

ALMACAL

XIII. Evolution of the CO luminosity function and the molecular gas mass density out to $z \sim 6$

Victoria Bollo^{1,★}, Céline Péroux^{1,2}, Martin Zwaan¹, Aleksandra Hamanowicz³, Jianhang Chen⁴, Simon Weng^{5,6}, Claudia del P. Lagos^{6,7,8}, Matías Bravo⁹, Rob J. Ivison^{1,6,10,11}, and Andrew Biggs¹²

¹ European Southern Observatory, Karl-Schwarzschild-Str. 2, 85748 Garching near Munich, Germany

² Aix Marseille Univ., CNRS, LAM (Laboratoire d’Astrophysique de Marseille), UMR 7326, F-13388 Marseille, France

³ Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA

⁴ Max-Planck-Institut für Extraterrestrische Physik (MPE), Giessenbachstrasse 1, D-85748 Garching, Germany

⁵ Sydney Institute for Astronomy, School of Physics A28, University of Sydney, 2006, NSW, Australia

⁶ ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), ACT 2611, Australia

⁷ International Centre for Radio Astronomy Research (ICRAR), M468, University of Western Australia, 35 Stirling Hwy, Crawley, WA 6009, Australia

⁸ Cosmic Dawn Center (DAWN), København, Denmark

⁹ Department of Physics & Astronomy, McMaster University, 1280 Main Street W, Hamilton, ON L8S 4M1, Canada

¹⁰ Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK

¹¹ School of Cosmic Physics, Dublin Institute for Advanced Studies, 31 Fitzwilliam Place, Dublin D02 XF86, Ireland

¹² UK Astronomy Technology Centre, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK

Received 29 November 2024 / Accepted 6 February 2025

ABSTRACT

Cold molecular gas, largely traced by CO emission, is the primary fuel for star formation, making it essential for understanding galaxy evolution. ALMA has made significant progress in the study of the cosmic evolution of cold molecular gas. Here, we exploit the ALMACAL survey to address issues relating to small sample sizes and cosmic variance, utilising calibration data from ALMA to compile a statistically significant and essentially unbiased sample of CO-selected galaxies. By employing a novel statistical approach to emission-line classification using semi-analytical models, we place strong constraints on the CO luminosity function and the cosmic evolution of molecular gas mass density (ρ_{H_2}) back to $z \sim 6$. The cosmic molecular gas mass density increases with redshift, peaking around $z \sim 1.5$, then slowly declines towards higher redshifts by ~ 1 dex. Our findings confirm the key role of molecular gas in fuelling star formation. The new ρ_{H_2} estimates allow us to revisit the cosmic baryon cycle, showing that the ratio of molecular gas-to-stellar mass density is consistent with the so-called bathtub model of baryons, which implies a continuous replenishment of gas. The cosmic gas depletion timescale, estimated on a global scale, is shown to be fairly constant at all redshifts. We emphasise the importance of surveys using multiple small fields rather than a single contiguous area to mitigate the effects of cosmic variance.

Key words. evolution – galaxies: evolution – intergalactic medium – galaxies: ISM – galaxies: luminosity function, mass function

1. Introduction

The baryon cycle is a key driver of galaxy evolution, involving the inflow and outflow of the gas that regulates star formation over cosmic time (Péroux & Howk 2020; Walter et al. 2020). Galaxies form and grow within the framework of the baryon cycle, whereby gas is accreted from the inter- and circum-galactic media, primarily in the form of neutral atomic hydrogen (H I), which eventually cools and condenses into molecular hydrogen (H₂), the raw material for star formation (Tumlinson et al. 2017; Tacconi et al. 2020).

Understanding the baryon cycle is crucial for interpreting two fundamental cosmic quantities: the cosmic star formation rate density (SFRD) and the stellar mass function. The cosmic SFRD, $\psi_{\star}(z)$, tracks the rate at which stars are formed across the Universe as a function of redshift. It peaks around two billion years after the Big Bang ($z \sim 2$), then declines towards the present day (Madau & Dickinson 2014). The cosmic stel-

lar mass function, $\rho_{\star}(z)$, represents the cumulative mass of stars within galaxies, providing insights into the efficiency of star formation and how galaxies build up their stellar content over time. Together, these two observables are key to understanding the history of star formation in the Universe.

