












Spectral stacking of radio-interferometric data

Lukas Neumann¹ , Jakob S. den Brok^{1,2}, Frank Bigiel¹ , Adam Leroy³, Antonio Usero⁴ , Ashley T. Barnes⁵,
Ivana Bešlić^{1,6} , Cosima Eibensteiner¹ , Malena Held¹ , María J. Jiménez-Donaire⁴ , Jérôme Pety^{6,7} ,
Erik W. Rosolowsky⁸ , Eva Schinnerer⁹ , and Thomas G. Williams¹⁰ 

¹ Argelander-Institut für Astronomie, Universität Bonn, Auf dem Hügel 71, 53121 Bonn, Germany
e-mail: lukas.neumann.astro@gmail.com

² Center for Astrophysics, Harvard & Smithsonian, 60 Garden St., 02138 Cambridge, MA, USA

³ Department of Astronomy, The Ohio State University, 140 West 18th Ave, Columbus, OH 43210, USA

⁴ Observatorio Astronómico Nacional (IGN), C/ Alfonso XII, 3, 28014 Madrid, Spain

⁵ European Southern Observatory, Karl-Schwarzschild Straße 2, 85748 Garching bei München, Germany

⁶ LERMA, Observatoire de Paris, PSL Research University, CNRS, Sorbonne Universités, 75014 Paris, France

⁷ Institut de Radioastronomie Millimétrique (IRAM), 300 rue de la Piscine, 38406 Saint Martin d'Hères, France

⁸ Dept. of Physics, University of Alberta, Edmonton, Alberta, T6G 2E1, Canada

⁹ Max Planck Institute for Astronomy, Königstuhl 17, 69117 Heidelberg, Germany

¹⁰ Sub-department of Astrophysics, Department of Physics, University of Oxford, Keble Road, Oxford OX1 3RH, UK

Received 12 February 2023 / Accepted 28 April 2023

ABSTRACT

Context. Mapping molecular line emission beyond the bright low- J CO transitions is still challenging in extragalactic studies, even with the latest generation of (sub-)millimetre interferometers, such as ALMA and NOEMA.

Aims. We summarise and test a spectral stacking method that has been used in the literature to recover low-intensity molecular line emission, such as HCN(1–0), HCO⁺(1–0), and even fainter lines in external galaxies. The goal is to study the capabilities and limitations of the stacking technique when applied to imaged interferometric observations.

Methods. The core idea of spectral stacking is to align spectra of the low S/N spectral lines to a known velocity field calculated from a higher S/N line expected to share the kinematics of the fainter line (e.g. CO(1–0) or 21 cm emission). Then these aligned spectra can be coherently averaged to produce potentially high S/N spectral stacks. Here we used imaged simulated interferometric and total power observations at different S/N levels, based on real CO observations.

Results. For the combined interferometric and total power data, we find that the spectral stacking technique is capable of recovering the integrated intensities even at low S/N levels across most of the region where the high S/N prior is detected. However, when stacking interferometer-only data for low S/N emission, the stacks can miss up to 50% of the emission from the fainter line.

Conclusions. A key result of this analysis is that the spectral stacking method is able to recover the true mean line intensities in low S/N cubes and to accurately measure the statistical significance of the recovered lines. To facilitate the application of this technique we provide a public Python package, called PYSTACKER.

Key words. methods: data analysis – techniques: interferometric – galaxies: ISM – radio lines: galaxies – radio lines: ISM

1. Introduction

Mapping extragalactic molecular line emission with high spatial resolution and sensitivity is still challenging even with the latest generation of (sub-)millimetre interferometers, such as the Atacama Large Millimeter/submillimeter Array (ALMA) and the Northern Extended Millimeter Array (NOEMA). In practice, for most nearby galaxies, only the low- J CO transitions, which are the brightest millimetre-wave lines, can be rapidly surveyed at a good resolution ($\lesssim 1''$) while also achieving widespread high-significance detections across the full disc of a typical star-forming galaxy (e.g. Leroy et al. 2021b). Recovering integrated intensities of fainter, and hence typically low signal-to-noise ratio (S/N) lines, such as HCN(1–0), HCO⁺(1–0), or N₂H⁺(1–0), is more challenging. These lines carry critical physical information on the composition, temperature, and density of the gas, but often have intensities 30 to >100 times fainter than the CO lines (e.g. Usero et al. 2015; Jiménez-Donaire et al. 2017). To measure the intensities of these other lines, ‘spectral stacking’ methods have become popular in recent years.

Stacking of astronomical data has been used for at least four decades (e.g. Cady & Bates 1980) and applied across wavelength regimes, from X-ray (e.g. Hickox et al. 2007; Chen et al. 2013) to sub-millimetre and radio wavelengths (e.g. Knudsen et al. 2005; Karim et al. 2011; Schrubba et al. 2011; Delhaize et al. 2013; Caldú-Primo et al. 2013; Bigiel et al. 2016; Lindroos et al. 2016; Jolly et al. 2020). In the past decade, spectral stacking has become a particularly important tool in millimetre studies of galaxies, allowing the recovery of otherwise undetected line emission. For example, Jiménez-Donaire et al. (2019), Bešlić et al. (2021), and Neumann et al. (2023) all use spectral stacking leveraging a CO emission prior to recover emission from faint high critical density emission lines, including HCN(1–0), HNC(1–0), or HCO⁺(1–0), across large areas in the discs of nearby galaxies. den Brok et al. (2021) and den Brok et al. (2022) used spectral stacking based on ¹²CO to obtain more significant constraints on lines tracing rarer CO isotopologues. And Schrubba et al. (2011) used 21 cm emission as a prior to construct extended, sensitive radial profiles of CO emission even in the outer parts of galaxies. These studies all demonstrate how spectral stacking

recovers more information about the distribution, composition, and physical conditions of the molecular gas in galaxies.

The basic idea of spectral stacking as often applied to nearby galaxies is to align all spectra by recentring them on the local mean velocity of the interstellar medium (ISM), which is measured using a high S/N prior (e.g. CO(1–0)) or the 21 cm line (Sect. 2). Then spectra from different parts of the galaxy can be coherently averaged with minimal contributions from noise in empty parts of the bandpass. By averaging in azimuthal rings, one can construct sensitive radial profiles. One can also average as a function of other quantities to test specific hypotheses or scaling relations (e.g. galactocentric radius, line intensity, surface density, or star formation rate). Carrying out this stacking on the spectra allows an important visual check that the averaged result indeed looks like an astrophysical spectral line (i.e. to first order a Doppler-broadened Gaussian line profile), and can even allow recovery of mean kinematic information via the width of the Gaussian.

While these techniques are simple in principle, a key uncertainty remains surrounding their application to the most powerful current millimetre-wave telescopes, ALMA and NOEMA. These facilities are interferometers, and the images they produce reflect both incomplete sampling of the u – v plane and a deconvolution process that often focuses on bright emission. While u – v plane stacking can alleviate both concerns in unresolved objects or those with simple geometries, stacking in the image plane remains the most practical option for extended, complex sources, such as nearby galaxies. Since stacking using these powerful telescopes represents a key way to push our knowledge of the physical state and makeup of the ISM, evaluating the accuracy of this technique when applied to recover faint low S/N lines from interferometer data is a key next step.

