our view, the best characteristic sample-wide value for R_{21} is the mean of the luminosity-weighted mean ratios for the individual galaxies. This is $\langle R_{21}^{\rm mean} \rangle = 0.64 \pm 0.09$ with 0.10 rms scatter from galaxy to galaxy. The uncertainty is the standard deviation between the galaxies. The value of the ratio agrees well with previous measurements of a wider population of galaxies, which tend to lie in the range 0.5–0.8 (Sections 1 and 4). We verified that no significant effects are found when different weighting schemes are used.

In principle, our choice of method could affect our derived mean R_{21} if, e.g. a few very bright regions show different R_{21} than the rest of the galaxy or there is a large diffuse component with different R_{21}. The small differences among different approaches in Fig. 1 and Table 2 show that this is mostly not the case for our sample. The galaxy-wide mean and intensity-weighted mean show good agreement for most galaxies. Moreover, we find an average ratio of $\langle R_{21}^{\rm mean,norm} \rangle = 0.62$ and a standard deviation of 0.12 when weighting

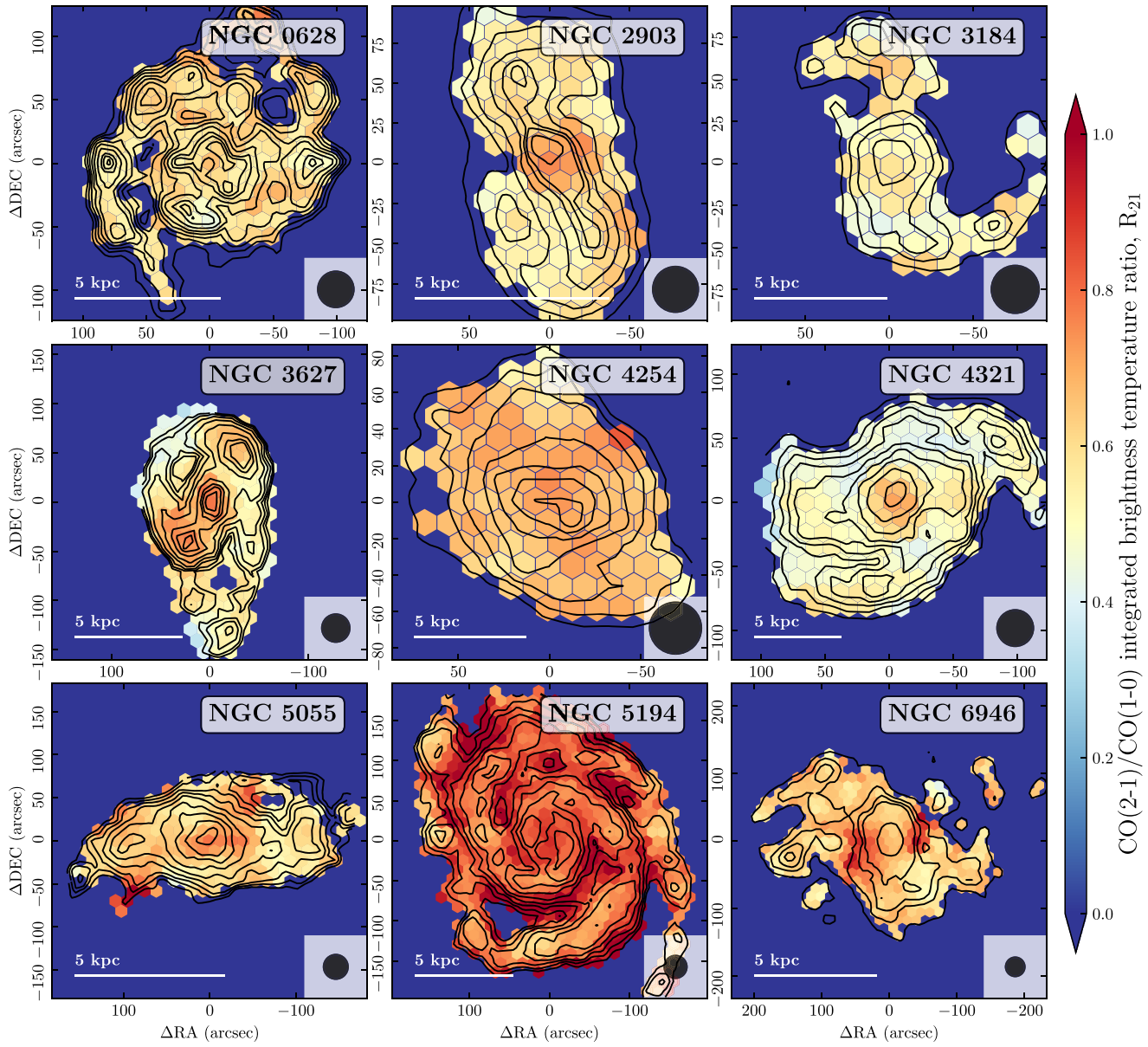

Figure 2. Maps of the CO(2–1)/CO(1–0) integrated brightness temperature ratio, log(R_{21}), for nine nearby star-forming spiral galaxies. These maps show R_{21} for all positions that have S/N > 10 integrated brightness temperature in both the CO(2–1) and CO(1–0) maps. The overlaid contours show the CO(1–0) integrated brightness temperature, with levels showing 20, 30, 40, 50, 60, 70, 80, 90, 95, 97.5, and 99.5 per cent of the peak value in the map (see Fig. A1). The white scale bar in the bottom left corner of each panel shows a linear scale of 5 kpc, without accounting for inclination, at the distance of each source (see Table 1). The black circle in the bottom right corner indicates the beam size of 27 arcsec. Note that the hexagonal grid shows points critically sampling the beam, i.e. adjacent points are spaced by one-half the beam size.

all lines of sight equally, compared to $\langle R_{21}^{\rm mean} \rangle = 0.64 \pm 0.09$ when weighting by the luminosity-weighted mean of each galaxy.

Fig. 2 shows the maps of the distribution of the CO line brightness temperature ratio R_{21} across the individual galaxies. We do find evidence for both radial and azimuthal variations. We explore the systematic variation of the ratio within and between individual galaxies in the following sections.

3.2 Radial variations of R_{21}

Many quantities, including the SFR, molecular gas fraction, and gas density vary as a function of galactocentric radius. In Fig. 3, we present radial profiles of R_{21} for our sample. We plot all individual lines of sight at 27 arcsec (∼1–2 kpc) resolution. Recall that for these data, adjacent points are spaced by one-half beam so that the points are not independent. Filled points show measurements with a signal-to-noise ratio, S/N > 3, on the line ratio, propagated. Open symbols indicate measurements with S/N < 3.

Coloured points in Fig. 3 show radial profiles of azimuthal-averaged ratios, with error bars indicating the uncertainty on this mean R_{21}. For these stacked profiles, we use all of the data, regardless of S/N. We plot all of the stacked profiles together in Fig. 4.

In Figs 3 and 4, our azimuthally averaged measurements of R_{21} show only a small dynamical range across individual galaxy discs for

Table 2. The CO(2–1)/CO(1–0) integrated brightness temperature ratio, R_{21}, for nine nearby star-forming spiral galaxies. The upper half of the table shows luminosity-weighted statistics (R_{21}). The lower half of the table shows number statistics, which treat each pixel equally ($R_{21}^{\rm num}$). We tabulate the galaxy name, values of the mean, and 5th, 16th, 50th (i.e. median), 84th, and 95th percentile ranges of R_{21} for our target galaxies. The second to last column gives the mean value and associated standard deviation of R_{21} across the whole sample. The last column lists the mean and percentiles when considering all significant (S/N > 3) lines of sight across the whole sample. Note that all values within this table have been calculated for positions with a significant detection in both the CO(1–0) and CO(2–1) integrated brightness temperature maps (i.e. S/N > 3). The bottom rows indicate the number of lines of sight (l.o.s.) for the individual galaxies. The number of all l.o.s. as well as those l.o.s that have both the CO(1–0) and CO(2–1) above the threshold of 3σ and 10σ.

NGC	0628	2903	3184	3627	4254	4321	5055	5194	6946	Galaxy average	All sightlines ($>3\sigma$)
$R_{21}^{\rm mean}$	0.61	0.59	0.55	0.59	0.66	0.51	0.63	0.83	0.66	0.63 ± 0.09	0.66
$R_{21}^{5\%}$	0.5	0.48	0.41	0.41	0.49	0.38	0.48	0.68	0.51		0.49
$R_{21}^{16\%}$	0.55	0.52	0.46	0.48	0.61	0.44	0.54	0.74	0.58		0.56
$R_{21}^{50\%}$	0.6	0.58	0.54	0.59	0.66	0.5	0.62	0.81	0.65		0.71
$R_{21}^{84\%}$	0.67	0.7	0.64	0.71	0.71	0.6	0.72	0.91	0.73		0.85
$R_{21}^{95\%}$	0.75	0.72	0.78	0.76	0.74	0.68	0.79	1.02	0.83		0.94
$R_{21}^{\rm num.mean}$	0.62	0.57	0.57	0.54	0.63	0.47	0.64	0.9	0.65	0.62 ± 0.12	0.65
$R_{21}^{\rm num.5\%}$	0.52	0.44	0.4	0.37	0.42	0.34	0.44	0.66	0.47		0.42
$R_{21}^{\rm num.16\%}$	0.55	0.49	0.46	0.43	0.53	0.4	0.51	0.73	0.54		0.49
$R_{21}^{\rm num.50\%}$	0.61	0.57	0.55	0.54	0.65	0.47	0.61	0.84	0.64		0.63
$R_{21}^{\rm num.84\%}$	0.7	0.65	0.68	0.66	0.71	0.54	0.75	1.06	0.74		0.82
$R_{21}^{\rm num.95\%}$	0.79	0.72	0.82	0.73	0.77	0.59	0.97	1.37	0.85		1.01
$N_{\rm l.o.s}^{\rm all}$	328	319	741	239	288	318	410	947	1824	5414	
$N_{\rm l.o.s}^{>3\sigma}$	292	188	302	200	207	304	370	705	884	3452	
$N_{\rm l.o.s}^{>10\sigma}$	153	101	65	136	95	162	201	475	321	1709	

some cases while for other sources the ratio tends to show a negative or positive gradient. In Table 3, we report the power-law fit relating R_{21} to galactocentric radius in each galaxy. We also report the p-value (of a linear relation in logarithmic space), which allows us to gauge the significance (p-value) of the radial gradient.

Six out of the nine galaxies show radial variations. Four show stronger radial variations. NGC 2903 shows an initial radial decline then a rise in R_{21} with increasing radius. NGC 3184 shows increasing R_{21} with increasing radius. NGC 3627 and NGC 4321 show a strong while NGC 4254 shows a moderate decreasing trend as a function of the radius.

Fig. 4 also shows that in six of our nine targets, R_{21} appears higher in the central kpc than at intermediate radii, ∼1–6 kpc. This central enhancement is most prominent in NGC 2903, NGC 3627, NGC 4321, and NGC 5055. Other galaxies, for example NGC 4254, show little or no central enhancement. On average, the central R_{21} bin (0–1.5 kpc) for our targets is 16 per cent higher (median is 15 per cent) compared to the luminosity-weighted average of the rest of the galaxy.

Outside a galactocentric radius of about 6 kpc, we find highly variable behaviour among our sample, with some galaxies showing increasing R_{21}, some showing decreasing R_{21}, and some being flat. Fig. 3 shows that these breaks in the profile often coincide with the emergence of a large amount of low signal-to-noise data. Using different techniques to bin the data and estimate the binned ratio yield large discrepancies. We are therefore hesitant to overinterpret them. Sensitive multiline observations of outer discs will help illuminate whether CO excitation does change dramatically in the outer parts of disc galaxies.

3.3 Azimuthal variations of R_{21}

Fig. 2 also shows variation in R_{21} at fixed galactocentric radius. The SFR surface density and gas column density also vary azimuthally, with the most striking features due to the influence of spiral arms and bars. NGC 3627, 5194, NGC 6946, and to some extent NGC 2903, 4321, and 5055 show clear spatial variations in the CO line ratio. NGC 5194 shows higher R_{21} ratios in the interarm. We note that this stands in contrast to previous findings. Koda et al. (2012) found a higher line ratio in the arm region as opposed to the interarm region in this galaxy. For NGC 3627, we find a higher line ratio in the centre and at the bar ends. NGC 6946 shows regions with enhanced line ratio towards the east and west of the centre. NGC 2903, 4321, and 5055 all show an increase of the CO line ratio in the central ∼1–2 kpc region. The other three sources do not show any clear spatial variations.

At 27 arcsec resolution, our ability to distinguish arm and interarm regions is limited, especially in the inner parts of galaxies where most of the molecular gas resides. The most straightforward imprint of azimuthal R_{21} variations on our data is to increase the observed scatter in R_{21} at fixed radius, e.g. as suggested in Fig. 3.

To quantify these azimuthal variations in R_{21}, we measure the scatter in the ratio at a fixed galactocentric radius. We disentangle the physical variation from scatter due to observational noise using a forward modelling process, which we describe in Appendix B. Briefly, we use a Monte Carlo approach and the known observational errors to determine how much physical variation must be present in each radial bin to match the observed scatter. We plot the results of this calculation in Fig. 5, where we repeat the modelling routine for every individual galaxy. We show best-fitting physical variation in R_{21} as a function of radius for each galaxy.

On average, the intrinsic scatter rises from $\lesssim$10 per cent in the inner bins to $\gtrsim$20 per cent outside a galactocentric radius of 6 kpc. This has roughly the same magnitude as the observed galaxy-to-galaxy scatter. Note, however, that we do not expect calibration uncertainties to play as large a role in the scatter observed *within* a galaxy. Taking this into account, the physical scatter within galaxies may be larger than the physical scatter among galaxies. Also note that our azimuthal scatter calculations consider each pixel equally. A luminosity-weighted calculation would suppress faint regions and lower the magnitude of the measured scatter. A reasonable overall conclusion from this is that in our data, point-to-point scatter has a magnitude greater than or equal to galaxy-to-galaxy variations, and appears stronger than radial variations.

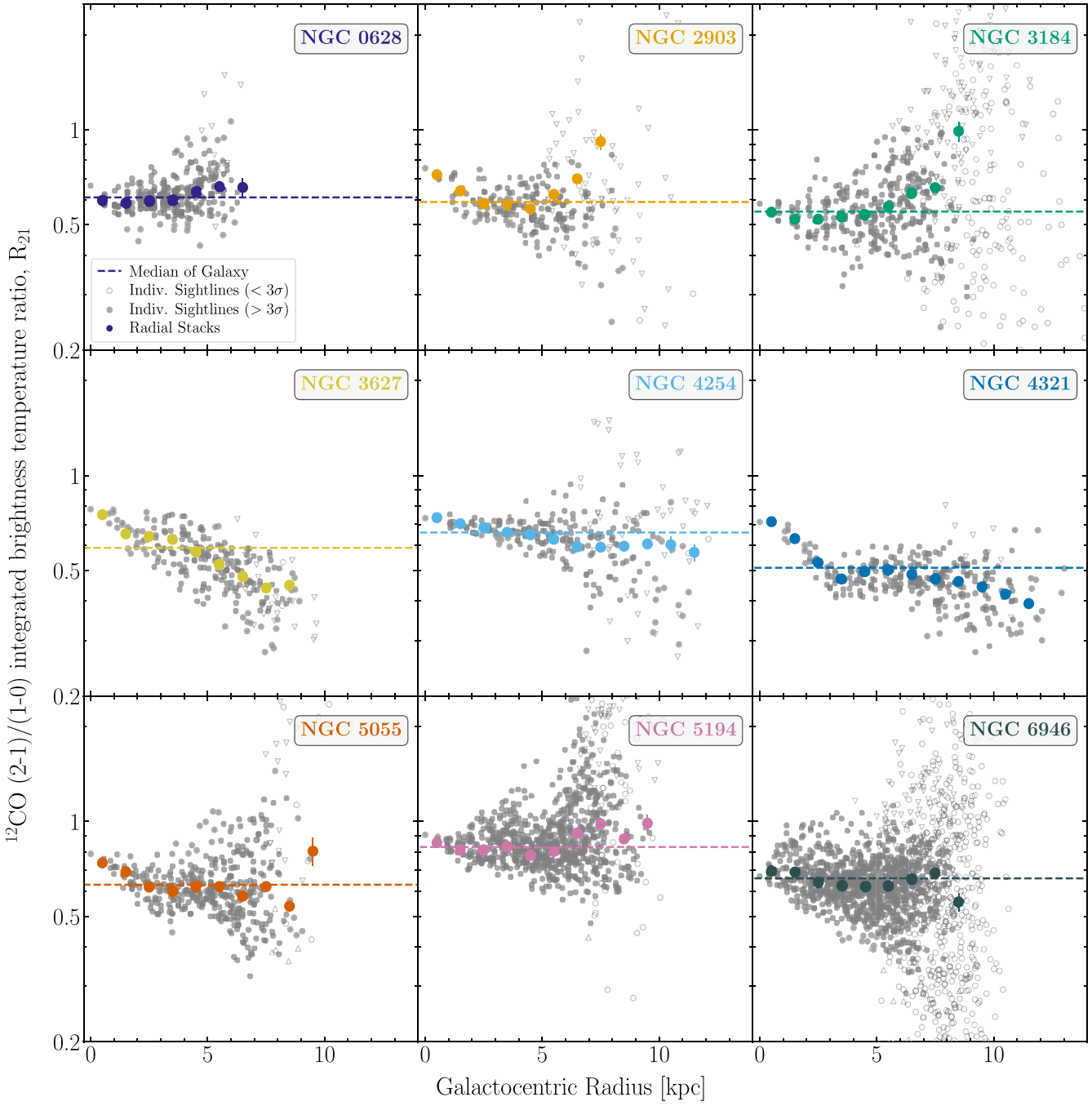

Figure 3. Radial profiles of the R_{21} ratio. The grey points correspond to individual sightlines with adjacent lines of sight spaced by one half the beamwidth. The line ratio is plotted on a logarithmic scale. Filled circles indicate data points where both lines have a signal-to-noise ratio above 3. Data with lower signal-to-noise ratio appear either as upper or lower limits. Open triangles show where one line is below S/N = 3. Open circles show points where both lines are below the S/N threshold of 3. Coloured circles present the stacked values of the line ratio calculated in 27 arcsec-wide radial bins following the method described in in Section 2.5. The galaxy-wide median value for each individual galaxy appears as a coloured dashed line. Six of our nine targets show clear central enhancements in R_{21}, but otherwise the stacked profiles show relatively small deviation from the galaxy-wide median.