Cold gas – atomic and molecular – plays a critical role in the evolution of these cosmic quantities (Carilli & Walter 2013; Saintonge & Catinella 2022). Atomic hydrogen (H I) is the largest, more diffuse gas reservoir in galaxies, which must cool and condense to form molecular clouds, from which stars ultimately form. At low redshift, H I is typically observed through its 21-cm emission (e.g. Zwaan et al. 2005; Jones et al. 2018), while at high redshift it is detected through absorption lines (e.g. Péroux et al. 2003). These methods provide a direct way to map the distribution of atomic gas and its role in galaxy evolution. However, molecular gas is a more direct precursor to star formation, which takes place in the dense core of cold gas clouds (Kennicutt & Evans 2012; Tacconi et al. 2020). Because H₂ lacks a permanent dipole moment, we generally

* Corresponding author; victoria.bollo@eso.org

rely on carbon monoxide (CO) as its tracer (Carilli & Walter 2013; Bolatto et al. 2013). Observatories such as the Atacama Large Millimeter Array (ALMA), the Northern Extended Millimeter Array (NOEMA) operated by the Institute for Radio Astronomy in the Millimetre Range (IRAM), the Very Large Array, the Atacama Pathfinder EXperiment (APEX), and the Owens Valley Radio Observatory (OVRO), among others, have enabled the detection of CO across a wide range of redshifts, allowing us to estimate the molecular gas content in galaxies (Hodge & da Cunha 2020).

Another key observable is the CO luminosity function (CO LF), which quantifies the distribution of CO luminosities in galaxies. It provides a valuable diagnostic for studying the molecular gas properties and its role in galaxy evolution. By constructing the CO LF across different cosmic epochs, previous works have traced the evolution of the molecular gas mass density across different redshifts, $\rho_{\text{H}_2}(z)$. For instance, the ALMA Spectroscopic Survey in the Hubble Ultra Deep Field (ASPECS) survey (Decarli et al. 2016; Walter et al. 2016) detected CO in galaxies across $z = 1\text{--}3$ using ALMA bands 3 and 6, covering areas of 4.6 and 2.9 arcmin 2 , respectively (Decarli et al. 2019, 2020). Boogaard et al. (2023) combined the results of ASPECS with NOEMA observations to constrain the bright end of the CO LF up to $z \sim 6$, based on observations of the Hubble Deep Field North (HDFN), spanning 8.5 arcmin 2 . Lenkić et al. (2020) explored CO emission lines in 110 main-sequence galaxies using fields from the original PHIBSS survey (Guilloteau et al. 1992), covering a total area of ~ 130 arcmin 2 . The COLDz survey detected CO in galaxies at $z = 2\text{--}3$ and at $z = 5\text{--}7$ using more than 320 h of time with the Very Large Array (Pavesi et al. 2018; Riechers et al. 2019, 2020a), spanning ~ 60 arcmin 2 . Audibert et al. (2022) exploited the ALMA calibrator catalogue to study the CO LF up to redshift $z \sim 2.5$ and to examine the role of radio emission in galaxy evolution. In addition to CO-based measurements, molecular gas masses have also been inferred from dust emission using scaling relations based on stellar mass or star formation rates (Berta et al. 2013; Scoville et al. 2017; Magnelli et al. 2020) and from the [C II] emission line at higher redshifts (e.g. Aravena et al. 2024; Dessauges-Zavadsky et al. 2020; Zanella et al. 2018). Despite these advancements, the molecular gas mass density at several redshift bins in these surveys exhibit significant measurement uncertainties, sometimes exceeding 1 dex, highlighting ongoing challenges in accurately measuring molecular gas content.

Accurate measurements of the cosmic molecular gas mass density face two primary obstacles: cosmic variance and Poisson uncertainties. Cosmic variance reflects variations in galaxy properties across different regions of the Universe, while Poisson errors arise from the low number of detected galaxies in each luminosity and redshift bin. For instance, bright sources – which are sparsely distributed – can significantly increase shot noise. On the other hand, faint sources contribute to clustering noise, where their uneven distribution over large cosmic structures impacts measurements of the matter distribution in the Universe.