The goal of this work is to provide such a demonstration. For this purpose we used the Common Astronomy Software Applications (CASA; CASA Team et al. 2022) ALMA simulator to simulate interferometric and total power observations of low S/N lines based on a known input model. The resulting simulated observations were imaged using the Physics at High Angular resolution in Nearby Galaxies (PHANGS)–ALMA pipeline (Leroy et al. 2021a). Then we applied the spectral stacking method and assessed how well the stacks recover the known input. In particular, we stacked via the galactocentric radius using simulated CO(2–1) data cubes built on real observations of galaxies from the PHANGS–ALMA survey (Leroy et al. 2021b).

We also present a new public Python package, PYSTACKER that can be used to easily apply these techniques. This utility complements the tool LINESTACKER presented by Jolly et al. (2020), which is also validated against simulation. Their work focuses on spectrally stacking many distinct sources in three dimensions, while our code emphasises stacking within an individual data set in the presence of a complex prior velocity field. Another stacking package called SPECTRAL-STACK¹, relies on Fourier shifting to align the spectra to be averaged. The advantage of this approach is that the noise properties and channel-to-channel correlations are preserved. However, it deals less well with edge effects.

2. Description of the spectral stacking method

The main goal of the spectral stacking technique is the recovery of low S/N lines by shifting the spectra to a known velocity field defined by a high S/N prior (e.g. CO(1–0) or CO(2–1) or the

H I 21 cm line) and then averaging the spectra based on another parameter such as environment, star formation rate, or line intensity. Our stacking method is based on den Brok et al. (2022) and Neumann et al. (2023) and our implementation is available as a Python package, called PYSTACKER². We describe the basic steps of the code in the following.

We begin with a set of data cubes of the same target with different S/N levels of the input line emission. First, we homogenise the data bringing all data cubes to the same coordinate grid and convolving to the same spatial resolution. Then we define a prior, typically the most significantly detected line, which is used to obtain the velocity field as the velocity at the peak intensity of each spectrum³ (Koch et al. 2018). We use this velocity field to redefine the spectral axis for each individual spectrum in the cube⁴ so that the emission of all lines should be centred at a velocity of 0 km s^{−1}. The result is sometimes referred to as a ‘shuffled’ cube, in reference to the shuffle task of the Groningen Image Processing System (GIPSY; van der Hulst et al. 1992).

Figure 2 shows the basic functioning scheme of the PYSTACKER package, which allows for two input options. The first is the PyStructure database, a numpy dictionary containing all the molecular line emission data. The PyStructure database is produced by a separate pipeline and already contains the velocity alignment analogous to the velocity shuffling performed by PyStacker. In the second option (the default for most users), the input can be data cubes in the form of .fits files, where each FITS file contains the position-position-velocity information of the respective spectral line. Here the user can provide a model velocity field used to shuffle the velocity field of the lines to be stacked. In both cases, a configuration file must be specified, which sets the parameters for the stacking. If data cubes are provided, they are sampled on hexagonal gridded pixels with half-beam spacing. Next, the significant pixels of the given prior are identified and the velocity field is shuffled based on the moment-1 of the prior (if not provided by the input model). After applying the velocity offsets to the molecular line data, the spectra inside the given bins are averaged. The user can specify in the configuration file if the prior-non-detected pixels are ignored or set to zero for the bin average. Afterwards, the stacked spectra of the prior are used to build a velocity mask for each stack, which is used to compute the integrated intensities (inside the mask) and the uncertainties (rms outside the mask). The final output is a numpy dictionary containing the stacked spectra along with their integrated intensities, uncertainties, and other quantities (see the documentation for the full output content).

One can average the shuffled spectra inside bins defined by any arbitrary quantity of scientific interest (e.g. galactocentric radius, CO(1–0) line intensity, or star formation rate). For instance, in this work we stack as a function of galactocentric radius (Sect. 3). If signal is present in the stacked line within a given bin, the averaged spectrum should then appear as a clear emission line (e.g. Fig. 1 right panel). For comparison, averaging across different parts of strongly rotating discs without first adjusting to the local velocity yields a broad lower signal-to-noise profile (e.g. see Fig. 2 in Schrubba et al. 2011).

² <https://github.com/PhangsTeam/PyStacker>

³ The code also allows inputting a model velocity field.

⁴ Recentring the spectrum itself has some associated subtleties, and can be done using either Fourier techniques or via regridding and oversampling. In this paper we use re-gridding techniques, but the choice of approach can affect the channel-to-channel correlation and noise properties of the stacked spectrum.

¹ https://github.com/low-sky/spectral_stack

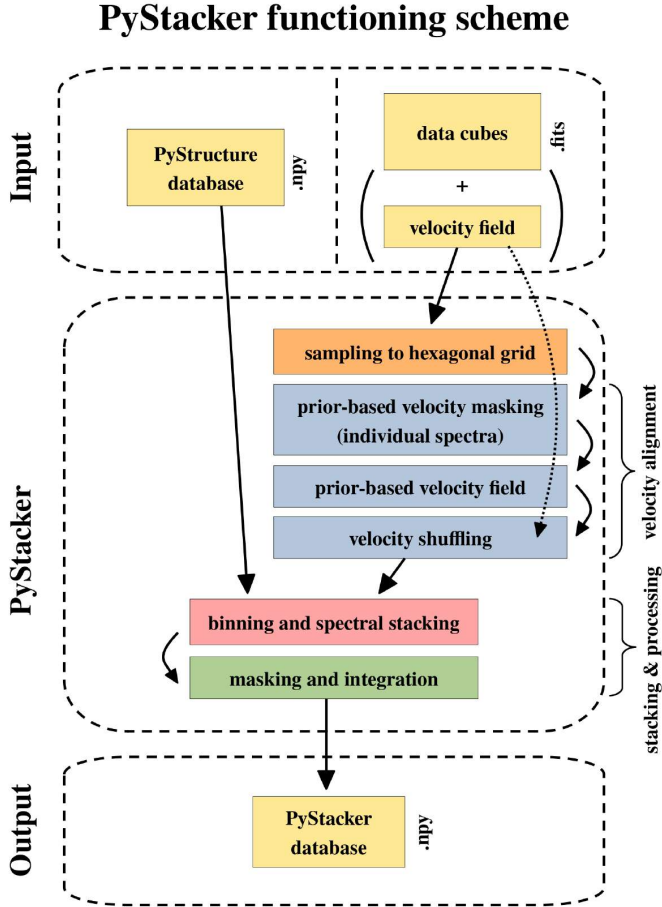


Fig. 1. Schematic of the PyStacker package functioning principle. The input can be either a so-called PyStructure database (at the moment of submission only internally used) or FITS files of data cubes containing the molecular line emission. The latter is applicable to all users. The PyStructure files does already include the re-sampled and velocity aligned data. If FITS files are provided, the PyStacker package will perform the velocity alignment given an input prior or take the input velocity field. In both cases, the spectra will be stacked according to the input stacking quantity and processed to retrieve average integrated intensities. The results are then returned as a Python dictionary, which can be read e.g. following the example script coming with the PyStacker package.

We can only reliably shuffle the spectra if the prior is actually detected in the respective spectrum. Thus, if the prior is only detected in a fraction of the spectra inside a bin, we rely on these spectra to infer the average of the bin. In this case we consider the emission in the spectra that could not be shuffled to be equal to zero, such that the average is always measured relative to all spectra inside a given bin⁵. We expect this to be a reasonable assumption when stacking a rare faint molecular line such as HCN(1–0) using CO as a prior⁶.