We also applied the same Monte Carlo based analysis on the complete data of all lines of sights as a whole. This way we can estimate the overall physical scatter. The physical scatter estimated after accounting for different calibration uncertainties for the individual instruments is about 8 per cent.

As stated before, for NGC 5194 (M51) we measure spatial variations that have the opposite sense of those reported by Koda et al. (2012). In the arm region, we find a R_{21} value ∼0.8, which is in agreement with the value found by Koda et al. (2012) in the arm region. However, in the interarm region, we find larger values (R_{21} ∼ 0.9–1), while they find lower values (R_{21} ∼ 0.4–0.6). We examine in detail and discuss possible causes for the difference in Appendix E, and find the disagreement to stem from differences in the CO (1–0) maps used for the analysis. In particular, the NRO map used by Koda

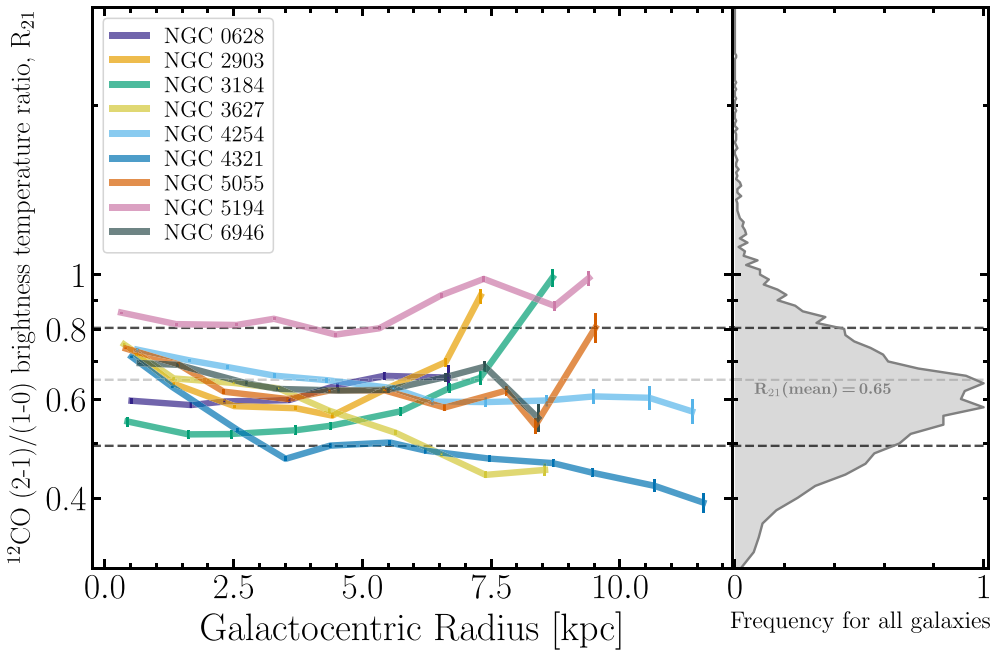

Figure 4. Stacked profiles of the CO line brightness temperature ratio, R_{21}, as a function of galactocentric radius, plotted using a *y*-axis logarithmic scale. Stacked radial profiles use a bin width of 1.5 kpc. We only plot stacked bins where both CO(1–0) and CO(2–1) integrated, stacked lines are detected above an S/N > 3, but otherwise include all data. Error bars indicate the propagated uncertainty of the integrated brightness temperature of the stacked spectra. The histogram, reproduced from Fig. 1, indicates the distribution of all lines of sight were both the CO(1–0) and CO(2–1) integrated line emissions are significantly (S/N > 3) detected. The light grey, dashed line indicates the sample-wide mean ratio when considering all lines of sight with S/N > 3. The darker, dashed lines indicate the 1σ scatter over all sightlines. The figure illustrates the galaxy-to-galaxy scatter that accounts for a large fraction of the overall variation of R_{21} in the sample.

et al. (2011) shows more emission in the interarm region than the IRAM 30-m PAWS CO(1–0) map. Beyond a galactocentric radius of 2 kpc, NGC 5194 shows the highest scatter in our sample and Fig. 2 does show the strongest arm–interarm contrast. At our resolution, this contrast manifests as rms physical scatter of 15−20 per cent between $r_{gal} = 2$ and 6 kpc. Koda et al. (2012) report a standard deviation of ∼0.1–0.15 and mean $R_{21} \approx 0.7$, so our numbers for both the mean line ratio and scatter appear to be overall slightly larger than theirs, which is mostly driven by the differing values within the interarm region.

As Fig. A1 shows, our 27 arcsec resolution only coarsely resolves the dynamical features in our targets. Spiral arms and bars are visible at large radii in many targets. However, we cannot distinguish the upstream and downstream sides of these features and they can be almost entirely suppressed in the inner galaxies. Physical conditions can vary dramatically across a spiral arm (e.g. Schinnerer et al. 2010, 2017). Thus, we expect our large beam to blur together regions with a wide range of temperature and densities, especially in the inner parts of galaxies. This effect is expected to be even stronger, when a bright arm region lies next to a faint interarm region, the wider spacing between, e.g. arms and other discrete regions in the outer parts of galaxies may partially explain the increased scatter at large radii. Future work at higher physical scales offers the prospect to give much more insight on local variations of R_{21}.

3.4 Correlations with CO brightness temperature and IR emission

We also compare R_{21} to the local intensities of CO(1–0) and CO(2–1) emission, the local TIR surface brightness, and the local 70–160 μm ratio. These observed quantities indirectly trace physical conditions that should affect CO excitation, so that this analysis can highlight the physical drivers of the R_{21} variations observed in the previous two sections.

We compare to these specific quantities because they are directly observable and also indirectly related to conditions which we expect to affect excitation. Though we observe at coarse physical resolution, we expect that CO(1–0) and CO(2–1) emission trace the molecular gas surface density and more indirectly trace gas volume density. High gas densities will be associated with thermalization and a higher R_{21}. The IR colour traces the dust temperature, which in turn is set by strength of the interstellar radiation field (Draine 2011). The radiation field also illuminates photon-dominated regions and should play a key role in heating the gas. All other things equal, we expect warmer gas to be more nearly thermalized. Along similar lines, the TIR surface brightness indicates the level of star formation activity. We expect that this indirectly relates to both the heating of the gas and the gas density, with denser gas forming more stars, on average.

Fig. 6 shows the mean normalized R_{21} (normalized with respect to the galaxy-internal luminosity-weighted mean; see Table 2) calculated from spectral stacks as a function of each quantity of interest. We show results for each galaxy separately and show results stacking the data in bins of CO(1–0) brightness temperature, CO(2–1) brightness temperature, TIR surface density, and 70–160 μm ratio. Table 3 reports the results form fitting a power law relating R_{21} to each quantity for all individual points. In addition, the *p*-value is indicated for the correlation in log-space, describing the tightness of the correlation. Finally, we also report the Spearman's rank correlation coefficient, r_s. For the stacking, we only included sightlines with S/N > 10 in both the CO(1–0) and CO(2–1) data to make sure that the

Table 3. Fits and correlation coefficients for individual galaxies. Results from fitting a power law of form $R_{21,\mathrm{norm}}^{\mathrm{fit}} = C \cdot x^m$ to the stacked ratios in Figs 6 and 7. The fit is performed as a linear fit in logarithmic space. We normalized the line ratio by the luminosity-weighted, galaxy wide mean (see Table 2). The Pearson p-value indicates the significance of the linear correlation in logarithmic space. We only performed the fit, if the p value is below 0.05. Furthermore, the Spearman's rank correlation coefficient, r_s, is given. For CO(1–0) the fitting range was set to exclude points below $I_{\mathrm{CO(1-0)}} < 10$ K km s^{-1}. A minus indicates that no fit could be made. For NGC 0628 and NGC 3184, we do not have $I_{\mathrm{CO(1-0)}} > 10$ K km s^{-1} data and for NGC 2903 we do not have IR data from *Herschel*.

		NGC 0628	NGC 2903	NGC 3184	NGC 3627	NGC 4254	NGC 4321	NGC 5055	NGC 5194	NGC 6946
CO(1–0)	m	–	0.17	–	0.23	0.73	0.18	0.16	–	–
[K km s^{-1}]	C	–	0.61	–	0.46	0.82	0.62	0.62	–	–
	p	–	0.048	–	0.030	0.035	6.4×10^{-4}	0.026	0.26	0.24
	r_s	–	1.0	–	1.0	1.0	1.0	1.0	0.60	0.80
CO(2–1)	m	−0.0027	0.14	0.13	0.24	0.14	0.17	0.076	–	–
[K km s^{-1}]	C	1.0	0.70	0.86	0.56	0.74	0.69	0.82	–	–
	p	0.0	0.0050	0.0	0.0016	0.034	0.0053	0.029	0.08	0.06
	r_s	−1.0	1.0	1.0	1.0	1.0	1.0	0.9	0.68	0.77
PACS 70/160	m	0.15	–	0.19	0.43	–	0.40	0.21	0.20	0.20
	C	1.2	–	1.3	1.5	–	1.6	1.3	1.2	1.2
	p	0.039	–	0.047	2.8×10^{-6}	0.07	0.00033	0.0083	0.00046	0.03
	r_s	0.9	–	1.0	1.0	1.0	1.0	0.82	1.0	0.83
Σ_{TIR}	m	–	–	–	0.20	0.080	0.20	0.082	0.020	0.1
[W kpc^{-1}]	C	–	–	–	6.7×10^{-8}	0.0016	1.1×10^{-7}	0.0014	0.20	0.00025
	p	0.40	–	0.27	0.00082	0.034	0.0023	0.0073	0.037	0.026
	r_s	0.4	–	0.8	1.0	1.0	1.0	0.94	0.86	0.94
$\Sigma_{\mathrm{TIR}}/\mathrm{CO}(1-0)$	m	0.24	–	0.2	–	0.11	–	–	0.17	–
[(W kpc^{-1})/(K km s^{-1})]	C	–	–	9.8×10^{-8}	–	0.00021	–	–	1.3×10^{-6}	–
	p	0.15	–	0.0	0.70	0.0	0.20	0.13	0.011	0.088
	r_s	1.0	–	1.0	1.0	1.0	1.0	1.0	1.0	0.8
$\Sigma_{\mathrm{TIR}}/\mathrm{CO}(2-1)$	m	–	–	−0.063	0.12	−0.053	–	–	–	–
[(W kpc^{-1})/(K km s^{-1})]	C	–	–	1.4×10^{2}	7.4×10^{-5}	64	–	–	–	–
	p	0.25	–	0.0	0.0	0.0	0.60	0.54	0.24	0.18
	r_s	1.0	–	−1.0	1.0	−1.0	−0.50	0.5	1.0	−1.0

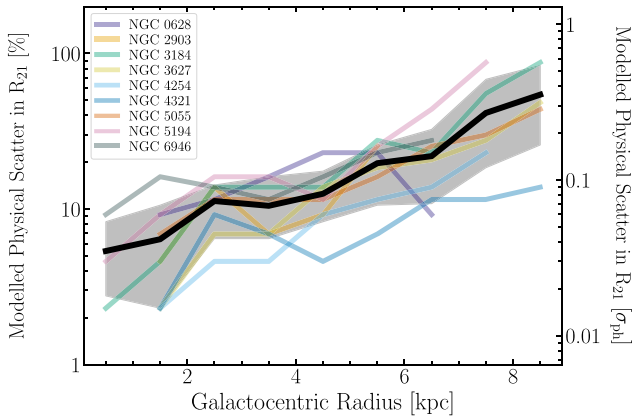

Figure 5. Inferred physical scatter in the R_{21} ratio for each radial bin. The plot shows the physical scatter in each radial bin inferred from our modelling. We have already removed the effects of observational noise from the plotted scatter via a forward modelling Monte Carlo calculation. The left y-axis indicates the scatter in percentage of the line ratio, while the right y-axis describes the scatter's actual value (converting using a fixed $R_{21} = 0.65$). Coloured lines show individual galaxies. The black line and grey region show the median and $\pm 1\sigma$ range combining all galaxies. We observe increasing scatter in R_{21} towards large galactocentric radii, with scatter $\gtrsim$20 per cent typical at radii >6 kpc. This increase in scatter may reflect a large variation in physical conditions at large radius or the inability to separate physically distinct regions at the 27 arcsec (∼1–2 kpc) resolution of our measurements.

trends in the ratio are not noise dominated. The black line indicates the binned mean line ratio and the grey band describes the binned standard deviation. Table 4 lists the results from fitting a power law to the binned mean line ratio indicating the strength of the trend.

CO(1–0) brightness temperature: The top left panel of Fig. 6 shows the stacked, normalized R_{21} as a function of CO(1–0) brightness temperature. The bin width of the stacks is set to 0.25 in logarithmic space. At higher intensities, we observe a tendency to find higher R_{21} in the highest brightness temperature bins. The highest CO(1–0) brightness temperature almost always appears in the galaxy centre, so this reflects the same central enhancements noted in the radial profiles. Overall, Fig. 6 reveals a positive relationship between R_{21} and CO(1–0) brightness temperature.

Because of the correlated axes, low signal-to-noise CO(1–0) measurements will lead to an artificial upturn at the low brightness temperature end driven by sorting predominantly noise measurements, as it can be seen in the top left panel. Supporting this view, no such feature appears in the radial, CO(2–1), or IR intensity plots.

CO(2–1) brightness temperature: The top right panel of Fig. 6 shows stacked R_{21} as a function of CO(2–1) brightness temperature. We also chose a bin width of 0.25 in logarithmic space. As with CO(1–0), we observe a positive correlation between R_{21} and CO(2–1). In general, we tend to find moderately higher R_{21} in high brightness temperature regions.