The choice of the observed region can introduce significant uncertainties when measuring cosmic properties, such as the number density of galaxies and LF. For example, Driver & Robotham (2010) used the Sloan Digital Sky Survey (SDSS) with galaxies at redshift $z = 0.03\text{--}0.1$ to study how the cosmic variance changes with survey volume and field shape, providing a formula to quantify these variations. Similarly, Moster et al. (2011) used N -body simulations to calculate the excess variance caused by cosmic effects, showing that

it significantly exceeds Poisson variance, which only accounts for random sampling. Both studies concluded that surveys that only probe limited volumes inevitably provide biased estimates of cosmic statistics. This makes it challenging to derive accurate and representative measurements of molecular gas over cosmic time.

Simulations play a crucial role in overcoming observational limitations by providing theoretical predictions of CO emission and molecular gas distribution in galaxies. Hydrodynamical simulations and semi-analytical models (SAMs) have successfully reproduced a wide range of galaxy properties such as UV-to-FIR LFs of galaxies, the number counts, and the redshift distribution of sub-millimetre galaxies (SMGs; Narayanan et al. 2012; Somerville & Davé 2015).

Current cosmological simulations do not directly include a cold gas component due to the complexity of the underlying physics. We therefore rely on post-processing to derive molecular gas properties. There has been a growing interest in using simulations to model the CO LF and the molecular gas density at several redshifts (Lagos et al. 2012, 2015, 2018; Popping et al. 2016; Vallini et al. 2018; Popping et al. 2019a,b; Maio et al. 2022; Béthermin et al. 2022; Bisigello et al. 2022; Garcia et al. 2023; Guo et al. 2023; Lagos et al. 2024; Ragone-Figueroa et al. 2024). However, reproducing the bright end of the CO LF remains difficult, and molecular gas density predictions have not yet converged.

Recent studies have explored the impact of cosmic variance on observables like the CO LF using simulated data. For instance, Popping et al. (2019b) compared observational data from the ASPECS survey with predictions from cosmological simulations, including post-processed IllustrisTNG (Pillepich et al. 2018) and the Santa Cruz semi-analytical model (Somerville et al. 2001). They found that cosmic variance has a stronger impact at lower redshifts ($z < 1$) since the same area of sky at higher redshifts has a larger volume. Similarly, Keenan et al. (2020) used simulations to find that for a survey of ~ 50 arcmin 2 with CO luminosities of around $10^{10} \text{ K km s}^{-1} \text{ pc}^2$, the uncertainties from Poisson noise and cosmic variance become comparable. Also, Gkogkou et al. (2022) found that using only Poisson variance under-estimates the total uncertainty by up to 80%, particularly for lower luminosity sources and larger survey areas. They reported that at redshifts $z < 3$, cosmic variance can introduce up to 40% uncertainty in molecular gas density estimates in small surveys (e.g. 4.6 arcmin 2 , similar to ASPECS), but this drops to below 10% for surveys covering more than 43.2 arcmin 2 .

Observations have begun to acknowledge cosmic variance as a key factor in explaining discrepancies between molecular gas density estimates across different surveys (Lenkić et al. 2020; Boogaard et al. 2023). This marks a shift from previous studies, which often neglected to include cosmic variance in uncertainty estimates, resulting in optimistic error estimates.

This paper aims to alleviate the limitation introduced by cosmic variance when estimating the molecular gas mass density by exploiting ALMA calibrator data, as part of the ALMACAL-22 survey (Zwaan et al. 2022). We make use of the sample selected and described in a previous paper of this series (Bollo et al. 2024), consisting of the highest-quality and deepest data compiled from 2016 to 2022. This study builds upon the methodology and results of the previous ALMACAL-CO survey (Hamanowicz et al. 2022). The proof-of-concept pilot of this earlier work successfully demonstrated the feasibility of using ALMA calibration data to characterise CO-selected galaxies and estimate their molecular gas content. We use the

statistical approaches and classification techniques introduced in that study in this extended analysis. With the larger dataset now available, we can refine our results and place stronger constraints on the CO LF and the evolution of the molecular gas mass density across cosmic time, fully aligning with the expectations set by the previous work.