Within any given bin n , we measure the average spectrum

$$T_{n,\text{stack}}(v) = \frac{1}{N_{\text{tot}}(n)} \sum_{i=0}^{N_{\text{det}}(n)} T_{n,i}(v), \quad (1)$$

⁵ The alternative, i.e. averaging over the prior-detected spectra only, tends to overestimate the stacked intensity. However, PyStacker allows us to also use this option, and we show its results in Fig. A.3.

⁶ In other applications of this method, a lower resolution cube or even a model rotation curve may sometimes be used as a prior to shuffle in cases where the brighter line has patchy coverage or limited S/N.

where $N_{\text{tot}}(n)$ and $N_{\text{det}}(n)$ are the total number and the prior-detected number of spectra in bin n . To compute the integrated intensities of the stacked spectra, we built a mask based on the high-S/N reference cube. We selected the velocity range of significant emission for each spectrum as described in Bešlić et al. (2021) and integrated the intensities over mask-selected velocity channels

$$W_n = \sum_{N_{\text{mask}}} T_{n,\text{stack}}(v) \cdot \Delta v_{\text{channel}}, \quad (2)$$

where $\Delta v_{\text{channel}}$ is the channel width and N_{mask} is the number of (independent) channels inside the mask. The nominal uncertainties of the integrated intensities (σ_W , studied in Sec. 3.1) are given by

$$\sigma_W = \text{rms} \times \sqrt{\frac{N_{\text{tot}}}{N_{\text{det}}}} \times \Delta v_{\text{channel}} \times \sqrt{N_{\text{mask}}}, \quad (3)$$

where rms is the root mean square of the emission-free channels (i.e. outside the mask) in the stacked spectrum. Since the stacked spectrum is computed from the prior-detected pixels, N_{det} , but divided by the total number of pixels in that bin, N_{tot} (Eq. (1)), the measured rms of the emission-free channels is biased low if $N_{\text{det}} < N_{\text{tot}}$. Therefore, we have to correct the rms by the factor $\sqrt{N_{\text{tot}}/N_{\text{det}}} (\geq 1)$ in Eq. (3) in order not to underestimate the rms, and thus σ_W . The correction factor mimics the increase in noise when adding up N_{tot} spectra with the same noise level.

3. Recovery of integrated intensities

We apply our method to simulated data cubes with known input to test how well spectral stacking can recover the integrated intensities of molecular line emission as a function of the noise level of the observations. Specifically, we use a set of simulations of molecular line emission produced to validate the PHANGS–ALMA data reduction pipeline (Leroy et al. 2021a). As described in Leroy et al. (2021a), the simulated CO(2–1) data cubes were produced using the CASA tasks `simdata` and `simobserve` using inputs based on real PHANGS–ALMA CO(2–1) images. The simulated observations mimic interferometric observations of the galaxy NGC 3059 similar to the PHANGS–ALMA survey⁷. The simulations included the creation of a simulated total power map constructed by convolving the input model to the resolution of the ALMA TP antennas and adding Gaussian noise of the expected magnitude for a real PHANGS–ALMA TP observation.

The input intensity cube, hereafter referred to as the template, is the actual masked NGC 3059 cube from PHANGS–ALMA. We show the integrated intensity map of this template data cube in the upper left panel of Fig. 2. These ‘true’ data are used to construct simulated 12 m, 7 m, and total power observations and imaged via the PHANGS–ALMA pipeline (for more details see Leroy et al. 2021a). Then we run these through the stacking pipeline in this study. The use of real data as a model means that there will be some observational noise in the true data, but we consider that as signal, and explore how well it gets recovered, and it should have only a modest impact on the analysis.

The simulations produce images for different combinations of the ALMA main array, the ACA 7 m antennas, and the

⁷ The NGC 3059 look-alike has been rotated so that the major axis of the galaxy is aligned with the declination. However, these modifications have no effect on our analysis.