TIR luminosity surface brightness: The bottom left panel of Fig. 3 shows stacked R_{21} as a function of TIR surface brightness. The bin width for the TIR surface brightness is 0.3 in logarithmic space. TIR surface brightness traces embedded star formation activity and

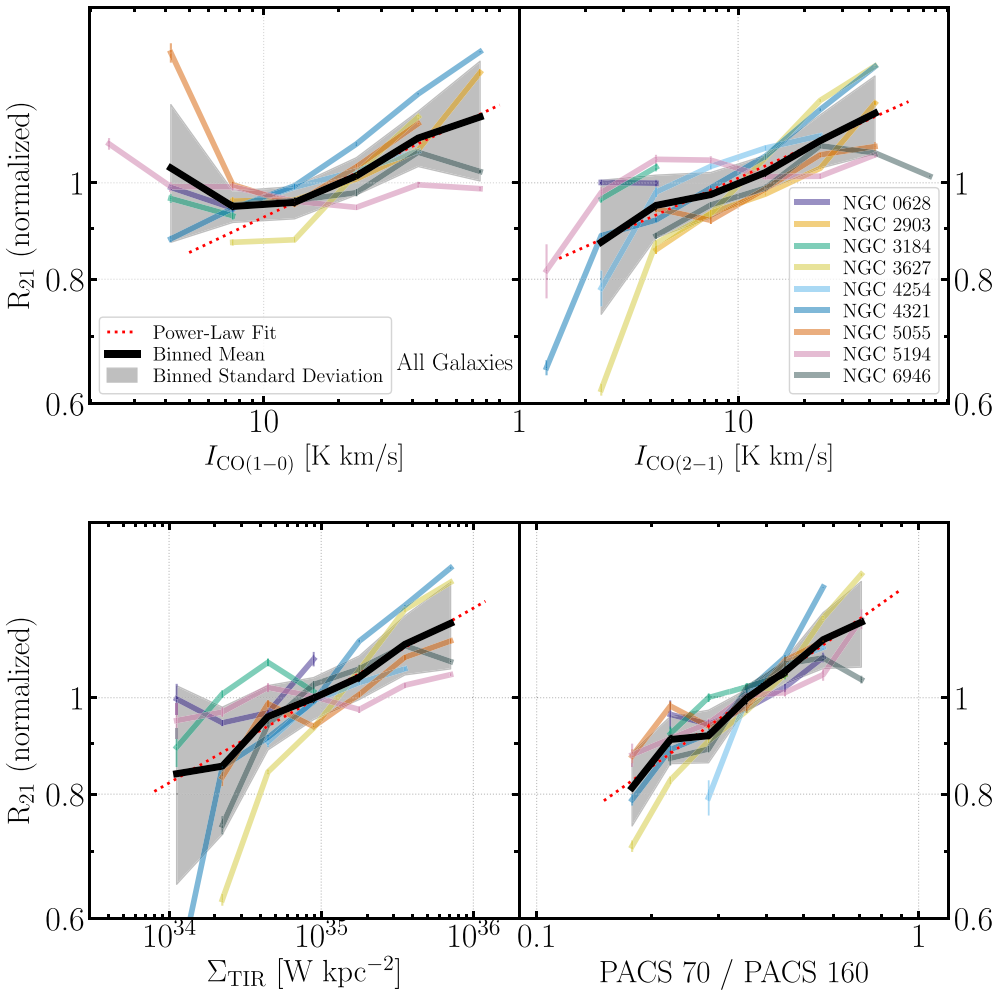

Figure 6. Stacked measurements of R_{21} on a logarithmic scale in bins of CO(1–0), CO(2–1) brightness temperature, TIR surface brightness, and IR colour. In each panel, we bin each galaxy by the quantity on the *x*-axis. Then, we measure stacked line ratios in each bin for each galaxy. For stacking in bins of CO(1–0), CO(2–1) brightness temperature, TIR surface brightness, and IR colour, we only included lines of sight with an S/N > 10 for both CO(1–0) and CO(2–1) in order to reduce noise effects and make the trend more robust. The individual lines are normalized by the median stacked CO ratio value of the corresponding galaxy. The black line is the binned mean combining all galaxies. Only bins with at least a stacking result from three galaxies are included in the mean line. We observe positive correlations between R_{21} and CO brightness temperature, IR surface brightness, and IR colour. These have the sense that R_{21} increases along with gas surface density, star formation activity, and dust temperature. The red dotted line indicates the power-law fit. The fitting range and results are listed in Table 4. We also report rank correlation coefficients in Table 3.

Table 4. Fitting results for stacked, normalized profiles combining all galaxies. Results from fitting a power law of form $R^{\rm fit}_{21,\rm norm} = C \cdot x^m$ to the mean of the stacked, normalized quantities in Figs 6 and 7 (black line in figure). The fitting range indicates the range along the *x*-axis over which the fit is performed.

Parameter (*x*-axis)	Fitting range (units of param.)	C	m	r_s
CO(1–0)	10–75	0.70	0.12	1.0
CO(2–1)	2–43	0.81	0.099	1.0
PACS 70/ PACS 160	0.2–0.7	1.3	0.27	1.0
$\Sigma_{\rm TIR}$	10^{34}–$10^{35.85}$	8.5×10^{-4}	0.088	1.0
$\Sigma_{\rm TIR}$/CO(1–0)	$10^{33.45}$–$10^{34.05}$	5.2×10^{-8}	0.20	1.0
$\Sigma_{\rm TIR}$/CO(2–1)	$10^{33.75}$–$10^{34.35}$	0.15	0.025	0.5

scales with molecular gas surface density, so we expect similar results to stacking by CO brightness temperature. Again, we observe a positive correlation where the IR-bright parts of our sample show moderately higher R_{21}. Because TIR tends to be measured at high signal-to-noise ratio and represents an independent quantity from CO(2–1) and CO(1–0) this correlation spans a larger dynamic range than the CO(1–0) and CO(2–1)-based stacks and should be less subject to systematics. As in the previous panels, we observe a positive correlation between TIR surface brightness and R_{21}. In regions with more star formation per unit area, R_{21} tends to be higher.

70–160 μm ratio: In the bottom right panel of Fig. 6, we plot R_{21} stacked as a function of IR colour. The stacks have a bin with of 0.1 in logarithmic space. IR colour traces dust temperature and the interstellar radiation field. The axes are not correlated, though the lowest bin may again suffer from some sampling and signal-to-noise concerns.

As above, we find a positive correlation between 70–160 μm ratio and R_{21}. R_{21} tends to be higher with stronger interstellar radiation field. This fits with an overall pattern that systems with more intense star formation activity also tend to have higher dust temperatures, denser gas, and more nearly thermal excitation in their CO lines. A higher dust temperature does correspond to a higher R_{21}.

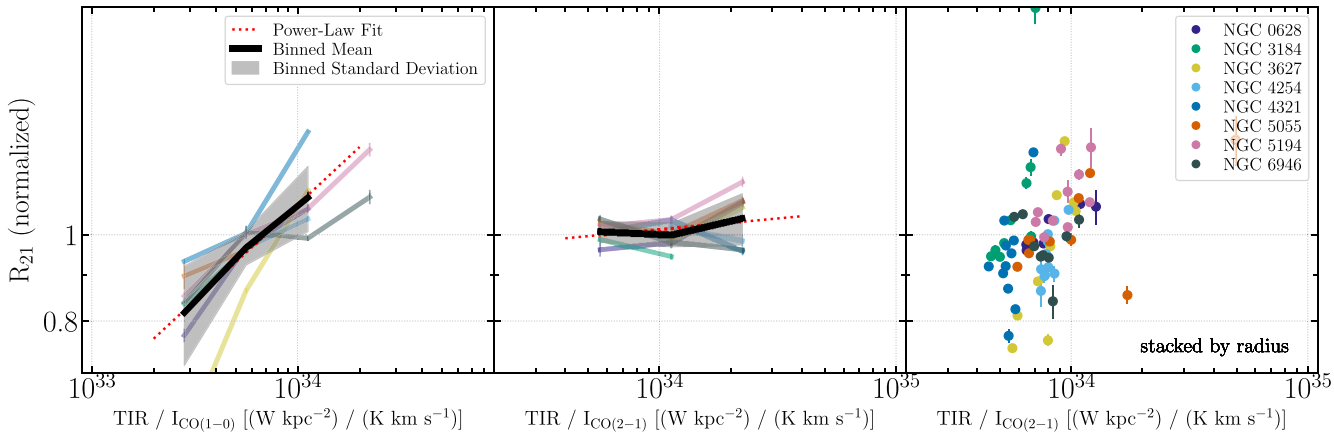

Figure 7. Stacked measurements of the R_{21} ratio as a function of TIR-to-CO ratios. The TIR-to-CO ratio is closely related to the SFR per unit gas mass, a key figure for many studies using both lines. The figure shows three plots of stacked R_{21} as a function of TIR-to-CO(1–0) and TIR-to-CO(2–1). In the left-hand panel, we show R_{21} as a function of TIR-to-CO(1–0), stacking by TIR-to-CO(1–0). In the middle panel, we show R_{21} as a function of TIR-to-CO(2–1), stacking by TIR-to-CO(2–1). In both of these stacks, the quantity being stacked correlates with the quantity used for stacking. However, the correlation with the CO(2–1) integrated brightness temperature is much weaker. In the right-hand panel, we show R_{21} as a function of TIR-to-CO(2–1), but now stacked by radius, an independent quantity, to remove this bias in the stacks. The underlying trend appears to be a moderate positive correlation between R_{21} and TIR-to-CO(2–1), consistent with the results in Fig. 6. The inner, high density parts of galaxies both show higher star formation per unit gas and higher excitation.

Taken together, the CO-bright, IR-bright, high 70–160 μm ratio regions of our targets show moderately higher R_{21} than the cooler, fainter regions. These trends appear significant, with most galaxies showing a trend with these external parameters. The overall magnitude of the trends is a ∼20−30 per cent change in the ratio across the sample.

Perhaps surprisingly, these trends appear *internal* to galaxies. They do not appear to explain the observed galaxy-to-galaxy offset in R_{21}. They can explain some of the internal radial and azimuthal variations observed. The residual galaxy-to-galaxy offsets must either be driven by different physics or be due to flux calibration uncertainties.

We also note, that while R_{21} in NGC 5194 shows discrepancies in spatial variation with previous findings (Koda et al. 2012), the trends discussed in this subsection are actually in agreement. The discrepancy is mostly due to differences in the fainter, interarm region, thus an agreement in the trends with environmental parameters spanning the entire galactic radial range is not unexpected.

3.5 TIR-to-CO ratio and R_{21}

Many CO surveys over the last two decades have focused on measuring the gas depletion time, star formation scaling relations, or related quantities. In these studies, the figure of merit is often the SFR per unit molecular gas mass. Both CO(1–0) and CO(2–1) line emissions are commonly used to estimate the molecular gas mass. We use a simple observational proxy, the TIR-to-CO ratio, to explore how R_{21} depends on the SFR per unit molecular gas. For the bin width, we chose 0.3 in logarithmic space.

In Fig. 7, we plot R_{21} stacked by the TIR-to-CO(1–0) ratio, the TIR-to-CO(2–1) ratio, and galactocentric radius. We explore all three stacks because of the correlated nature of the axes. We might expect an artificial correlation between R_{21} and TIR-to-CO(1–0) when stacking by TIR-to-CO(1–0). In noisy or scattered data, low CO(1–0) data points will scatter to both high R_{21} and high TIR-to-CO(1–0) values, potentially creating an artificial correlation. A similar effect could introduce an artificial anticorrelation comparing R_{21} to TIR-to-CO(2–1). Because the stacking approach uses values for individual data points to assign them to bins, it will not necessarily reduce this effect via averaging.

In the left-hand panel of Fig. 7, we plot R_{21} stacked by the TIR-to-CO(1–0) ratio. The profiles show a clear positive correlation between R_{21} and TIR-to-CO(1–0) for all galaxies. Correlated axes could, however, be tilting the trend in this direction.

In the central panel of Fig. 7, we instead plot R_{21} as a function of TIR-to-CO(2–1), binned using the TIR-to-CO(2–1) ratio. That is, we change the line used for the stack. Again, the correlated axes potentially affect the stack, this time producing a mild correlation between R_{21} and TIR-to-CO(2–1) for most galaxies.

With this in mind, the right-hand panel of Fig. 7 where R_{21} is plotted as a function of the TIR-to-CO(2–1) ratio, i.e. the same axes as in the left-hand panel, but now stacked by galactocentric radius. Radius represents an independent variable that should minimize bias in the stacks. This figure shows more scatter and a somewhat smaller dynamic range compared to the previous two stacks. When stacking by radius, there is an overall tendency for TIR-to-CO(2–1) and R_{21} to be positively correlated. The behaviour is less universal than we saw when stacking by TIR-to-CO(2–1).

Overall, this result appears consistent with the results in the previous section. We tend to find high R_{21}, TIR-to-CO(1–0), 70–160 μm ratio, and CO brightness temperature in the inner parts of galaxies.

4 DISCUSSION

4.1 Comparison to single-pointing literature measurements

In order to compare our results to literature values, we compiled and homogenized single-pointing CO observations from a number of publications (Boselli et al. 1994; Wiklind, Combes & Henkel 1995; Chini, Kruegel & Lemke 1996; Leon, Combes & Menon 1998; Lavezzi et al. 1999; Curran, Aalto & Booth 2000; Böker, Lisenfeld & Schinnerer 2003; Albrecht et al. 2004; Strong et al. 2004; Evans et al. 2005; Albrecht, Krügel & Chini 2007; Combes, Young & Bureau 2007; Ocaña Flaquer et al. 2010; Cappellari et al. 2011). To create a set of reference measurements:

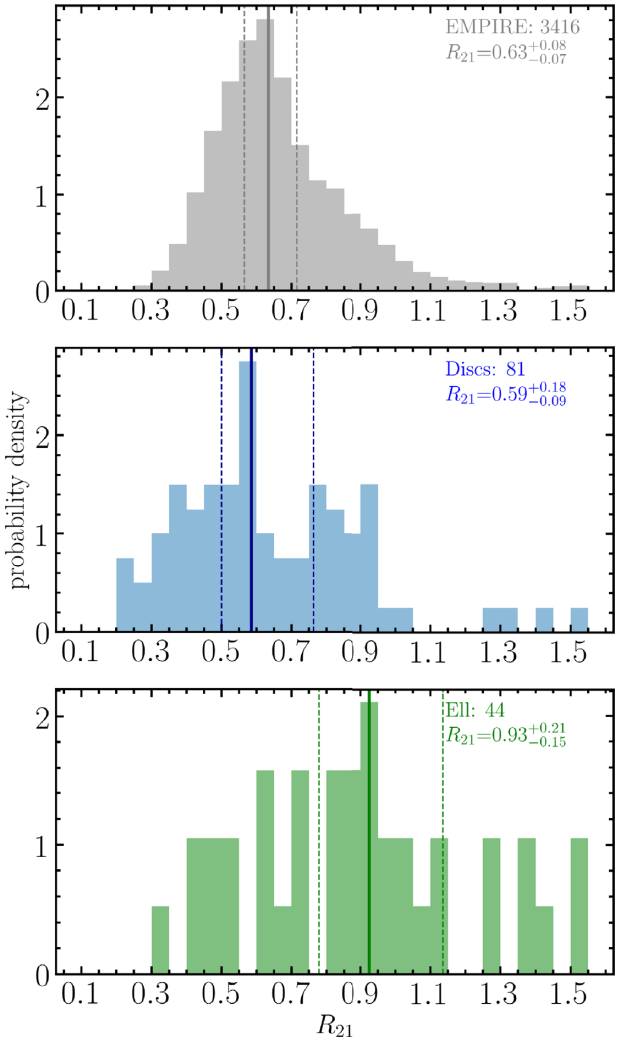

Figure 8. Distributions of R_{21} from literature studies. The distribution of R_{21} for individual lines of sight in EMPIRE (top panel) and a compilation of R_{21} estimates from the literature broken into late-type disc (middle panel) and early-type elliptical (bottom) galaxies (Data from Boselli et al. 1994; Wiklind et al. 1995; Chini et al. 1996; Leon et al. 1998; Lavezzi et al. 1999; Curran et al. 2000; Böker et al. 2003; Albrecht et al. 2004, 2007; Strong et al. 2004; Evans et al. 2005; Combes et al. 2007; Ocaña Flaquer et al. 2010; Cappellari et al. 2011).

(i) We first tabulate CO(1–0) and CO(2–1) line brightness temperatures, errors, and beam sizes for 659 galaxies drawn from the references above. When necessary, we converted from antenna temperature scale to main beam temperature scale using the efficiencies provided in the respective publication. Note that we further limited our compilation of measurements using the criteria below. Our final number of reference galaxies in Fig. 8 is 125.

(ii) For each target, we obtained optical blue band 25 mag arcsec^{-2} isophotal diameters (D_{25}), axial ratios from RC3 (de Vaucouleurs et al. 1991), and morphological types through the *NASA Extragalactic Database* (NED).