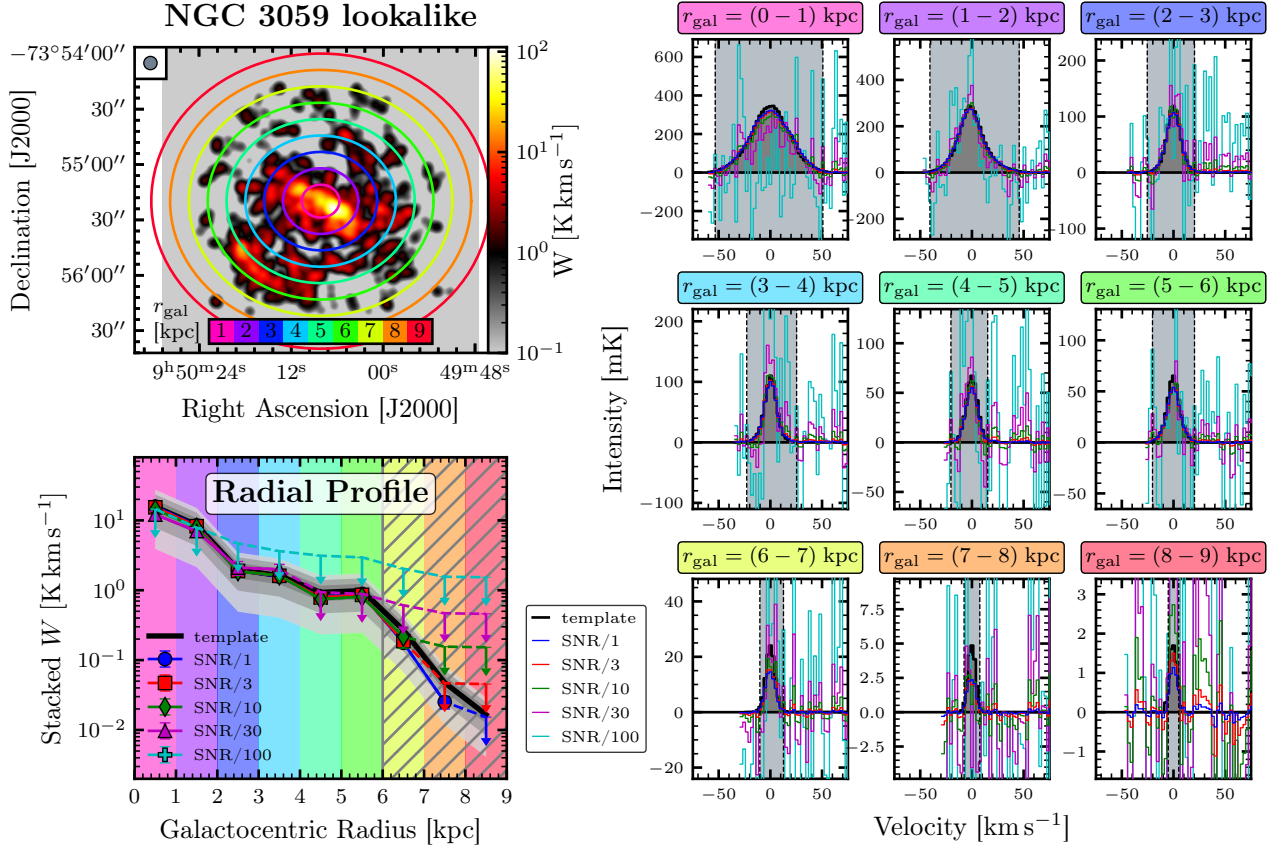


Fig. 2. Spectral stacking of the NGC 3059 template galaxy via galactocentric radius. *Top left:* moment-0 map of the true (i.e. known) simulated CO(2–1) emission mimicking the intensity distribution of NGC 3059. The data have been convolved to the common highest resolution over all array configurations (7.2'') indicated by the grey circle in the upper left. The coloured rings show the loci of the galactocentric radius. *Right panels:* stacked spectra in 1 kpc radial bins from the centre out to 9 kpc as illustrated in the top left panel. The dark grey histogram shows the true spectra (i.e. those obtained by stacking the input template). The coloured lines show the stacked spectra from the respective simulated data cubes. The grey-shaded area indicates the velocity range used to compute the integrated intensities. *Bottom left:* stacked integrated intensities corresponding to the spectra shown in the right panels plotted against galactocentric radius. The black line shows the true radial trend, and the coloured lines show the recovered trend of the simulated data cubes. Solid points indicate data above 3σ ; downward pointing arrows denote 3σ upper limits. The grey shaded areas show differences to the true trend in levels of $\pm\{25, 50, 75\}\%$. The hatched area denotes the regime where the prior S/N/1 is detected ($S/N \geq 3$) in less than 20% of the pixels.

total power data: 12m+7m+tp, 12m+7m, 7m+tp, 12 m and 7 m (Sect. 3.2). They also produce cubes with a range of different signal-to-noise levels, which we refer to as S/N/1, S/N/3, and so on. The S/N/1 cube mimics the sensitivity of a typical PHANGS–ALMA CO(2–1) observation, while S/N/3, S/N/10, S/N/30, and S/N/100 have a factor of 3, 10, 30, and 100 lower S/N⁸, respectively, but leave the noise the same. Here, we take these cubes and rescale them with the respective factors to obtain cubes at the same intensity but different noise, and thus S/N levels. At the common PHANGS–ALMA sensitivity and for a brightness distribution similar to NGC 3059, S/N/10 could be representative for $^{13}\text{CO}(1-0)$, S/N/30 for HCN(1–0) or HCO^+ and S/N/100 for fainter lines such as $\text{N}_2\text{H}^+(1-0)$.

The S/N levels of the moment-0 maps resulting from the various S/N cubes range from 2.5 (minimum), 341.7 (maximum) for the S/N/1 data; over -1.9 (minimum), 32.5 (maximum) for the S/N/10 cube; and down to -3.8 (minimum), 3.7 (maximum) for the version with 100 times higher noise. This means we can

study cubes that contain significant emission across most of the field of view all the way to almost pure noise cubes.

The angular resolution of the 12m+7m+tp cube is 2.7'' and higher than that of the 7m and 7m+tp cubes at 7.2'', which correspond to a linear scales of 264 pc and 702 pc, respectively, at a distance of 20.2 Mpc. In order to compare the 12m+7m+tp results more directly with the other arrays, we convolve all cubes to a common 7.2'' resolution and focus on the 12m+7m+tp at 7.2'' resolution for most of the analysis. We note that the convolution from the native 12m resolution (2.7'') to the common best resolution (7.2'') might smear out some of the significant compact emission, and thus potentially reduce the efficiency of the stacking. However, we checked, in our case, that the recovered stacks from the native resolution cubes are consistent with the stacks from the convolved cubes.

We apply the spectral stacking method described in Sect. 2 to the five data cubes at different S/N levels described above. We stack the spectra by galactocentric radius from 0 to 9 kpc in 1 kpc increments as illustrated in Fig. 2. We note that there is very little emission (less than 20% of the pixels in the S/N/1 moment-0 map contain significant emission) outside of 6 kpc.

⁸ The original exercise in Leroy et al. (2021a) actually scales the signal down by factors of 3, 10, 30, 100.

Therefore, we limit most of the discussion to the inner 6 kpc and consider this the typical extent of the molecular gas disc. For the outer radii, an alternative approach could be to average over a larger region, for example binning everything beyond the radius of 6 kpc, in order to potentially recover more of the fainter emission at the cost of spatial information. However, in this case we do not recover more emission due to the steep drop in emission beyond 6 kpc. We use the S/N/1 data (i.e. the simulated data mimicking PHANGS–ALMA CO(2–1) observations) as a prior to account for the varying velocity field across the galaxy and to determine the channels of significant emission to compute the integrated intensities. For each configuration, the respective S/N/1 cube is used as a prior. This means that to stack the 12m+7m+tp cubes, we use the 12m+7m+tp S/N/1 cube as the prior; to stack the 7m+tp data, we use the 7m+tp S/N/1 cube, and so on. This approach is similar to how we typically handle real observational data, where different lines have been observed with the same interferometric set-up. In this case, by construction, the velocity fields of the different line cubes are identical. In reality, we may expect small velocity offsets between different spectral lines leading to slightly broader, and thus potentially less significant stacked lines, though this effect is expected to be small when studying various molecular lines, which should share similar kinematics. The stacked spectra are shown in the right panels of Fig. 2. Since in this case we know the true velocity distribution from the template cube, we repeat the same procedure using the true velocity field (Appendix A).

The resulting radial profiles of the stacked integrated intensities are shown in the bottom left panel of Fig. 2. Overall, we find that the radial trend is well recovered across most of the molecular disc down to the S/N/10 data cubes. In the S/N/30 cube we are still able to recover the radial trend out to 4 kpc, where, in the 3 to 4 kpc bin, the median moment-0 S/N is 0.54. For the noisiest data used here (S/N/100) we obtain only upper limits, which highlights that it is extremely challenging to map molecular discs of nearby galaxies in line emission that is ~ 100 times fainter than CO(2–1) (e.g. the popular Galactic dense gas tracer N_2H^+). PHANGS–ALMA integrated for ~ 1 min per field. This exercise implies that to achieve the S/N ~ 10 required for reliable stacked detections, integrations ~ 100 times longer, ~ 2 h per pointing, would be required. Another approach could be to modify the binning (e.g. by averaging spectra over larger regions). Although this could lose spatial information, and so was not performed here, this approach could potentially recover otherwise undetected emission. Therefore, we recommend adapting the binning parameters to the strength and distribution of the studied line emission. We note that NGC 3059 is a relatively low-luminosity galaxy, and the situation may be more optimistic in somewhat brighter targets.