(iii) We filtered the data to only include significant detections (S/N > 3) where the beam covered an appreciable part of the galaxy. Some values in the literature are upper limits only, and we do not consider these here. We further require that the full width at half-maximum of the smaller beam, typically CO(2–1), covers at least 40 per cent of a CO scale length (see next point). This typically amounts to requiring that the CO beam covers at least 0.1 D_{25}.

(iv) From D_{25}, we estimate a CO scale length (r_{CO}), assuming: $r_{CO} = 0.23\,D_{25}/2$ (Young et al. 1995; Leroy et al. 2008, 2009; Lisenfeld et al. 2011; Davis et al. 2013). From this scale length and the known beam sizes, we assume an exponential disc and performed an aperture correction to estimate the full luminosity of the galaxy in each line (see e.g. Puschnig et al. 2020). Note that because we focus on the line ratio, the accuracy of the extrapolation to the full galaxy is not important. Only matching the effective area covered by the beams matter.

(v) We calculate R_{21} as the ratio of the estimated full-galaxy CO(2–1) luminosity to the full-galaxy CO(1–0) luminosity. After the cuts based on signal-to-noise ratio and extent, this leaves us with 125 measurements, 81 for late-type ('disc') galaxies and 44 for early-type ('elliptical') galaxies.

Fig. 8 shows histograms of these literature R_{21} measurements. In the figure, we divide the literature sample into disc-like and elliptical galaxies. For disc galaxies we find $R_{21} = 0.59^{+0.18}_{-0.09}$. This agrees well with our results for the EMPIRE sample, though the literature distribution appears much broader. Some of this additional scatter likely reflects uncertainty in calibration. Another part of the scatter reflects the comparatively lower signal-to-noise ratio of these data compared to EMPIRE. A full meta-analysis disentangling the sources of physical and observational scatter for the literature measurements is beyond the scope of this paper. Here, we emphasize the general good agreement between our smaller set of high-quality, resolved measurements and the literature.

The lower panel in Fig. 8 shows that elliptical galaxies exhibit higher excitation. This might be expected if the deeper potential well leads to wider line profiles and thus lower optical depth. Regardless of the explanation, literature observations of early-type galaxies indicate higher apparent R_{21} but also enormous scatter.

Recently, Saintonge et al. (2017) studied R_{21} in 28 galaxies that are part of the xCOLDGASS galaxy sample. Combining IRAM and APEX observations they found a mean ratio of $R_{21} = 0.79 \pm 0.03$ with scatter of 0.15–0.23. xCOLDGASS includes both disc-like and elliptical galaxies, so the relevant comparison is to our full compilation. For all literature data, we find a mean $R_{21} = 0.72$ with a scatter of ± 0.15. This agrees reasonably well with the xCOLDGASS results, especially given the heterogeneous nature of the literature data. The EMPIRE results have lower mean R_{21} compared to the xCOLDGASS IRAM–APEX overlap sample. As discussed by Saintonge et al. (2017), this may somewhat reflect the central focus of the xCOLDGASS pointings. Or it may reflect a greater contribution of early-type galaxies to their sample. Future larger mapping surveys will be needed to help synthesize our knowledge of resolved and galaxy-integrated R_{21}.

4.2 Comparison to previous mapping results

Combining HERACLES with previously existing CO(1–0) data, Leroy et al. (2009) initially found a mean $R_{21} \approx 0.8$ with evidence for central enhancements. Subsequently, improved main beam efficiencies for the IRAM 30-m telescope became available and Usero et al. (2015) carried out pointed spectroscopy using the IRAM 30-m that obtained improved CO(1–0) comparison data. Based on comparing HERACLES to the Usero et al. (2015) data and a collection of earlier CO(1–0) measurements, Leroy et al. (2013) found a median $R_{21} \approx 0.67$ with a large scatter of 0.16 dex or ~40 per cent among individual measurements. Our mean R_{21} has an almost identical value to that in

Leroy et al. (2013), but the measured scatter using EMPIRE CO(1–0) is smaller. This likely reflects the much better calibration using EMIR compared to the archival CO(1–0) data, though the smaller sample size may also play a role. Also using HERACLES, but now attempting to homogenize literature mapping data, Rosolowsky et al. (private communication) found that a ratio of ∼0.7 was typical of the inner parts of galaxies, while 0.5 was more common in outer discs (see also Rosolowsky et al. 2015). Our results yield slightly higher R_{21} at large galactocentric radius (see Fig. 4). Again, we expect the EMPIRE CO(1–0) data to be of higher quality than the HERACLES maps, but the EMPIRE sample size is small.

Our results also agree reasonably well with previous mapping-based results for other nearby galaxies. For example, Crosthwaite & Turner (2007) found $R_{21} \approx 0.8$ in the central 1 arcmin × 1 arcmin of NGC 6946, while we find a value of $R_{21} = 0.7$ for the same galaxy centre. Our measured value of 0.7 for NGC 5194 (M51) agrees with the typical ratio found by Koda et al. (2012) while studying the arm/interarm contrast of the ratio. Other resolved mapping results include $R_{21} \approx 0.8$ for M33 (Druard et al. 2014) with no obvious radial trends. Lundgren et al. (2004) and Koda et al. (2020) found $R_{21} \approx 0.77$ for M83 which closely resembles the EMPIRE targets in morphology and stellar mass, showing a similar decreasing trend in R_{21} with galactocentric radius. Again, our EMPIRE results tend towards the low side of the literature value, but well within the previously measured range.

4.3 R_{21} variations in EMPIRE

We measure the characteristic value, scatter, and dependence of R_{21} on environment. We do find scatter in R_{21} from galaxy-to-galaxy and within galaxies. In certain galaxies, we identify significant correlations with galactocentric radius and other observable quantities.

One recurring theme in our analysis is that the magnitude of these variations is weak. Put simply, the dynamic range in R_{21} across our sample remains small compared to many of the uncertainties associated with measuring the ratio. The small magnitude of these variations somewhat diminish the utility of R_{21} as a diagnostic of the physical conditions in the gas. Of course, CO(2–1) and CO(1–0) both still represent the most widely used tracers of molecular gas at low redshift. Detailed knowledge of how R_{21} behaves is crucial to our knowledge of molecular gas in galaxies.

With that caveat in mind, we discuss the major sources of R_{21} variation in EMPIRE:

Observed galaxy-to-galaxy scatter: In each of our analyses, galaxy-to-galaxy scatter in R_{21} appears to play a role. The galaxy being considered appears to matter independent of radial gradients, correlation with local conditions, or azimuthal variations.

We checked for correlations between global galaxy properties and R_{21} that might explain the galaxy-to-galaxy scatter, including comparing to stellar mass, SFR, metallicity, inclination, distance, and morphological type. We found no significant correlation that could explain the observed galaxy-to-galaxy scatter. We emphasize, however, that EMPIRE represents an extremely small sample with a limited range of stellar mass, SFR, and metallicity values. EMPIRE is fundamentally a mapping project, not a representative survey of the local galaxy population.

One plausible explanation for much of the observed galaxy-to-galaxy is uncertainty in the flux calibration, which we discuss in Section 2.3 and Appendix C. Given the estimated uncertainties in the amplitude calibration of each data set, ∼6 per cent for EMPIRE, ≈ 5 per cent for PHANGS-ALMA and M51 Large Program, and ∼20 per cent for HERACLES, we expect ∼7−20 per cent scatter in R_{21} based on calibration uncertainty alone. We measure rms scatter of 10−15 per cent from galaxy-to-galaxy, so it seems highly likely that much of the galaxy-to-galaxy scatter that we observe is caused by flux calibration uncertainties.

Building a quantitative understanding of galaxy-to-galaxy variations in R_{21} places strong requirements on the data. Given the small dynamic range in the ratio, one needs high signal-to-noise ratio and high precision absolute flux calibration. To avoid uncertain aperture corrections, one needs to observe and cover the same area in both lines. Although EMPIRE and the IRAM–APEX subset from xCOLDGASS (Saintonge et al. 2017) represent good first steps, obtaining such carefully calibrated, high signal-to-noise data sets still represents a future goal.

Trends within galaxies: Within galaxies, we find a clear, but weak systematic variation of R_{21} as a function of environment. We examined correlations with CO brightness temperature, TIR surface brightness, IR colour, and TIR-to-CO ratio. After accounting for biases and disregarding low signal-to-noise regions, these all show the same trend, i.e. higher R_{21} values in regions with higher gas surface density, hotter dust, and more star formation.

Galaxy centres: One major driver for these trends is that we observe a higher R_{21} in the centre of galaxies compared to the discs. The average enhancement is 15 per cent compared to the galaxy-wide luminosity-weighted mean, but several individual cases show much stronger nuclear enhancements. NGC 2903, NGC 327, NGC 4321, and NGC 5055 all show strong (∼50 per cent) central enhancements in R_{21}.

Though not uniquely associated with bars, these central enhancements do seem strongest in the strongly barred members of our sample. NGC 2903, NGC 3627, and NGC 4321 all have prominent bars that visibly interact with the molecular gas. In these cases our coarse resolution likely causes us to underestimate the strength of the R_{21} enhancement, because the nuclear star-forming regions, where we expect the R_{21} enhancements to be strongest, are compact (often ∼0.5 kpc) compared to our 27 arcsec (∼1–2 kpc) beam.

In unbarred galaxies, we often observe flatter R_{21} profiles, e.g. in NGC 628, NGC 3184, or NGC 4254. NGC 5194 (M51) remains an ambiguous case, with our newer EMIR mapping data showing evidence for a flatter R_{21} profile than the HERA maps (see Appendix C).

A similar drop in R_{21} from the centre of the galaxies towards their discs has also been found by many previous studies. Milky Way studies show values close to unity in the central kpc of the Galaxy, dropping to 0.75 at 4 kpc and to ∼0.6 at 8 kpc from the Galaxy centre (Sakamoto et al. 1997; Sawada et al. 2001). Similarly, Casoli et al. (1991b) report a value of ∼1 in the nuclei of nearby spirals (at ∼500 pc scales) compared to 0.5–7 in their discs. Studying the nearby spiral IC 342, Eckart et al. 1990) found a drop from $R_{21} \approx 1.1$ to $R_{21} \approx 0.7$–0.8 around 500 pc from the starburst nucleus. A high average ratio of ∼0.9 is also found by Braine & Combes (1992) in the central kpc of 36 nearby galaxies. Similar radial trends have been found by Saito et al. (2017) when studying the spatially resolved R_{21} ratio in NGC 1614. Using RADEX modelling they find a radial kinetic temperature gradient that mirrors the observed R_{21} trend. Using HERACLES with lower quality CO(1–0) data but a larger sample, Leroy et al. (2009, 2013) noted a similar trend in resolved maps of nearby galaxies.

Furthermore, variation in the CO line ratio could be driven to some parts by the presence of an active galactic nuclei (AGNs) within the galaxy. Four of the galaxies in our sample, NGC 3627, 4321, 5055, and 5194, are know to host an AGN.

Correlation with physical conditions: We observe positive correlations of R_{21} with CO brightness temperature, TIR surface

brightness, and IR colour. All of these quantities tend to decrease with increasing galactocentric radius, so these trends likely express the same underlying physics as the radial gradients.

Physically, the IR colour reflects the interstellar radiation field heating the dust. At the typical densities associated with molecular clouds, gas and dust have different temperatures and are not collisionally coupled (e.g. Draine 2011). However, the radiation field traced by the dust temperature should also relate to the radiation field illuminating molecular clouds and so indirectly relate to excitation of the gas. We would expect more intensely illuminated clouds to have high temperatures and be more nearly thermalized.

Similarly, the TIR surface density traces the heating of the ISM because it indicates the amount of reprocessed, mostly ultraviolet emission. High TIR surface density may also indirectly trace gas density, because high gas densities tend to be associated with high SFRs (see e.g. more discussion in Jiménez-Donaire et al. 2019). We would expect denser, higher temperature gas to be more nearly thermalized and show a higher R_{21}.

As mentioned above, optical depth effects may also play a role. The line ratio of optically thin gas exceeds unity and a component of diffuse, optically thin gas will drive R_{21} to higher values.

A systematic dependence of R_{21} on, e.g. $\Sigma_{\rm TIR}$ has implications for the slope of derived scaling relations. For example, the scaling relation between $\Sigma_{\rm TIR}$ and $I_{\rm CO}$ corresponds to the molecular version of the Kennicutt–Schmidt law. If R_{21} varies systematically with $\Sigma_{\rm TIR}$ then one expects to derive different slopes if using CO(1–0) or CO(2–1).

Our results show that this is the case, but also that the effect is modest. For example, in Fig. 6, R_{21} changes by ∼40 per cent as $\Sigma_{\rm TIR}$ changes by a factor of 100. This would translate to a difference in slope of ∼0.07 for a power law relating the two quantities. It seems reasonable to infer that using CO(2–1) instead of CO(1–0) will change the slope of the $I_{\rm CO}$–$\Sigma_{\rm TIR}$ relation by 0.05–0.1.

We caution that the implications of R_{21} for the underlying physical scaling relation, e.g. between $\Sigma_{\rm mol}$ and $\Sigma_{\rm SFR}$, are less clear. The sensitivity of R_{21} to these local physical conditions implies that physical conditions in the molecular gas are changing. Variations in density, temperature, and optical depth will imply changes in the CO-to-H_2 conversion factor, $\alpha_{\rm CO}$, for both CO(1–0) and CO(2–1). Unfortunately, on its own R_{21} does not heavily constrain $\alpha_{\rm CO}$. Future work using a large set of lines and independent constraints on $\alpha_{\rm CO}$ will help map R_{21} variations to $\alpha_{\rm CO}$ variations for both lines.

Scatter at fixed radius and resolved patterns: Density and radiation field also vary at fixed galactocentric radius, e.g. due to the effects of spiral density waves and stellar bars. The arm–interarm contrast and small-scale structure of R_{21} have been the focus of several recent papers (Koda et al. 2012; Law et al. 2018). Though our ∼1–2 kpc resolution limits our ability to isolate small-scale variations in R_{21}, we attempt to quantify the scatter in R_{21} at fixed galactocentric radius in our sample using a forward modeling technique.

NGC 5194 (M51) shows the largest intrinsic scatter of any target, presumably due to its well-defined grand-design structure. Past studies have already highlighted a strong arm–interarm contrast in the CO line ratio in M51 (Koda et al. 2012; Vlahakis et al. 2013). The contrast is also strongly visible in our analysis, but we find an opposite trend (we find a high interarm and low arm R_{21} ratio; see Section 3.3 and Appendix E). Among the literature there is disagreement between the relative and quantitative trend of R_{21} between arm versus interarm. Interestingly, when studying NGC 6946, Crosthwaite & Turner (2007) found $R_{21} > 1$ in the interarm as opposed to smaller values in the molecular arm regions. While we cannot confirm such large average absolute values with our observations, the regions that show enhancement in our data overlaps with theirs (we find an average $R_{21} \sim 0.9$ in the interarm, with 35 per cent of the points in that region with S/N > 3 showing $R_{21} > 1$, and only 8 per cent have $R_{21} > 1.2$). Furthermore, Crosthwaite et al. (2002) and Lundgren et al. (2004) have investigated M83 and both report higher line ratio values in the interarm region as well. However, the validity of this trend within M83 has recently been disputed by Koda et al. (2020), who studied the source with ALMA observations. Differences of R_{21} between arm and interarm have been found in several galaxies, however, different studies have presented opposing trends. We believe that this is caused at least partially by coarse spatial resolution and insufficient data quality, and should be investigated in more detail in the future.

From point of view of the data, our study differs from Koda et al. (2012) in that we use both new CO(1–0) and CO(2–1) maps obtained using the IRAM 30-m EMIR receiver. The primary difference appears to come from the use of the new CO(1–0) map. We defer a detailed comparison among M51 data sets to the presentation of the new IRAM LP (den Brok et al., in preparation) and new Submillimeter Array observations (Jimenez Donaire et al., in preparation). As emphasized in Section 2.2 and Appendix E, we use what we consider the best available map.

Interpreting in terms of ISM physics and molecular cloud conditions, our finding of enhanced R_{21} in interarm regions as compared to arm regions implies more excited, perhaps more diffuse, warm and optically thin gas in the interarm regions. In the interarm region, the heating is most likely more efficient due to different cloud composition. Furthermore, Lundgren et al. (2004) suggest that photon-dominated regions (PDRs) around cool stars could be responsible for bright CO emission. This is, because the PDR radiation field is softer, thus the CO can be heated photoelectrically at lower A_V (Spaans et al. 1994). Another possibility would be 'CO-loud' gas (Lundgren et al. 2004). Small amounts of optically thin gas could already cause strong emisivity in CO(2–1) (Wiklind et al. 1990). These explanations might be consistent with the extended diffuse component identified in M51 by Pety et al. (2013). They find that ∼50 per cent of the total CO emission originates from larger spatial scales (>1.3 kpc), which would be consistent with emission from a diffuse disc of gas at a scale height of ∼200 pc. A similar fining was made by Caldú-Primo & Schruba (2016) studying M31. The large scatter found may also reflect the influence of M51's ongoing interaction with the companion galaxy NGC 5195.

Uncertain behaviour at large radii: Despite our use of spectral stacking, many of our R_{21} estimates remain uncertain at low brightness temperature and large radius. In the stacked radial profiles, we see suggestions of large deviations to both low and high R_{21} in some of our targets. Similarly, in the lowest brightness temperature bins of IR colour and TIR surface brightness we see hints of significant deviations. It could well be that molecular gas in the outer parts of galaxies is either optically thin, leading to high R_{21}, or cool, leading to low R_{21}. More sensitive observations of both lines will be required to ascertain the behaviour of the ratio in the faint CO emission from the outer discs of galaxies.

4.4 Comparison to radiative transfer models

Following up the work of Leroy et al. (2017), Puschnig et al. (in preparation) have established a set of molecular radiative transfer models, i.e. the `Dense Gas Toolbox` (Puschnig 2020), which predicts line ratios for a medium with an underlying density distribution (e.g. a lognormal distribution rather than from a single density).

Using CO line optical depths as previously published by Cormier et al. (2018) for EMPIRE galaxies (they find $\tau^{12} \approx 6$ for 12CO(1–0)), we now examine the impact of three physical quantities on R_{21}: temperature, mean density, and width of the lognormal density distribution. The interplay between these quantities can be studied through an interactive tool.[3] The models show that R_{21} is most sensitive to regimes with mean densities lower than $\sim 10^3$ cm^{-3}. Below that value all three quantities are degenerate. However, above that density the line ratio may only be driven further up by higher temperatures, regardless of the width of the density distribution (that is proportional to line width or Mach number). We also recognize that values of $R_{21} > 0.8$ are only predicted for temperatures above 35 K, regardless of the mean gas density. For NGC 5194, we may thus conclude that the mean gas density and temperature must be $n_{\rm H_2} > 10^3$ cm^{-3} and $T_{\rm kin} > 35$ K, throughout the whole disc.

5 SUMMARY

We measure the ^{12}CO(2–1)/^{12}CO(1–0) brightness temperature ratio, R_{21}, across the star-forming discs of nine nearby galaxies. We measure CO(1–0) emission from maps obtained by the IRAM 30-m telescope in the context of the EMPIRE survey (Bigiel et al. 2016; Jiménez-Donaire et al. 2019) and CO(2–1) emission from a mixture of ALMA and IRAM 30-m data (den Brok et al., in preparation). We use IRAM 30-m CO (2–1) maps obtained as part of HERACLES (Leroy et al. 2009) and a new IRAM Large Program targeting M51. We use ALMA maps obtained as part of the PHANGS-ALMA survey (Leroy et al. 2021b). We measure the distributions and mean values of R_{21} across individual lines of sight, integrated over galaxies, stacking by radius, and stacking as a function of other local conditions. Our main results are:

(i) The luminosity-weighted mean R_{21} for individual galaxies ranges from 0.48 to 0.73. Within individual galaxies, we observe a typical range of ± 0.1. Over the whole sample, treating galaxies equally we find a mean R_{21} of 0.63 ± 0.09.

(ii) We compiled and homogenized a set of CO observations from the literature. For 81 disc galaxies, these literature measurements yield $R_{21} = 0.59^{+0.18}_{-0.09}$, in good agreement with our mean value.

(iii) Seven of our nine targets show a central enhancement in R_{21} compared to the disc-averaged value (median enhancment $\sim$15 per cent). The magnitude of the deviation varies from galaxy-to-galaxy, but variation at larger radii can be much larger than the ones found towards the centre. Both central enhancements and radial gradients in R_{21} are in agreement with previous work.

(iv) We find significant correlations between R_{21}, CO brightness temperature, TIR surface density, and 70–160 μm ratio. All of these have the expected trend of an increasing ratio when the gas density and radiation field increase.

(v) R_{21} also shows azimuthal variation. Using a forward modelling approach, we estimate the intrinsic scatter in R_{21} at fixed galactocentric radius to be $\sim$ 20 per cent at our $\sim$1–2 kpc resolution.

(vi) These physical trends are not sufficient to explain the majority of the galaxy-to-galaxy variations observed. Given the scale of our calibration uncertainties, we cannot completely rule them out as one of the dominant drivers for these trends. Instead, the magnitude of this scatter appears consistent with being driven by absolute flux calibration uncertainties.

[3] http://www.densegastoolbox.com/explorer/

ACKNOWLEDGEMENTS

The IRAM 30-m large programme EMPIRE was carried out under project number 206-14 (PI Bigiel), the ^{12}CO(1–0) observations under projects 061-15 and 059-16 (PI Jiménez-Donaire) and D15-12 (PI Cormier). IRAM is supported by INSU,CNRS (France), MPG (Germany), and IGN (Spain). MJJD acknowledges support from the Smithsonian Institution as a Submillimeter Array (SMA) Fellow. FB, JPu, AB, and JdB acknowledge funding from the European Union's Horizon 2020 research and innovation programme (grant agreement No 726384/Empire). AU acknowledges support from the Spanish funding grants AYA2016-79006-P (MINECO/FEDER), PGC2018-094671-B-I00 (MCIU/AEI/FEDER), and PID2019-108765GB-I00 (MICINN). The work of AKL and MJG is partially supported by the National Science Foundation under Grants No. 1615105, 1615109, and 1653300. AKL also acknowledges partial support from NASA ADAP grants NNX16AF48G and NNX17AF39G. ES, DL, and ST acknowledge funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 694343). CMF acknowledges support from the National Science Foundation under Award No. 1903946. JMDK gratefully acknowledges funding from the Deutsche Forschungsgemeinschaft (DFG) 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). JPe acknowledges support by the Programme National 'Physique et Chimie du Milieu Interstellaire' (PCMI) of INSU,CNRS with INC/INP, co-funded by CEA and CNES. Furthermore, we thank J. Koda for making available the NRO 45-m CO(1–0) data of NGC 5194.

DATA AVAILABILITY

The HERACLES and EMPIRE survey data used in this article are publicly available in IRAM repository, at https://www.iram.fr/ILPA/LP001/ and https://www.iram.fr/ILPA/LP015/, respectively. The PHANGS-ALMA CO maps will be available as part of the first public data release in the first half of 2021 and will be available from the ALMA archive and https://www.phangs.org. The M51 IRAM Large Program will be publicly available during the first half of 2021 via the IRAM Large Program Archiva. The remaining data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

Albrecht M., Chini R., Krügel E., Müller S. A. H., Lemke R., 2004, A&A, 414, 141
Albrecht M., Krügel E., Chini R., 2007, A&A, 462, 575
Aravena M. et al., 2010, ApJ, 718, 177
Aravena M. et al., 2014, MNRAS, 442, 558
Aravena M. et al., 2016, MNRAS, 457, 4406
Bigiel F. et al., 2016, ApJ, 822, L26
Böker T., Lisenfeld U., Schinnerer E., 2003, A&A, 406, 87
Bolatto A. D., Wolfire M., Leroy A. K., 2013, ARA&A, 51, 207
Bonato M. et al., 2018, MNRAS, 478, 1512
Boselli A., Gavazzi G., Combes F., Lequeux J., Casoli F., 1994, A&A, 285, 69
Bothwell M. S. et al., 2013, MNRAS, 429, 3047
Braine J., Combes F., 1992, A&A, 264, 433
Caldú-Primo A., Schruba A., 2016, AJ, 151, 34

Cappellari M. et al., 2011, MNRAS, 413, 813
Carter M. et al., 2012, A&A, 538, A89
Casoli F., Boisse P., Combes F., Dupraz C., 1991a, A&A, 249, 359
Casoli F., Dupraz C., Combes F., Kazes I., 1991b, A&A, 251, 1
Chini R., Kruegel E., Lemke R., 1996, A&AS, 118, 47
Combes F., Young L. M., Bureau M., 2007, MNRAS, 377, 1795
Cormier D. et al., 2018, MNRAS, 475, 3909
Crosthwaite L. P., Turner J. L., 2007, AJ, 134, 1827
Crosthwaite L. P., Turner J. L., Buchholz L., Ho P. T. P., Martin R. N., 2002, AJ, 123, 1892
Curran S. J., Aalto S., Booth R. S., 2000, A&AS, 141, 193
de Vaucouleurs G., de Vaucouleurs A., Corwin H. G. Jr., Buta R. J., Paturel G., Fouque P., 1991, Third Reference Catalogue of Bright Galaxies. Springer, New York
Daddi E. et al., 2010, ApJ, 713, 686
Davis T. A. et al., 2013, MNRAS, 429, 534
Draine B. T., 2011, Physics of the Interstellar and Intergalactic Medium. Princeton Univ. Press, Princeton, NJ
Draine B. T. et al., 2007, ApJ, 663, 866
Druard C. et al., 2014, A&A, 567, A118
Eckart A., Downes D., Genzel R., Harris A. I., Jaffe D. T., Wild W., 1990, ApJ, 348, 434
Evans A. S., Mazzarella J. M., Surace J. A., Frayer D. T., Iwasawa K., Sanders D. B., 2005, ApJS, 159, 197
Galametz M. et al., 2013, MNRAS, 431, 1956
Gallagher M. J. et al., 2018a, ApJ, 858, 90
Gallagher M. J. et al., 2018b, ApJ, 868, L38
Hasegawa T., 1997, in Latter W. B., Radford S. J. E., Jewell P. R., Mangum J. G., Bally J., eds, Proc. IAU Symp. 170, The CO 2-1/1-0 Ratio. Kluwer, Dordrecht, p. 39
Hasegawa T. et al., 1997, in Okuda H., Matsumoto T., Rollig T., eds, ASP Conf. Ser. Vol. 124, Diffuse Infrared Radiation and the IRTS. Astron. Soc. Pac., San Francisco, p. 244
Herrera C. N. et al., 2020, A&A, 634, A121
Heyer M., Dame T. M., 2015, ARA&A, 53, 583
Jiménez-Donaire M. J. et al., 2017, ApJ, 836, L29
Jiménez-Donaire M. J. et al., 2019, ApJ, 880, 127
Kennicutt R. C. et al., 2011, PASP, 123, 1347
Koda J. et al., 2011, ApJS, 193, 19
Koda J. et al., 2012, ApJ, 761, 41
Koda J. et al., 2020, ApJ, 890, L10
Lavezzi T. E., Dickey J. M., Casoli F., Kazès I., 1999, AJ, 117, 1995
Law C. J., Zhang Q., Ricci L., Petitpas G., Jiménez-Donaire M. J., Ueda J., Lu X., Dunham M. M., 2018, ApJ, 865, 17
Leon S., Combes F., Menon T. K., 1998, A&A, 330, 37
Leroy A. K., Walter F., Brinks E., Bigiel F., de Blok W. J. G., Madore B., Thornley M. D., 2008, AJ, 136, 2782
Leroy A. K. et al., 2009, AJ, 137, 4670
Leroy A. K. et al., 2013, AJ, 146, 19
Leroy A. K. et al., 2017, ApJ, 835, 217
Leroy A. K. et al., 2021a, ApJS, preprint (arXiv:2104.07665)
Leroy A. K. et al., 2021b, preprint (arXiv:2104.07739)
Lisenfeld U. et al., 2011, A&A, 534, A102
Lundgren A. A., Wiklind T., Olofsson H., Rydbeck G., 2004, A&A, 413, 505
Mangum J. G., Emerson D. T., Greisen E. W., 2007, A&A, 474, 679
Mangum J. G., Darling J., Henkel C., Menten K. M., MacGregor M., Svoboda B. E., Schinnerer E., 2013, ApJ, 779, 33
Ocaña Flaquer B., Leon S., Combes F., Lim J., 2010, A&A, 518, A9
Parkin T. J. et al., 2013, ApJ, 776, 65
Peñaloza C. H., Clark P. C., Glover S. C. O., Shetty R., Klessen R. S., 2017, MNRAS, 465, 2277
Peñaloza C. H., Clark P. C., Glover S. C. O., Klessen R. S., 2018, MNRAS, 475, 1508
Pety J. et al., 2013, ApJ, 779, 43
Puschnig J., 2020, Dense Gas Toolbox. Available at: http://www.densegastoolbox.com/ (accessed April 10, 2021)
Puschnig J. et al., 2020, A&A, 644, A10
Rosolowsky E., Leroy A. K., Usero A., Loeppky J., Walter F., Wilson C., Heracles Team N. T., 2015, American Astronomical Society Meeting Abstracts #225, p. 141.25
Saintonge A. et al., 2017, ApJS, 233, 22
Saito T. et al., 2017, ApJ, 835, 174
Sakamoto S., Hayashi M., Hasegawa T., Handa T., Oka T., 1994, ApJ, 425, 641
Sakamoto S., Hasegawa T., Handa T., Hayashi M., Oka T., 1997, ApJ, 486, 276
Sandstrom K. M. et al., 2013, ApJ, 777, 5
Sawada T. et al., 2001, ApJS, 136, 189
Schinnerer E., Weiß A., Aalto S., Scoville N. Z., 2010, ApJ, 719, 1588
Schinnerer E. et al., 2013, ApJ, 779, 42
Schinnerer E. et al., 2017, ApJ, 836, 62
Schruba A. et al., 2011, AJ, 142, 37
Schruba A. et al., 2012, AJ, 143, 138
Schuster K. F., Kramer C., Hitschfeld M., Garcia-Burillo S., Mookerjea B., 2007, A&A, 461, 143
Shirley Y. L., 2015, PASP, 127, 299
Solomon P. M., Vanden Bout P. A., 2005, ARA&A, 43, 677
Spaans M., Tielens A. G. G. M., van Dishoeck E. F., Bakes E. L. O., 1994, ApJ, 437, 270
Strong M., Pedlar A., Aalto S., Beswick R. J., Curran S., Booth R., 2004, MNRAS, 353, 1151
Sun J. et al., 2018, ApJ, 860, 172
Usero A. et al., 2015, ApJ, 150, 115
Vlahakis C., van der Werf P., Israel F. P., Tilanus R. P. J., 2013, MNRAS, 433, 1837
Walter F., Brinks E., de Blok W. J. G., Bigiel F., Kennicutt R. C., Jr, Thornley M. D., Leroy A., 2008, AJ, 136, 2563
Wiklind T., Rydbeck G., Hjalmarson A., Bergman P., 1990, A&A, 232, L11
Wiklind T., Combes F., Henkel C., 1995, A&A, 297, 643
Yajima Y. et al., 2021, PASJ, 73, 257
Yoda T. et al., 2010, PASJ, 62, 1277
Young J. S. et al., 1995, ApJS, 98, 219

APPENDIX A: OVERVIEW OF MAPS

Fig. A1 displays maps of our nine target galaxies. The leftmost column shows $\log_{10}R_{21}$, the CO(2–1)/(1–0) integrated brightness temperature ratio (see Section 3.1). Columns two and three show the integrated CO(1–0) and CO(2–1) brightness temperature maps (see Sections 2.2 and 2.3). Columns four and five show the *Herschel* 70 and 160 μm intensity maps (see Section 2.4). The last column shows the TIR surface brightness (see Section 2.5).[4]

All of the maps have already been convolved to share the same, 27 arcsec angular resolution. They have all been projected on to a hexagonal grid with a grid spacing equal to half the beamsize (13.5 arcsec). The maps in Fig. A1 only show sightlines that have significant ($S/N > 3$) integrated brightness temperature detections in both CO(2–1) and CO(1–0).

[4]NGC 2903 lacks *Herschel* data.