We highlight the differences between the recovered stacks and the true values in Fig. 3. Based on the computed uncertainties of the stacked integrated intensities, σ_w , we clip at S/N (W/σ_w) levels of 3, 5, and 10. As expected, we find that with stronger σ_w -clipping the stacked line intensities show better agreement with the true values such that for data above 10σ the maximum discrepancy is $<35\%$, and $<15\%$ ($<8\%$) in the inner 6 kpc (4 kpc). We also systematically find values that are too low at larger radii (i.e. apparently significant measurements that do not agree within the uncertainties with the true values for $r_{\text{gal}} > 4$ kpc). This offset might be explained by the low fraction of spectra contributing to the stacks in these bins (see Table A.1). For the outer bins, $r_{\text{gal}} > 4$ kpc, less than half of the spectra inside each bin could be used for stacking, and as a result, we may potentially miss some emission hidden in the noise that we

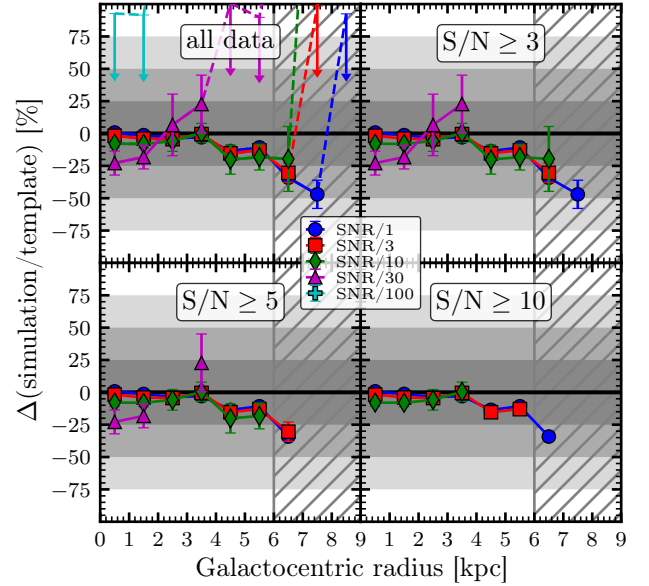


Fig. 3. Agreement between measured and expected radial trends at different sigma-clipping levels. Shown is the relative differences between the measured radial stacks from the simulated 12m+7m+tp cubes and the true stacks, as follows from the bottom left panel of Fig. 2, by subtracting the template trend. The difference between the different panels is that the data of the resulting stacks is clipped at 3, 5, or 10 S/N ($W\sigma_w$). The stacking procedure is always the same. The hatched area denotes the regime, where the prior S/N/1 is detected in less than 20% of the pixels.

are not able to recover. However, we find a very similar discrepancy if the velocity field is perfectly known (Fig. A.2), which suggests that the offset is at least partly arising from the imaging and not the stacking procedure. Nevertheless, we find, over all S/N cubes, an agreement between the significant stacked line intensities and the true values within 23% in bins where the prior is at least moderately ($\geq 36\%$ of the pixels) detected (i.e. within 6 kpc). These results demonstrate that the quality of the stacking results is linked to the significance of the prior used to align the velocity field and the imaging of the interferometric data.

3.1. Uncertainties

For interpreting the results, it is crucial to have a robust measure of the uncertainties and the resulting S/N in order to infer if a data point is significant or not. We measure the uncertainties of the stacked integrated intensities from the standard deviation in the emission-free channels following Eq. (3). Here we check whether this uncertainty matches the uncertainty obtained by propagating the noise measured in the cube.

To do so, we take the S/N/100 cube, which does not contain any significant spectra and consider it as a pure noise cube. We compute the rms in each pixel as the standard deviation across the corresponding pixel. Next, we bin the noise map in radial increments, analogous to radial stacking, and propagate the uncertainty to obtain the expected rms of the stacked spectrum in each bin. The propagated uncertainty is computed as the average rms in each bin corrected for the number of pixels by dividing by the square root of the number of pixels in that bin. Finally, the expected uncertainty is computed analogously to the measured uncertainty (Eq. (3)), but using the cube-propagated rms. We re-scale the cube-propagated rms to the respective noise cubes by multiplying by the respective noise level factors, and plot the measured against the expected uncertainties for all S/N

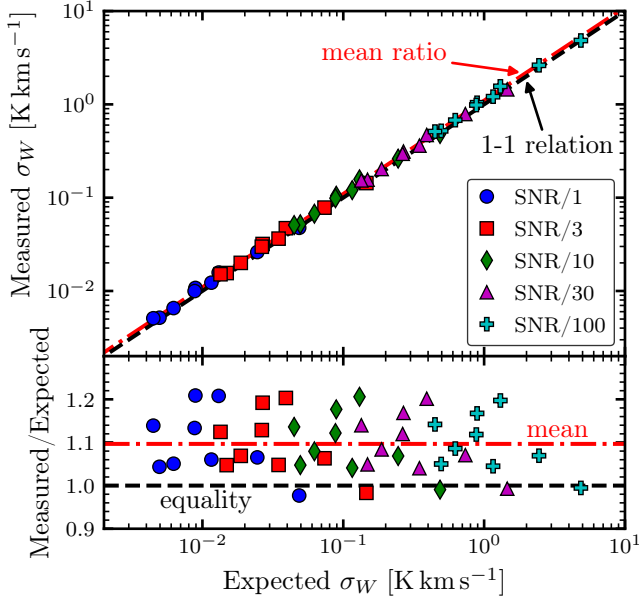


Fig. 4. Measured vs expected uncertainties. *Top:* Comparison between measured (12m+7m+tp) and expected uncertainties of the integrated stacks. The measured uncertainty is obtained from the emission free channel of the stacked spectra following Eq. (3). The expected uncertainty is inferred via Gaussian error propagation from the S/N/100 cube treated as a noise cube. *Bottom:* Ratio of the measured to the expected uncertainties.

cubes (Fig. 4). In the Appendix we also show the resulting S/N of the stacks (i.e. W/σ_W) and compare the measured and expected S/N (Fig. A.1).

We find that the measured uncertainties are strongly correlated with the expected uncertainties, but slightly biased by $\sim 10\%$ on average and little scatter within $\pm 10\%$. The slightly too large measured uncertainties could arise from some emission remaining in the assumed emission-free channels after masking, which contributes to the rms estimation. These results demonstrate that we measure trustworthy statistical uncertainties on the stacked integrated intensities.

3.2. Array configurations

Interferometric observations filter out the extended emission of the source if not combined with single-dish data. Using the simulated observations, we can study how well interferometric data alone can recover line emission in radial bins, and so test whether total power data are needed to obtain accurate stacking results. We repeat the above-described spectral stacking method using data obtained from combining different telescope array configurations: 7m+tp (the ACA including total power data), 12+7 m (the main array and ACA 7 m antennas), 12 m (the main array alone), and 7 m (the ACA 7 m data alone) (see Leroy et al. 2021a, for more information).