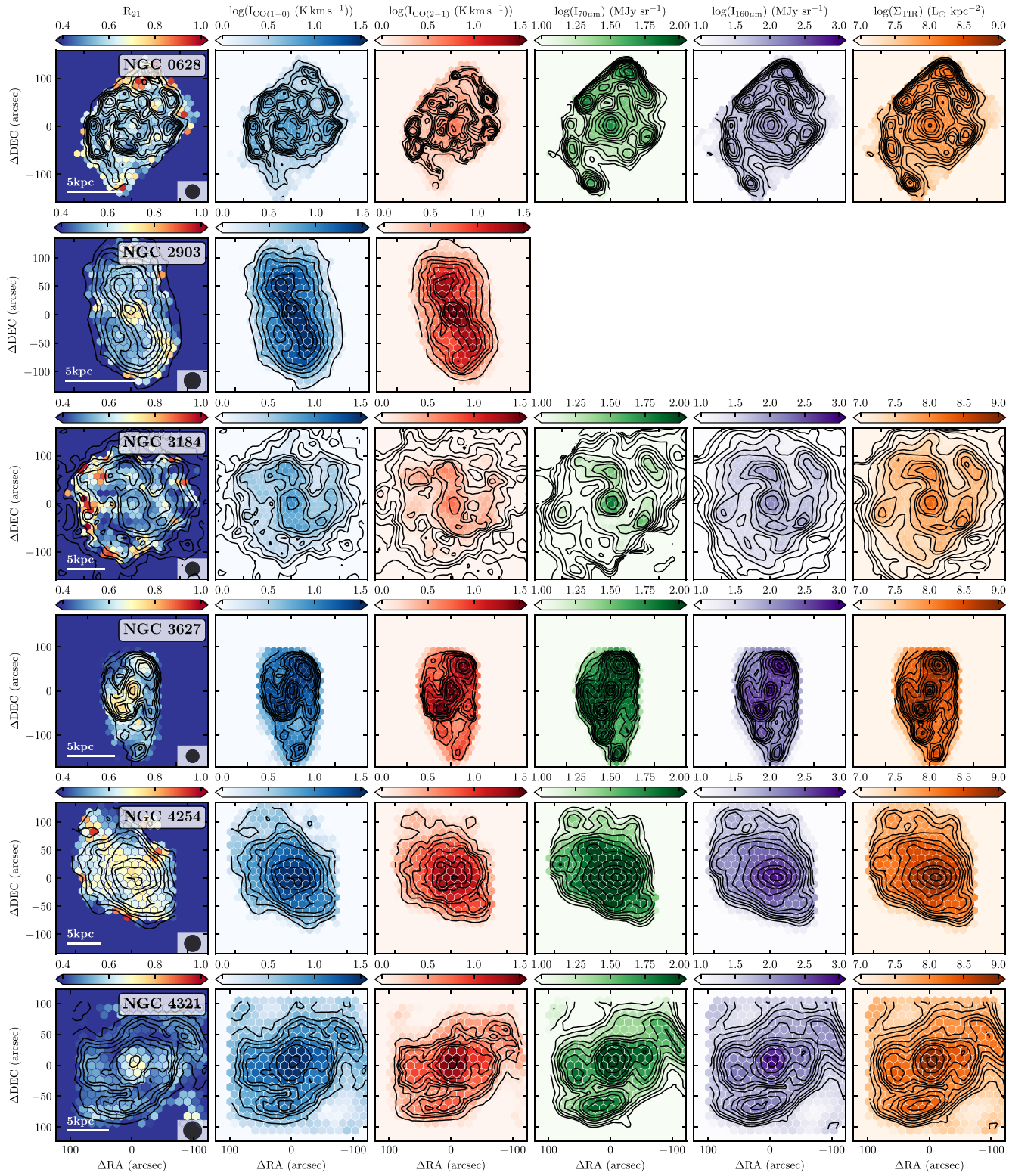

Figure A1. Multiwavelength maps used in this paper. From left to right, each row shows maps of the CO(2–1)/CO(1–0) integrated brightness temperature ratio, $\log(R_{21})$, CO(1–0) and CO(2–1) integrated brightness temperature, *Herschel* 70 and 160 μm fluxes, and the TIR surface brightness. There are no *Herschel* data available for NGC 2903. We plot values for positions that have significant ($S/N > 3$) integrated brightness temperature detections within both the CO(2–1) and CO(1–0) maps. For reference, we overlay CO(1–0) integrated brightness temperature contours on the R_{21} map. The remaining panels have contours of their respective colour scale, in levels of 20, 30, 40, 50, 60, 70, 80, 90, 95, 97.5, and 99.5 of the peak value.

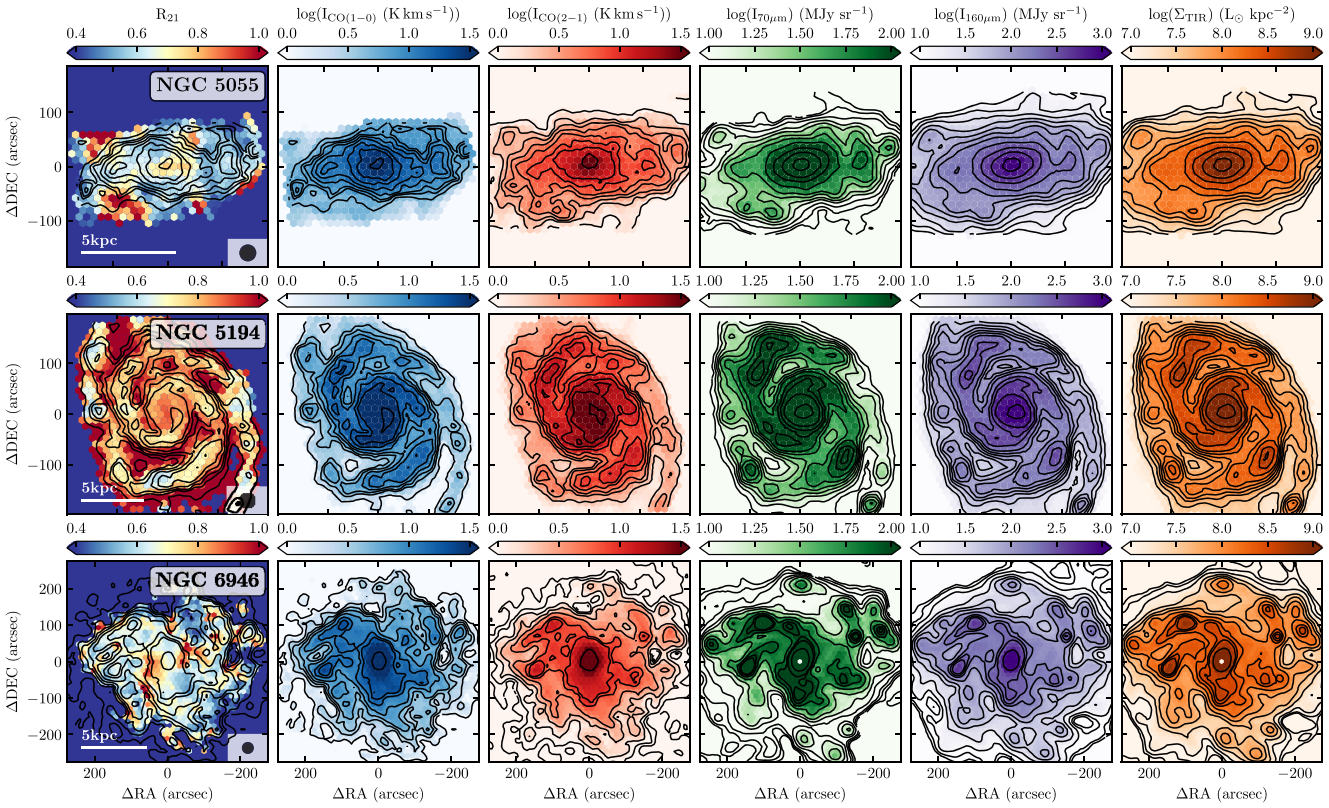

Figure A1 – *continued*

APPENDIX B: FORWARD MODELLING TO INFER PHYSICAL SCATTER IN R_{21}

This section presents our modelling used to estimate the scatter of the R_{21} ratio at a fixed galactocentric radius. In our modelling, we choose to disentangle the intrinsic scatter per radial bin from the observational noise, and model the value of the scatter that best describes our data separately.

The observed scatter in R_{21} reflects a combination of statistical and physical scatter. Fortunately, we have accurate estimates of the statistical uncertainties. To estimate the physical scatter, we carry out a forward modelling analysis that leverages this knowledge.

To do this, we assume that the true physical distribution of R_{21} is lognormal. This appears to be a reasonable assumption based on the observed distributions, e.g. in Fig. 1. The lognormal distribution has strictly speaking no physical meaning, but gives a simple representation of the scatter. Then, our modelling process proceeds as follows:

(i) We normalize all measured R_{21} ratios to the median value of the ratio in the distribution.

(ii) We simulate a set of new data. Each data set has a known physical scatter between 0 and 2 dex. We add Gaussian noise to each new data set based on the known observational uncertainties for the measurement in the data set.

(iii) We compare the observed distribution to the simulated distribution and select the simulated data set that best matches our observational data. To selected the best match, we use the Kolmogorov–Smirnov statistics.

(iv) We adopt the physical scatter in the best-matching model distribution as our best estimate of the true physical scatter.

This estimate accounts for the known scatter due to statistical uncertainties, which can be substantial.

APPENDIX C: CALIBRATION UNCERTAINTIES IN HERACLES

In the main text, we emphasize the importance of knowing the calibration uncertainties for accurate R_{21} estimation. Here, we revisit topics related to the calibration of the HERACLES CO(2–1) maps. Table C1 shows the list of sources for which we have complementary CO(2–1) data besides HERACLES.

HERACLES was obtained using the HERA receiver array on the IRAM 30-m telescope (Schuster et al. 2007). HERA consists of two nine-receiver arrays, one for each polarization, for a total of 18 pixels. HERACLES was calibrated using the standard IRAM 30-m chopper wheel calibration and converted from antenna temperature to main beam temperature using best estimates for the IRAM forward and main beam efficiencies. The bandwidth of HERA does not allow observations of a Galactic line calibrator. The overhead to observe a flux calibrator with every pixel during each few-hour observing block was prohibitive.

Measured gain variations: Leroy et al. (2009) assessed the uncertainty in the HERACLES calibration by building maps from different polarizations and observing sessions. By comparing the intensity of bright-point sources, they estimated an overall calibration uncertainty of 20 per cent.

After that, a more rigorous check was added to the HERACLES pipeline to assess the relative flux calibration of the individual receiver pixels. We took the final cube created from all pixels. Then, we took the location of each observation for each individual receiver

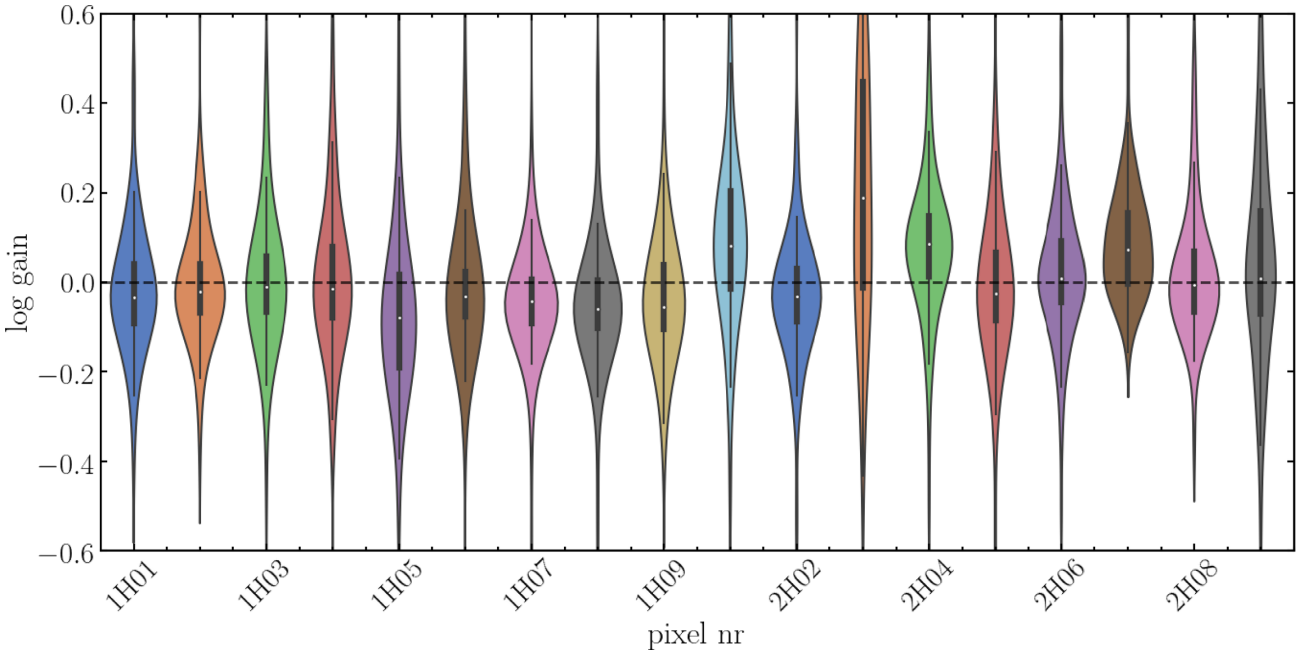

Figure C1. Histograms of log gains per HERA pixel and polarization.

Table C1. Overview of adopted CO(2–1) single dish data sets that we use as complimentary to EMPIRE CO(1–0) for the CO line brightness temperature ratio.

Galaxy	HERA[a]	ALMA[b]	EMIR[c]
NGC 0628	✓	✓	
NGC 2903	✓	✓	
NGC 3184	✓		
NGC 3627	✓	✓	
NGC 4254	✓	✓	
NGC 4321	✓	✓	
NGC 5055	✓		
NGC 5194	✓		✓
NGC 6946	✓		

[a]Part of HERACLES (Leroy et al. 2009).
[b]Part of PHANGS-ALMA-survey (Leroy et al. 2021b).
[c]Part of M51 IRAM 30-m Large Program (den Brok et al., in preparation).

pixel. In this way, we simulated the spectrum that we would expect to observe with that pixel. We compared this expected spectrum to the real observed spectrum for that pixel. Based on this comparison, we calculate the best-fitting multiplicative factor, the 'pixel gain', to match that pixel to the overall cube. The accuracy of the gain measurements is accessible via the χ^2 values obtained from the comparison between simulated and observed spectrum. We measured a gain for each array pixel and observing session, labelling the observing session by the day of the observations.

The measured pixel gain represents the offset in calibration between that pixel and the overall array. This factor does not capture absolute variations in the calibration, it measures how internally well calibrated the pixels are relative to one another.

Fig. C1 shows histograms of the pixel gain for each pixel. We only plot pixel gains with high accuracy, i.e. their χ^2 values lie within $\pm 1\sigma$ of the Gaussian log-χ^2 distribution. Lower signal-to-noise cases typically represent observations of faint galaxies or empty sky and do not contain the signal needed to fit for the pixel gain.

The figure displays that the gain shows typical rms variation of ± 0.10 dex. Some pixels are less stable than others, with the second polarization (HERA2, labelled '2H') showing more scatter than the first polarization.

If the pixel gains were uncorrelated, random, and the coverage of each pixel were spread evenly across the maps, then we expect that the pixel gain uncertainties should average and the calibration uncertainty associated with individual receiver variations would be $\sqrt{18} \approx 4.2$ times lower than the mean individual pixel gain. This represents a lower limit to the calibration uncertainty, which we estimate at ± 0.024 dex or ± 6 per cent.

In fact, the gains do show some correlation, so that there do not appear to be 18 truly independent realizations. As mentioned, the two polarizations often appear offset from one another, with the typical offset on any given day of 0.08 dex.

Based on this, we find an upper and lower limit uncertainty of ± 0.10 and ± 0.024 dex, respectively, corresponding to a flux calibration uncertainty between ± 6 and ± 25 per cent. This will not include any additional terms that are covariant among all pixels, like correction for the atmosphere and beam efficiency effects.

Note that although the HERACLES observing strategy attempted to maximize the number of different pixels observing each part of the sky, local variations in the calibration will be worse due to the fact that not all pixels see all locations.

As an aside, note that we already used these calculated pixel gains to identify and flag the worst receiver-day combinations before producing the maps made publicly available and used in Schruba et al. (2011, 2012), Leroy et al. (2013), and Sandstrom et al. (2013).

Comparisons in Galaxy with two maps: As a more direct, alternative check, we took the overlap between PHANGS-ALMA, HERACLES, and the new IRAM 30-m map of M51 in our sample and directly calculated how these CO(2–1) maps compared to one another (see Figs C2 and C3).