Figure 5 presents the radially stacked line intensities from the above-listed configurations relative to the template values. The 12m+7m+tp configuration should recover all spatial scales and can be considered the benchmark for the other configurations. We find that the 7m+tp data performs similarly to the 12m+7m+tp, though with a significantly larger scatter, which is expected due to the lower sensitivity. For the pure interferometric data (i.e. 12+7m, 12 m, and 7 m), we systematically find stacked

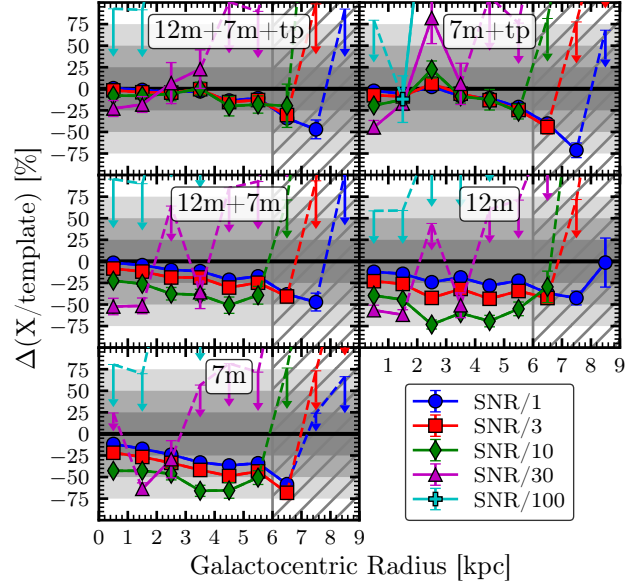


Fig. 5. Flux recovery using different array configurations. Comparison between radial stacking obtained from different array configurations ($X = \{12m + 7m + tp, 7m + tp, 12m + 7m, 12m, 7m\}$), as indicated in the top of each panel. Shown is the ratio between the radial stacks obtained from the simulated data cubes at the given combination of telescope arrays and the true template values against the galactocentric radius. Solid points show data above 3σ and downward pointing arrows denote 3σ upper limits. The hatched area denotes the regime, where the prior, i.e. S/N/1, is detected in less than 20% of the pixels.

line intensities that are too low at all radii, especially when considering 7 m only, where we miss 10–20% across all bins even for the S/N/1 data. Most interestingly, we find a trend with S/N: the lower the S/N of the cube, the larger the bias of the stacks. In the most extreme case (i.e. 7 m S/N/10) the radial profile is detected out to 6 kpc, but yields 30–50% lower line intensities, than obtained with the 12m+7m+tp configuration. These results are in line with the conclusions about the spatial filtering of interferometric data drawn in Leroy et al. (2021a) and enforce the need for total power observations in order to cover the flux information from small spatial scales.

3.3. Weighted stacking

The benefit of the above-described methodology is the potential recovery of faint emission while conserving the flux in each bin. However, the drawback is that we might stack a few highly significant spectra with many noisy spectra, eventually leading to non-detection in the stacked spectra. To overcome this, we can go beyond the ‘equal weight per spectrum’ stacking described above, and weigh the spectra such that we obtain statistically more significant stacked spectra⁹. We compute the weighted stacks by multiplying the spectra ($T_{n,i}(v)$) with the associated weights (w_i) within each bin n . Then, we sum up the weighted spectra and divide them with the sum of the weights:

$$T_{n,stack}(v) = \frac{\sum_{i=0}^{N_{det}} T_{n,i}(v) \cdot w_i}{\sum_{i=0}^{N_{det}} w_i}. \quad (4)$$

⁹ Though keep in mind that this weighted stacking scheme is in general not flux-conserving as opposed to the unweighted stacking introduced before.

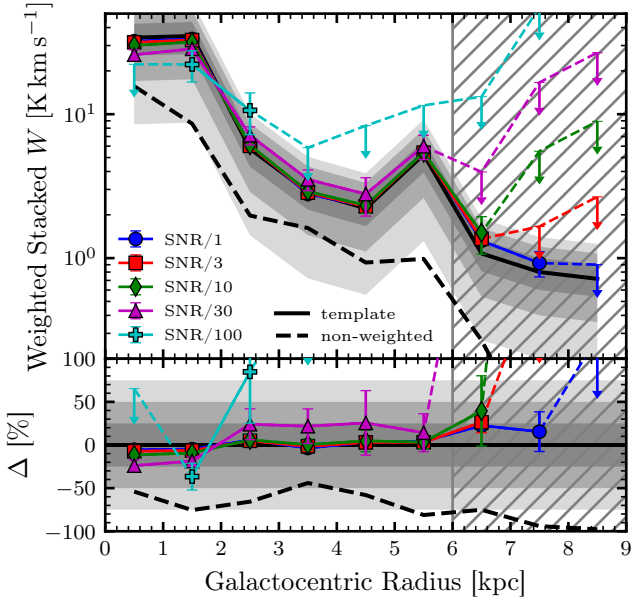


Fig. 6. Intensity-weighted stacking. *Top*: radial stacking similar to Fig. 2, but using the intensity of the prior, i.e. the “S/N/1” integrated intensities, as a weight following Eq. (4). The black solid line shows the template, i.e. the true, radial profile, where the template line intensities are used as the weight. The black dashed line shows the non-weighted radial trend of the template stacking also shown in Fig. 2. *Bottom*: deviation, in per cent, of the stacked radial trend from the template profile. The hatched area denotes the regime, where the prior, i.e. S/N/1, is detected in less than 20% of the pixels.

A useful weighting quantity could be the S/N or the line intensity of the prior. Here we showcase the latter, adopting an intensity-weighted stacking. Thus, we obtain the radial trend of the prior-bright (e.g. CO-bright) regions. We applied the intensity-weighted stacking to the above-introduced simulated observations analogous to the non-weighted stacking. The radial stacking results are presented in Fig. 6. We find that the stacked line intensities of the simulated cubes, excluding S/N/100, are consistent within 20% with the true (weighted) trend. In comparison with the unweighted stacking (Figs. 2 and 3), we find better agreement and no negative bias at low detection fraction of the prior. Thus, weighted stacking can indeed recover faint emission at larger galactocentric radii. However, we note that weighted stacking does not conserve flux and must be interpreted with care, in particular when comparing to stacks derived with another (e.g. non-weighted) method. Moreover, since, by construction, the intensity-weighted stacking used here is computing the weighted average stack over the detected pixels only (i.e. N_{det}), the measured upper limits in the outer bins are much larger than what is obtained in the unweighted case.

3.4. Stacking versus averaging integrated intensities

Instead of averaging stacked spectra, it can be more convenient to average the integrated intensities within the same region or bin. With this approach, typically referred to as binning, the main distinction to the stacking method is that we do not align the velocity field using a prior. Instead, we take advantage of the prior to create velocity masks for each individual spectrum (i.e. for each line of sight), which defines the velocity range over which each spectrum is integrated. The result is an integrated intensity (moment-0) map using a prior inferred velocity (field)

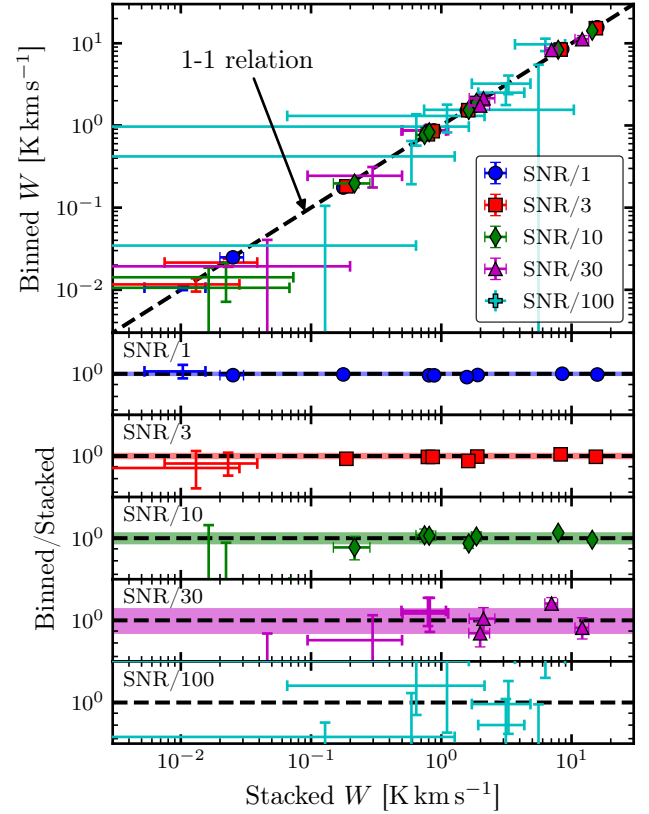


Fig. 7. Binning vs stacking. *Top*: Binned means vs stacked integrated intensities in matched radial bins. The dashed line marks the 1-to-1 relation. Different markers indicate the values recovered from the different S/N cubes. Data above $3\sigma_W$ are shown as markers, else only the error bars are plotted. *Bottom panels*: Ratio of binned mean and stacks to the stacked integrated intensities, separately for each S/N/X cube ($X = 1, 3, 10, 30$). The shaded areas indicate the respective 1 σ scatter of the $3\sigma_W$ data.

mask (see e.g. Gallagher et al. 2018b; Bešlić et al. 2021; den Brok et al. 2022; Neumann et al. 2023, for details about the masking process). Afterwards, we average the integrated intensities inside a given bin.

We apply the above averaging approach to the 12m+7m+tp data sets at different noise levels using the same radial bins, and compare the resulting average line intensities with the stacked line intensities computed as in Sect. 2. We find that the two approaches lead to very similar average line intensities inside a given bin, without bias and small scatter of {1, 2, 5, 10}% considering the significant measurements (S/N ≥ 3) of the S/N/1, 3, 10, and 30 cubes (Fig. 7). In agreement with Gallagher et al. (2018a), we conclude that spectral stacking and averaging masked moment-0 maps yield the same results within 10%. However, we note that spectral stacking still offers the great advantage of also recovering mean line shape, and thus mean kinematic information, which cannot be obtained from the averaged integrated intensities.

4. Conclusions

We performed spectral stacking of simulated interferometric data of the galaxy NGC 3059 as a function of galactocentric radius at different noise levels and combining different telescope arrays. Our main results are the following:

1. Spectral stacking is able to recover the integrated intensities across most of the molecular disc where the prior is predominantly detected. In the most extreme case we detect a stacked spectrum, even in bins where the integrated intensities of the moment-0 map have a median S/N of 0.54. For this specific galaxy, all data above 3σ and 10σ agrees within 23% and 15%, respectively, with the expected values if the prior is detected in at least 36% of the spectra contributing to the stack (i.e. within the inner 6 kpc).
2. Using interferometric data only (i.e. without total power information) can filter out up to 30% of the emission even at the typical PHANGS-ALMA sensitivity and even if the prior is predominantly significant. Even more extreme, for lines that are 10 times (e.g. HCN(1–0)), the 12m-only or 7m-only configurations miss ~50% of the emission in the stacked spectra throughout the full molecular gas disc.
3. The critical limitation of the spectral stacking method is connected to the quality of the prior used to align the velocity field and potentially the imaging procedure. If the prior is not detected across most of the bin, we expect to systematically find stacked line intensities that are too low. This might be improved by using low-resolution priors (e.g. H I 21 cm line) or model priors, which provide a completely defined velocity field. However, we show that the discrepancy can also arise from the imaging of the interferometric data (e.g. if the deconvolution is not able to extract faint emission).

This provides a concrete proof of concept that the stacking method works using combined interferometric and total power data on extended sources. A key result of this analysis is that, at the typical PHANGS-ALMA set-up, the spectral stacking method is able to recover the average integrated intensities within ~23% accuracy, if the prior is detected in at least ~36% of the bin's spectra. We also show that the noise estimated from the line-free parts of the stacked spectra captures the uncertainties of the line intensities with little bias (on average 10% biased high) such that 3σ data can confidently be considered significant detection.

Acknowledgements. We would like to thank the anonymous referee for their insightful comments that helped improve the quality of the paper. L.N. acknowledges funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 516405419. AKL gratefully acknowledges support by grants 1653300 and 2205628 from the National Science Foundation, by award JWST-GO-02107.009-A, and by a Humboldt Research Award from the Alexander von Humboldt Foundation. CE acknowledges funding from the Deutsche Forschungsgemeinschaft (DFG) Sachbeihilfe, grant number BI1546/3-1. J.P. acknowledges support from the Programme National “Physique et Chimie du Milieu Interstellaire” (PCMI) of CNRS/INSU with INC/INP co-funded by CEA and CNES. E.R. acknowledges the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), funding reference number RGPIN-2022-03499. E.S. acknowledges funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement no. 694343).

References

- Bešlić, I., Barnes, A. T., Bigiel, F., et al. 2021, *MNRAS*, **506**, 963
 Bigiel, F., Leroy, A. K., Jiménez-Donaire, M. J., et al. 2016, *ApJ*, **822**, L26
 Cady, F. M., & Bates, R. H. T. 1980, *Opt. Lett.*, **5**, 438
 Caldu-Primo, A., Schruba, A., Walter, F., et al. 2013, *AJ*, **146**, 150
 CASA Team (Bean, B., et al.) 2022, *PASP*, **134**, 114501
 Chen, C.-T. J., Hickox, R. C., Alberts, S., et al. 2013, *ApJ*, **773**, 3
 Delhaize, J., Meyer, M. J., Staveley-Smith, L., & Boyle, B. J. 2013, *MNRAS*, **433**, 1398
 den Brok, J. S., Chatzigiannakis, D., Bigiel, F., et al. 2021, *MNRAS*, **504**, 3221
 den Brok, J. S., Bigiel, F., Sliwa, K., et al. 2022, *A&A*, **662**, A89
 Gallagher, M. J., Leroy, A. K., Bigiel, F., et al. 2018a, *ApJ*, **868**, L38
 Gallagher, M. J., Leroy, A. K., Bigiel, F., et al. 2018b, *ApJ*, **858**, 90
 Hickox, R. C., Jones, C., Forman, W. R., et al. 2007, *ApJ*, **671**, 1365
 Jiménez-Donaire, M. J., Bigiel, F., Leroy, A. K., et al. 2017, *MNRAS*, **466**, 49
 Jiménez-Donaire, M. J., Bigiel, F., Leroy, A. K., et al. 2019, *ApJ*, **880**, 127
 Jolly, J.-B., Knudsen, K. K., & Stanley, F. 2020, *MNRAS*, **499**, 3992
 Karim, A., Schinnerer, E., Martínez-Sansigre, A., et al. 2011, *ApJ*, **730**, 61
 Knudsen, K. K., van der Werf, P., Franx, M., et al. 2005, *ApJ*, **632**, L9
 Koch, E., Rosolowsky, E., & Leroy, A. K. 2018, *RNAAS*, **2**, 220
 Leroy, A. K., Hughes, A., Liu, D., et al. 2021a, *ApJS*, **255**, 19
 Leroy, A. K., Schinnerer, E., Hughes, A., et al. 2021b, *ApJS*, **257**, 43
 Lindroos, L., Knudsen, K. K., Fan, L., et al. 2016, *MNRAS*, **462**, 1192
 Neumann, L., Gallagher, M. J., Bigiel, F., et al. 2023, *MNRAS*, **521**, 3348
 Schruba, A., Leroy, A. K., Walter, F., et al. 2011, *AJ*, **142**, 37
 Usero, A., Leroy, A. K., Walter, F., et al. 2015, *AJ*, **150**, 115
 van der Hulst, J. M., Terlouw, J. P., Begeman, K. G., Zwitter, W., & Roelfsema, P. R. 1992, in *Astronomical Data Analysis Software and Systems I*, ed. D. M. Worrall, C. Biemesderfer, & J. Barnes, *Astronomical Society of the Pacific Conference Series*, **25**, 131