On average, we find consistent results for the mean CO(2–1)/CO(1–0) line ratio when we change the CO(2–1) data set used in the overlapping data set (HERA: $\langle R_{21} \rangle = 0.62 \pm 0.14$; ALMA:

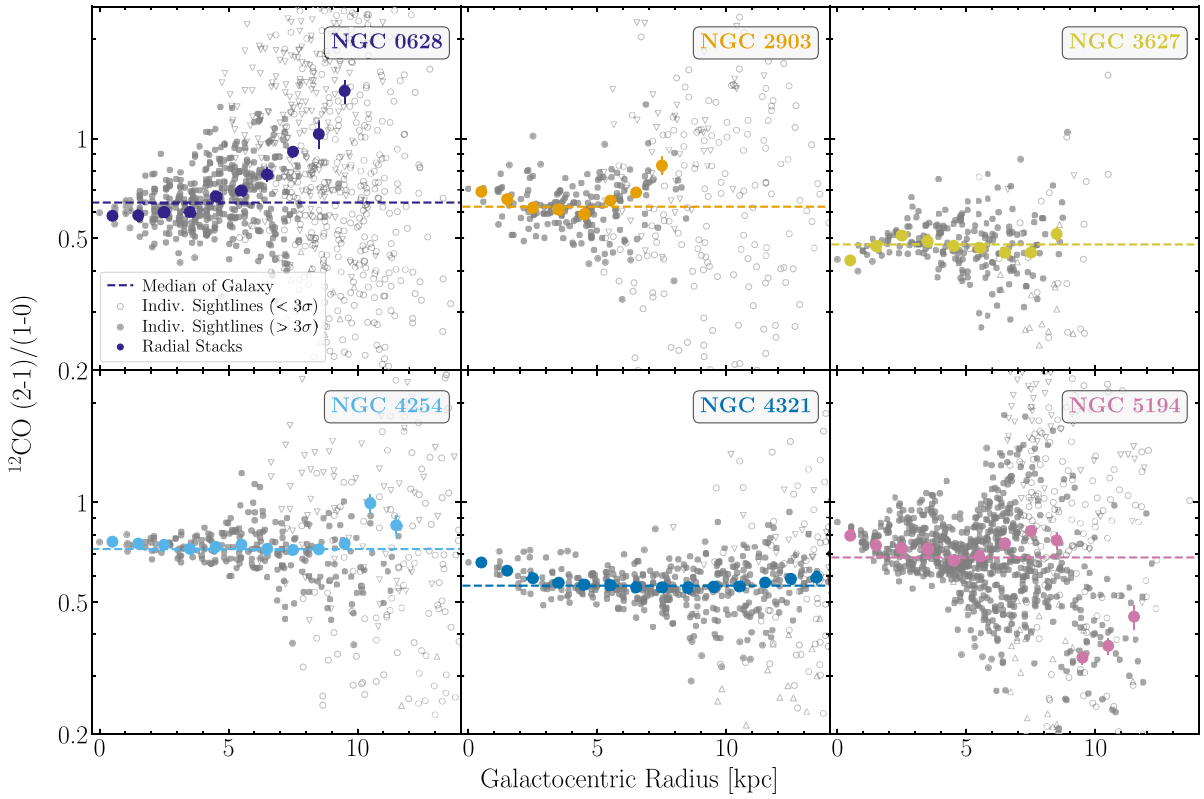

Figure C2. Radial profiles of the CO line brightness temperature ratio, R_{21}, as a function of galactocentric radius for all the galaxies in the EMPIRE sample, similar to Fig. 3. Here, we substituted the CO(2–1) data from HERACLES for ALMA or the new M51 Large Program data. Upper and lower limits of individual sightlines are indicated by upward and downward triangles. We present the stacked values of the ratio per 27 arcsec radial bin. The dashed line gives the mean line ratio within the galaxy.

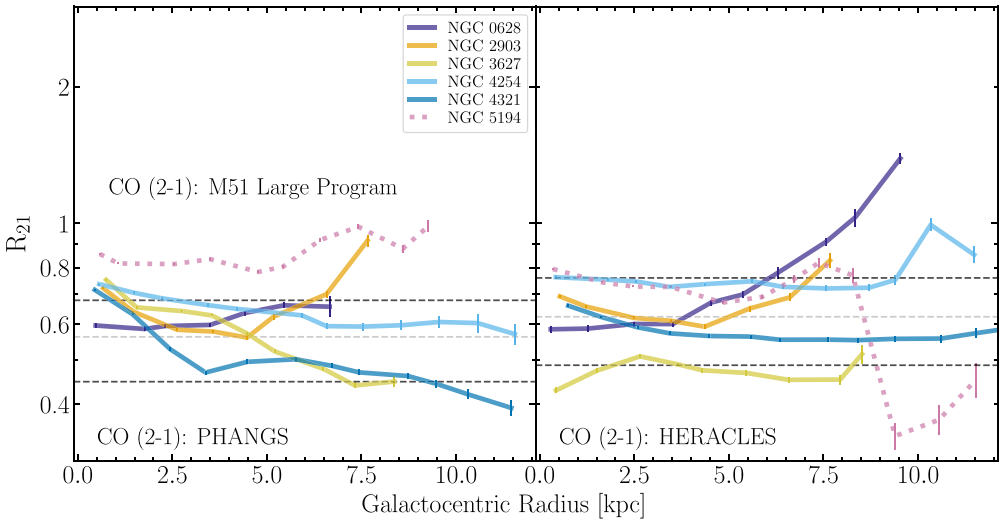

Figure C3. Radial profiles of the CO ratio, R_{21}, as a function of galactocentric radius for all the galaxies in the EMPIRE sample, similar to Fig. 4. The grey, dashed line gives the mean line ratio across all galaxies plotted, while the black, dashed lines indicate the 1σ deviation. For the CO(2–1) data set, HERA data were substituted by ALMA and in the case of NGC 5194 with EMIR 1-mm observations.

$\langle R_{21} \rangle = 0.56 \pm 0.11$). That is, there does not seem to be strong evidence that the overall amplitude scale of HERA is biased relative to ALMA or the new 30-m observations obtained with EMIR.

We do find relatively strong discrepancies in the maps for two galaxies: NGC 3627 and NGC 5194 (see Fig. C4). Compared to the new EMIR map by den Brok (in preparation), the HERACLES map of NGC 5194 is low by a factor of 0.89. Meanwhile the NGC 3627 shows an offset of 0.73 from the ALMA map on average, but also a qualitatively different radial structure.

These were the two earliest galaxies observed with HERA. NGC 5194 was observed as part of commissioning (Schuster et al. 2007) and NGC 3627 as part of a pilot programme that explored the viability of HERACLES. As a result, they did not yet adopt the rotation, cross-scanning, and offset that became part of the later

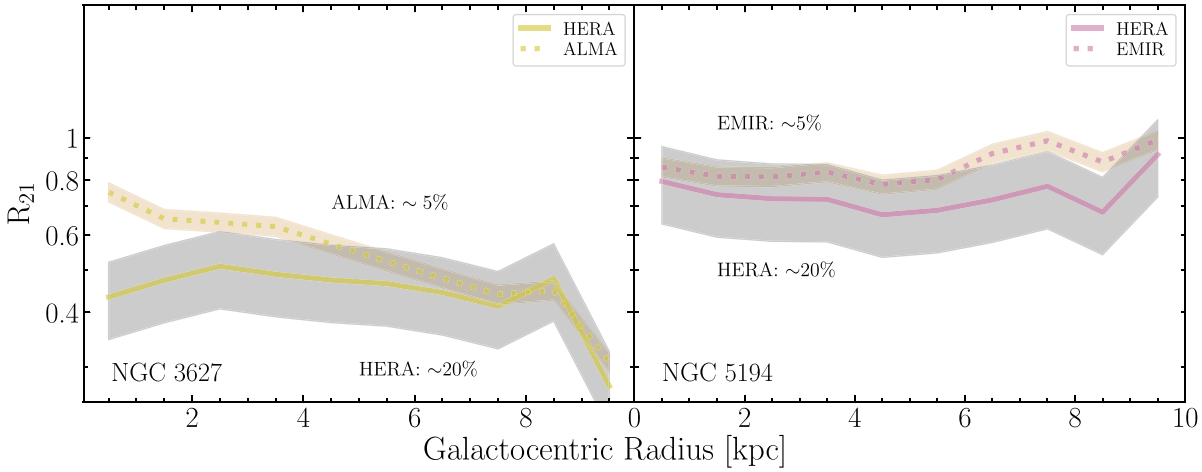

Figure C4. Side-by-side comparison when substituting HERA CO(2–1) with ALMA or EMIR data in the case of NGC 3627 and NGC 5194. The calibrational uncertainties for HERA are estimated to be around 20 per cent (Leroy et al. 2009), while for EMIR (Carter et al. 2012) and ALMA (Bonato et al. 2018) it is around 5 per cent. We note that the HERA maps in particular of NGC 5194 are less reliable, as this constitutes a commission observation under difficult observing conditions. The two cases show the strongest discrepancies when substituting the CO(2–1) data sets.

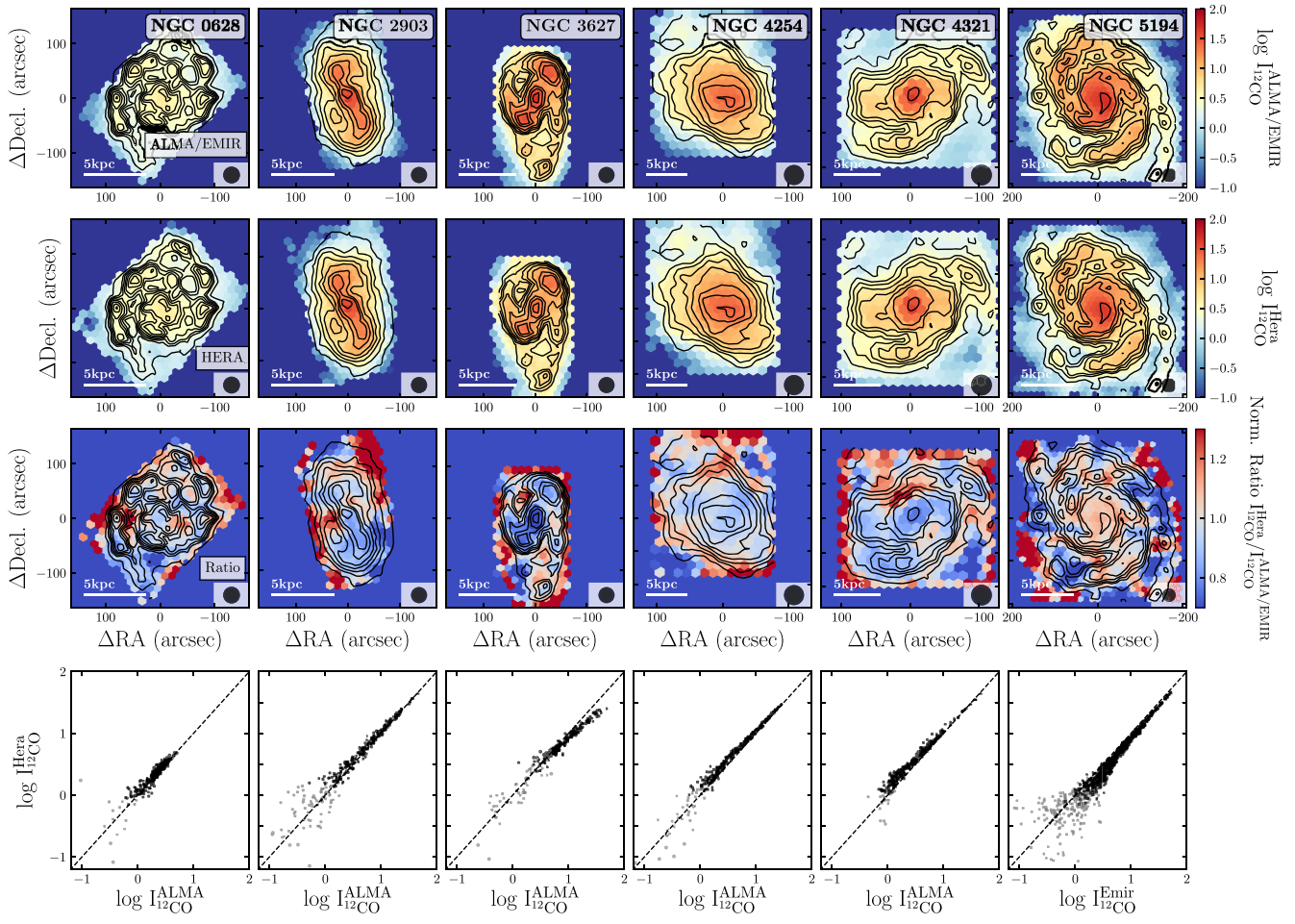

Figure C5. In-depth analysis and comparison of the the sources for which we have two CO(2–1) data sets. The top row shows the latest available CO(2–1) data sets (from ALMA or EMIR in the case of NGC 5194). The second row shows the HERACLES CO(2–1) 2D maps. The third row shows the relative spatial variation of the ratio of the two CO(2–1) observations. In the bottom row, the integrated brightness temperatures of individual sightlines are plotted against each other. The dotted line indicates the 1:1 relation. Black points indicate data points for which both CO(2–1) data are detected with S/N > 3.

HERACLES observing strategy. Our recommendation is that the new EMIR and ALMA maps supersede the HERA data for these targets, and we have adopted this approach in this paper.

For the remaining galaxies with two maps, NGC 628, NGC 2903, NGC 4254, and NGC 4321, we find better agreement. A more detailed comparison is shown in Fig. C5, where the spatial variation in 2D of the ratio of the two different CO(2–1) observations is shown. Overall pairs of CO(2–1) maps mostly show similar morphologies. The global offset in calibration for NGC 3627 and NGC 5194 discussed above is striking. We also see some second-order variations in morphology between the maps, e.g. in the centre of NGC 3627 and NGC 4321. Our best estimate is that these reflect pixel gain variations in HERA, which are inducing second-order local calibration uncertainties.

APPENDIX D: STACKED CO LINE MEASUREMENTS

We apply a stacking technique to improve the S/N when measuring R_{21} as a function of other quantities. This technique is summarized in Section 2.5 and described in more detail by Schruba et al. (2011), Jiménez-Donaire et al. (2017), and Cormier et al. (2018). In Figs D1 and D2, we show one application, the stacked CO(1–0) and CO(2–1) brightness temperature as a function of galactocentric radius in NGC 0628. This illustrates the procedure used to stack other galaxies and to stack by other quantities. It also highlights some of the uncertainty associated with the lowest brightness temperature bins, an issue raised in the main text.

Table D1. Signal-to-noise ratio for the radial stack bins (see Fig. D1 for an example showing the individual radial bins for one galaxy). The median of a given bin over the nine galaxies is taken, as well as the 5 and 95 percentiles. Each radial bin has a width of 1 kpc.