Appendix A: Additional material

Table A.1 lists the detection fraction of the prior $F_{\text{det}} = N_{\text{det}}/N_{\text{tot}}$ in each radial bin, where N_{det} is the number of prior-detected pixels and N_{tot} is the total number of pixels in that bin. In Fig. A.1, we compare the S/N measured from the stacked spectra and the expected S/N that is inferred from the S/N/100 cube considered as a pure noise cube as described in Sect. 3.1. In Fig. A.2, we show the spectral stacking as a function of galactocentric radius similar to Sect. 3 but using the template data cube as prior instead of the S/N/1 cube. In this case we have perfect alignment of the velocity field, and are not limited by the significance of the prior. In Fig. A.3, we show the radial trend obtained by taking the average spectrum over the prior-detected spectra only (i.e. by dividing the summed spectra by N_{det} instead of N_{tot} ; Section 2).

Table A.1. Prior detection fraction per radial bin.

r_{gal} [kpc] (1)	N_{tot} (2)	N_{det} (3)	F_{det} [%] (4)
0 – 1	23	22	95.7
1 – 2	72	61	84.7
2 – 3	128	79	61.7
3 – 4	176	131	74.4
4 – 5	230	107	46.5
5 – 6	278	101	36.3
6 – 7	328	54	16.5
7 – 8	372	11	3.0
8 – 9	322	6	1.9

Notes. (1) Radial bins, as illustrated in Fig. 2. (2) Total number of spectra, i.e. pixels in moment-0 map, inside the respective bin. (3) Number of spectra, where the prior, i.e. “S/N/1”, has been detected thus allowing velocity shuffling and spectral stacking. (4) Fraction of spectra used for stacking.

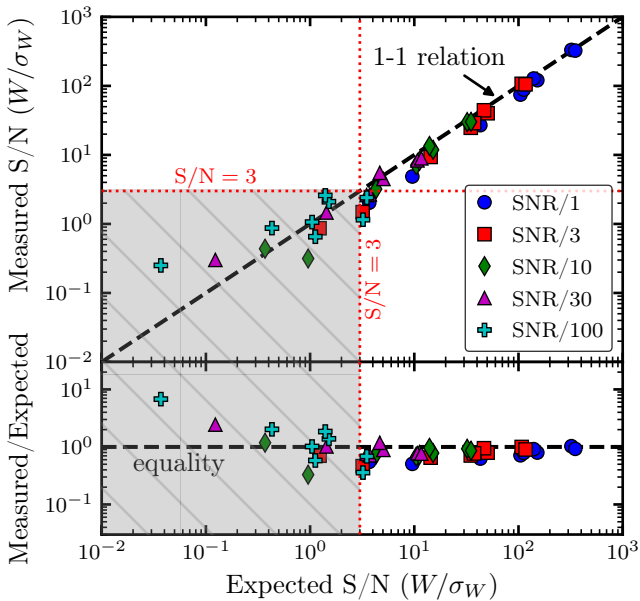


Fig. A.1. Measured vs expected signal-to-noise ratio. *Top:* Comparison between measured (12m+7m+tp) and expected signal-to-noise ratio of the integrated stacks. *Bottom:* Ratio of the measured to the expected S/N against the expected uncertainties.

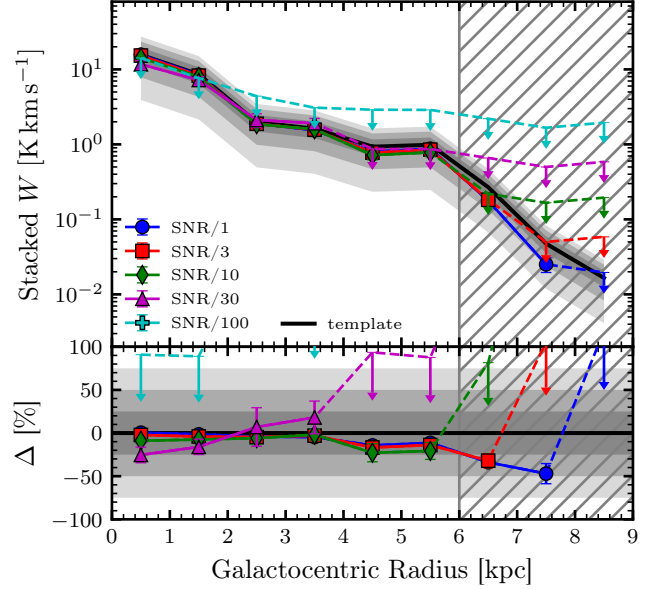


Fig. A.2. Template velocity field. Radial stacking similar to Fig. 2, but using the template (i.e. the true intensity distribution) as prior to aligning the velocity field.

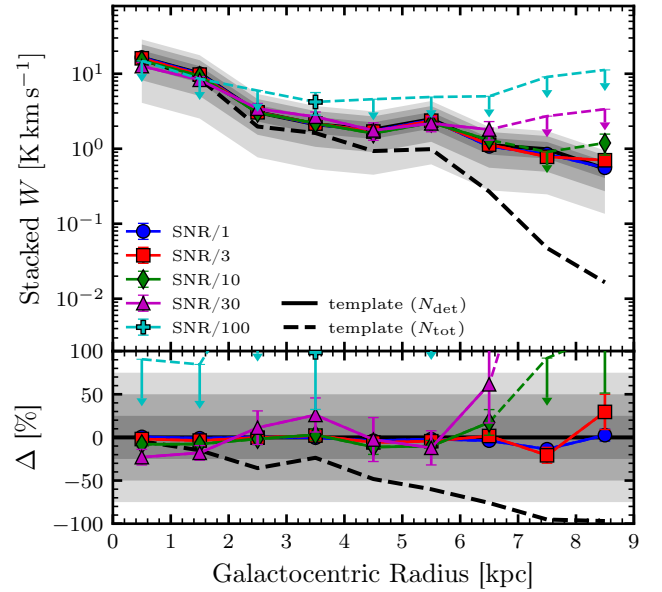


Fig. A.3. Average spectrum over prior-detected pixels. Radial stacking similar to Fig. 2, but computing the average over the prior-detected spectra only, as given by Eq. A.1.

Equation 2 then changes to

$$T_{n,\text{stack}}(v) = \frac{1}{N_{\text{det}}(n)} \sum_{i=0}^{N_{\text{det}}(n)} T_{n,i}(v). \quad (\text{A.1})$$

We find that the recovered stacks, computed from the prior-detected pixels, agree very well and without significant bias with the expected values measured in the prior-detected pixels (indicated by the solid line in Fig. A.3). However, using this method does not recover the true mean radial trend (dashed line), but, by construction, only considers the pixels, where the prior is detected, and is thus biased high, especially at radii where the prior detection fraction is low.