Center radial bin (kpc)	0.5	1.5	2.5	3.5	4.5	5.5	6.5	7.5	8.5	9.5	10.5	11.5
Median SNR	142	110	103	96	89	67	54	40	28	9.5	4.1	10.1
5 perc.	54	60	62	59	63	49	23	20	11	6.5	0.8	0.8
95 perc.	323	215	157	148	97	94	99	78	59	47	36	28

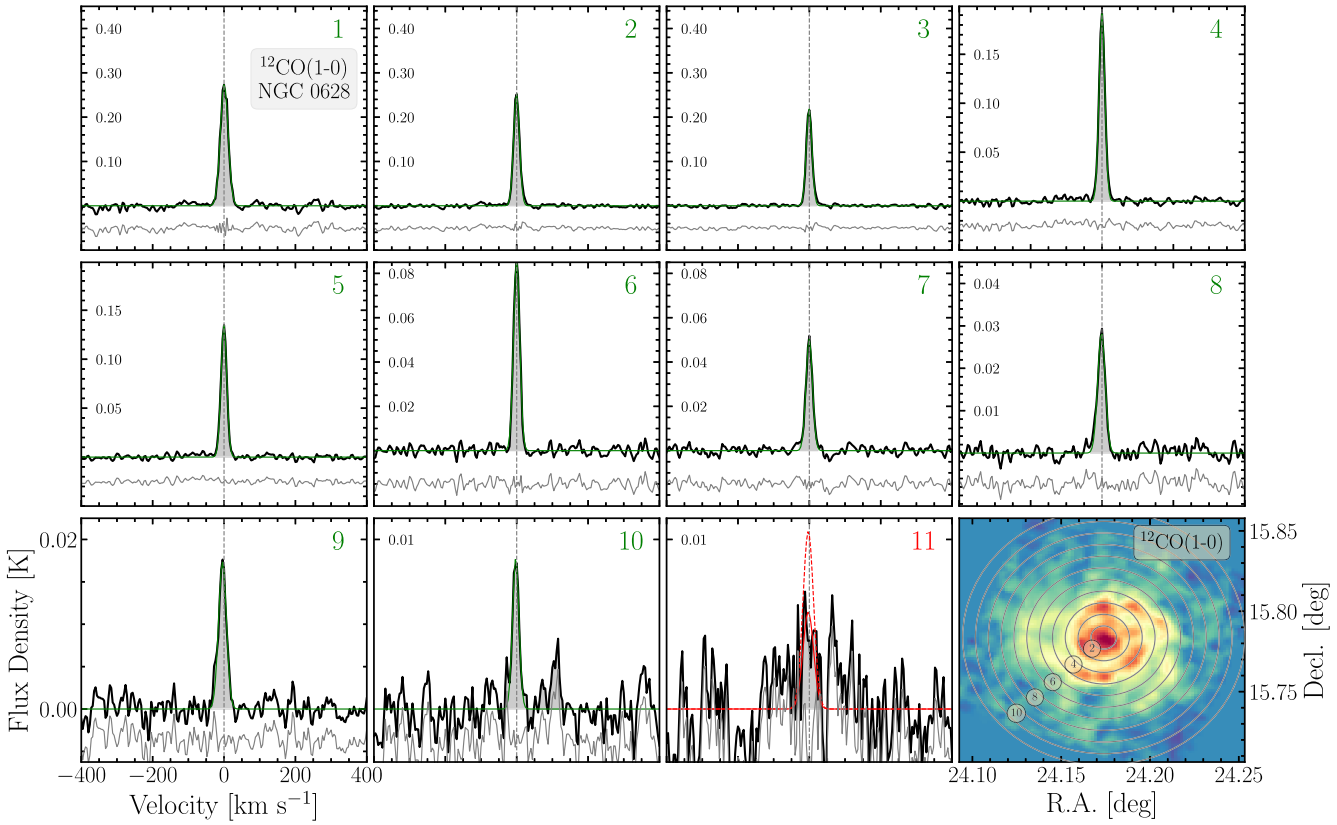

Figure D1. Illustration of spectral stacking. Spectra of CO(1–0) brightness temperature stacked as function of galactocentric radius in NGC 0628. The outer radius, in kpc, of each kpc-wide ring is given in the top right corner, with the colour indicating whether the line peak has S/N > 5 (green) or not (red). Shaded area shows the region over which we integrate to determine the line flux. The light grey dotted line indicates $v = 0\,\mathrm{km\,s^{-1}}$ position. The bottom right panel illustrated the rings used for the stack.

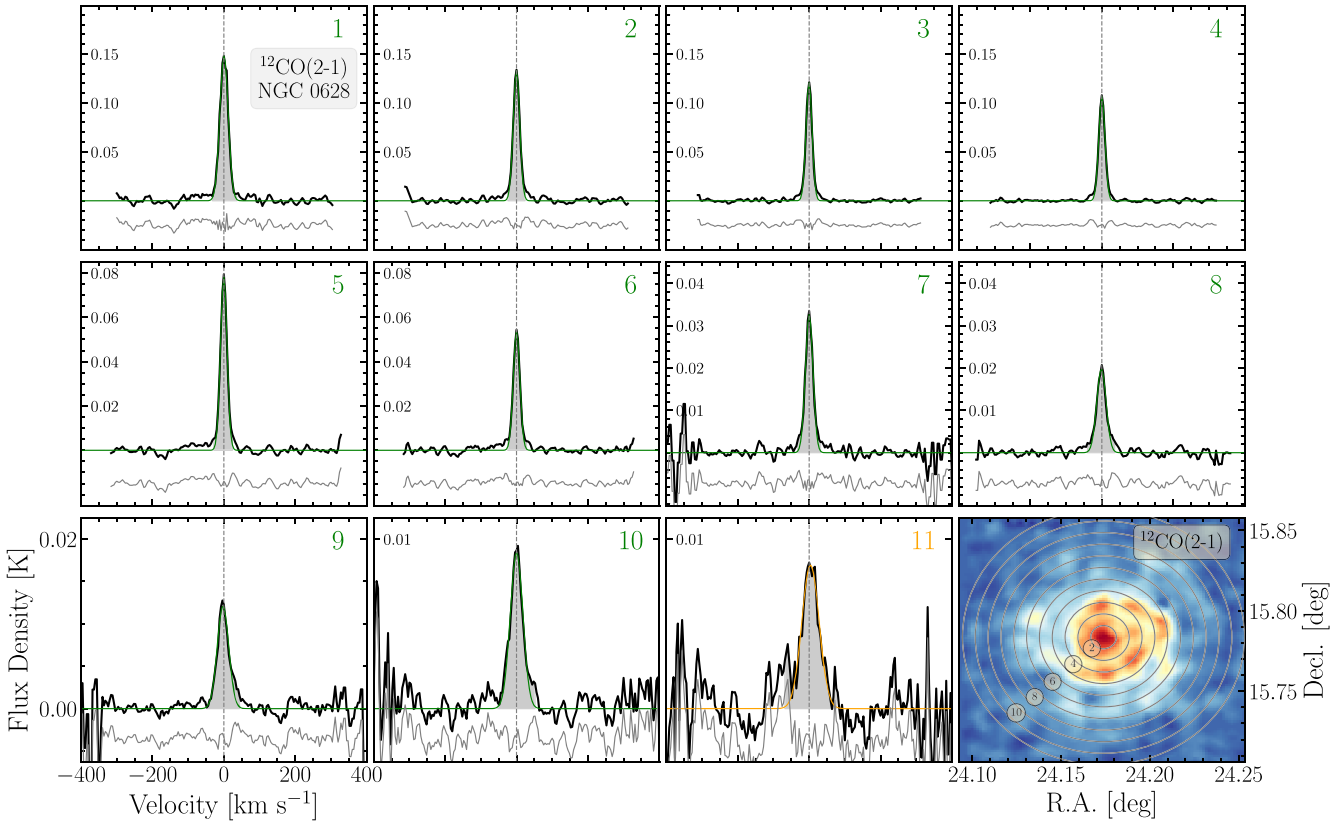

Figure D2. Illustration of spectral stacking**.** As Fig. D1 but now showing the CO(2-1) line stacked in bins of galactocentric radius.

Both figures show that stacked by the local H I velocity produced high signal-to-noise ratio, coherent spectra out to ∼10 kpc. The grey region shows the integration area used to determine the integrated brightness temperature. This clearly corresponds to all real astrophysical signal out to 9 kpc, with the algorithm used to identify the line width doing a good job.

Outside 9 kpc, we begin to see some breakdowns in the approach. In the 10th radial bin, which covers $r_{\rm gal}$ = 9–10 kpc, the CO(1–0) spectrum includes a second, fainter, CO(1–0) emission peak to the right of the main emission line. This second peak appears displaced by approximately 100 km s^{-1} from the main peak. This could represent a noise spike, a problem with the H I velocity field, a problem with baseline subtraction, or real signal. No analogous feature appears in the CO(2–1) spectrum. In this case, we manually extended the integration range to cover the additional emission line, but the profile becomes uncertain in this bin. This uncertainty is higher than the statistical uncertainty. In the next panel, we see that by 10–11 kpc, uncertainties in the baseline produce large 'ripples' in the spectrum for both CO(1–0) and CO(2–1). Though formally the S/N of the data remain high (there is extended emission over a large velocity range), these results remain uncertain.

These sorts of systematic uncertainties tend to crop up in the outer parts of the binned data. These breakdowns in the stacking procedure and baseline uncertainties contribute to some of the scatter in profiles at low intensity but are not necessarily reflected in the statistical scatter.

APPENDIX E: AZIMUTHAL R_{21} VARIATION IN NGC 5194

In this study, we investigated the spatial variation of the CO line brightness temperature ratio across the full discs of the galaxies. The source NGC 5194 is not only unique in the sense that it shows strong differences between the arm and interarm regions, but the trend we find disagrees with a previous study by Koda et al. (2012). We find larger R_{21} values (∼0.9–1) in the interarm region compared to the molecular arm region (∼0.7–0.8). This trend stands in contrast to the one reported in Koda et al. (2012). Here, we investigate the origin of the discrepancy. As we used different observations than the previous study, it is essential to determine, which data set causes the discrepancy. In particular, we used CO(1–0) observations from the PAWS survey (Pety et al. 2013) and CO(2–1) observations from the M51 Large Program (den Brok et al., in preparation). Koda et al. (2012) on the other hand used CO(1–0) observations from NRO (Koda et al. 2011) and CO(2–1) observations from HERACLES (Leroy et al. 2009).

To properly analyse the azimuthal variation, we followed the same procedure as described in Koda et al. (2012) to determine the variation of R_{21} as a function of the spiral phase. Fig. E1 shows (left-hand panel) the result, where we looked at the line ratio using all possible permutations of the CO(1–0) (PAWS and NRO) with CO(2–1) (M51 LP and HERACLES) data sets. All observations were convolved to a common resolution of 24 arcsec. The red and blue hashed regions indicate the location of the spiral arm, as given in Koda et al. (2012) (see also right-hand panel of Fig. E1; molecular arm: blue, ψ = 60–100° and 230–270°; leading edge: red, ψ = 80–120° and 250–290°). Note that the y-axis shows the normalized line

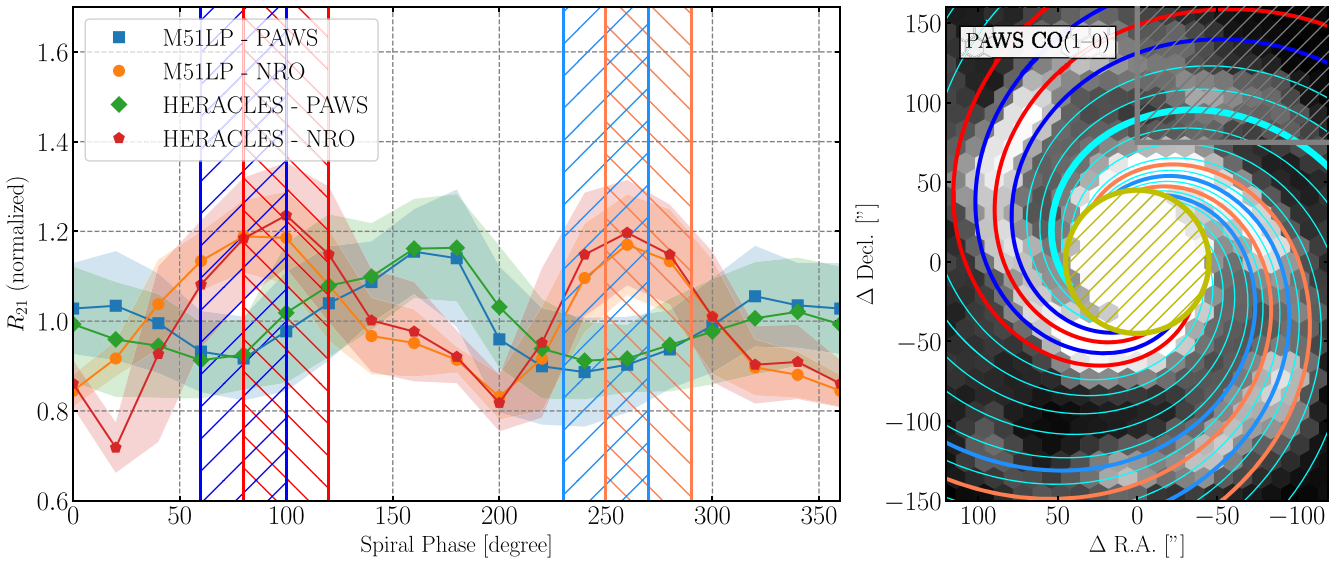

Figure E1. Normalized CO line ratio as function of spiral phase. We follow the spiral phase analysis of the integrated CO line brightness temperature ratio, as described in Koda et al. (2012), to study the azimuthal variation. (*Left-hand panel*) Here, we show all four permutations of the two different CO data sets for both transitions. Each line ratio is normalized by the corresponding mean. The shaded area shows the standard deviation of the line ratio binned by spiral phase. The blue, hashed band shows the molecular arm and the red, hashed band indicates the trailing star-forming arm, as provided by Koda et al. (2012). It is evident, that upon changing the CO(1–0) data set (PAWS versus NRO), the trend of the line ratio changes, while changing the CO(2–1) data sets (M51 LP versus HERACLES), has no effect. (*Right-hand panel*) The CO(1–0) brightness temperature map showing the PAWS data. The spiral phases are plotted in increments of 20°. Points within the central hashed inner 45 arcsec area as well as the hashed region in the north-west are excluded from the spiral phase bins. We applied an S/N = 10 threshold.

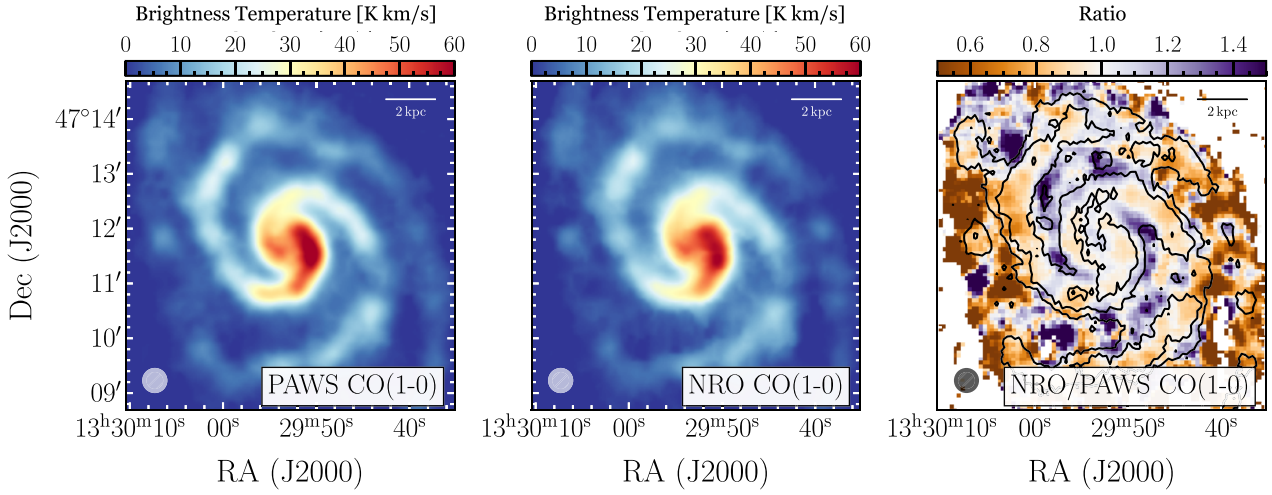

Figure E2. CO(1–0) line brightness temperature maps. Side-by-side comparison of (*left*) the PAWS CO(1–0) line brightness temperature map (Pety et al. 2013) and (*right*) the NRO 45-m map (from Koda et al. 2011). Both show similar line brightness temperature values, but the NRO map shows a weaker contrast between arm and interarm region (especially at the position of the outer arms), which we believe causes the discrepancy in the azimuthal line brightness temperature trend discussed in Section 3.3 and Appendix E. Both maps were created by integrating the full cube over the same masked velocity range. (*right*) The ratio of the CO(1–0) line of the NRO map and the PAWS map. The contours are drawn at S/N = 20, 40, and 100 of the PAWS CO(1–0) map. The discrepancy is strong in the interarm region, where the ratio is clearly >1.

ratio. It is evident from this plot, that the discrepancy is caused by the use of a different CO(1–0) data set. Substituting the CO(2–1) HERACLES data with the M51 LP observations does not change the azimuthal trend at all. Fig. E2 shows the two different CO(1–0) maps side by side. The left-hand panel shows the PAWS brightness temperature map, while the map on the right illustrates the NRO map. The NRO map has a native resolution of 19.7 arcsec. For the comparison, we convolved it to the resolution of the PAWS CO(1–0) map at 24 arcsec. We integrated both the NRO and PAWS cube over the same velocity range. Already from visual inspection, it is evident that the contrast between arm and interarm is higher in the PAWS than in the NRO map (especially for the position of the outer arms). The discrepancy can be caused by inproper error beam handling, different or unstable T_{sys}, variable S/N, or scanning artefacts. To investigate the exact cause for the discrepancy is beyond the scope of this project.

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