This paper is organised as follows. In Sect. 2, we describe the ALMACAL-22 survey and summarise the calibration process, the selection of the sample, and the imaging. Sect. 3 details the steps taken to search for serendipitous detections of CO emission lines, including the source finding algorithm used (Sect. 3.1), the completeness correction (Sect. 3.2), the estimation of the fidelity (Sect. 3.3) and the redshift (Sect. 3.4), and the determination of the final catalogue (Sect. 3.5). Sect. 4 presents the volume estimates (Sect. 4.1), the CO LF (Sect. 4.2), the Schechter fits (Sect. 4.3), the molecular gas mass density (Sect. 4.4), the results of adopting the lowest CO transition possible (Sect. 4.5), and the uncertainties of our results (Sect. 4.6). Sect. 5 compares our findings for the CO LF (Sect. 5.1) and the molecular gas mass density (Sect. 5.2) with observations and simulations, discusses the effect of cosmic variance (Sect. 5.3), and provides a complete census of the baryon cycle (Sect. 5.4). In Sect. 6, we summarise our key conclusions. Throughout this paper, we use $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_M = 0.3$, and $\Omega_\Lambda = 0.7$. The present-day cosmological critical density is $\rho_{0,\text{crit}} \simeq 277.4 h^2 \text{ M}_\odot \text{ kpc}^{-3}$.

2. ALMACAL-22 and sample selection

ALMACAL-22 exploits the archival ALMA calibrator sources and their fields observed since 2016 to produce scientific outcomes (Zwaan et al. 2022). ALMA calibrators are typically bright quasars, observed for a few minutes in every ALMA PI-led observing project. Each calibration scan has a set-up that matches the PI requirements, so the full ALMACAL survey has a diverse range of configurations in terms of spatial and spectral resolution, sensitivity and integration time. Most of the calibrators in ALMACAL are bright sub-millimetre point sources classified as blazars (Bonato et al. 2018, Weng et al., in prep.).

This work uses the latest data release, which comprises the ALMA calibrator data taken up until 2022 May, so-called ALMACAL-22 (Bollo et al. 2024). The full ALMACAL survey comprises more than 30 TB of calibrator data from ALMA Cycle 1 to Cycle 10, over more than 1000 fields. Bollo et al. (2024) defined a pruned sample to represent the highest quality data, where each entry exceeds ten minutes of accumulated integration time. We focus the analysis on this pruned sample, which includes only observations made with the 12-metre array and files with low root-mean-square noise values ($\text{RMS} < 0.01 \text{ mJy}$ per channel). We summarise the main calibration steps, sample selection and data processing below.

A dedicated ALMACAL pipeline (Oteo et al. 2016) automates the processing of calibrator data and produces images of delivered ALMA datasets. The pipeline calibrates using `scriptForPI.py`, extracts calibrator data, applies self-calibration to correct phase and amplitude variability, and removes the bright calibrator. A point source model is applied during self-calibration, resulting in calibrated visibilities and the creation of data cubes for the ALMACAL-22 dataset.

Due to the uneven nature of the ALMACAL data, a pruning procedure was applied to select the highest-quality data. The pruning steps are based on key properties like integration time, frequency coverage, spatial resolution and RMS noise. The on-source integration of observations covering the same frequencies was set to accumulate more than 10 minutes to reach a meaning-

Table 1. Properties of the CO transitions in the ALMACAL-22 survey

Transition	ν_{rest} [GHz]	Redshift	Volume $\log[\text{cMpc}]^3$	N_{ind}	CV %
CO(1–0)	115.27	0–0.37	4.85	173	4.85
CO(2–1)	230.54	0–1.70	6.07	240	1.75
CO(3–2)	345.54	0–3.15	6.30	259	1.51
CO(4–3)	461.04	0–4.40	6.39	274	1.20
CO(5–4)	576.27	0–5.80	6.44	287	1.07
CO(6–5)	661.47	0–7.10	6.47	284	1.40
Total				7.06	299

Notes. N_{ind} is the number of independent sightlines in which a given transition can be detected. CV is the percentage of cosmic variance estimated using Eq. (7) (see Sect. 5.3 for more details).

ful sensitivity level. Only observations with noise levels consistent with theoretical sensitivity were selected. From an initial 34909 measurement sets (ms), only the $\sim 20\%$ with the highest quality remains in the pruned sample. The total number of data cubes and calibrators in the pruned sample is listed in Bollo et al. (2024), Table 1. In addition to the pruning, there is potentially a bias related to clustering effects around the calibrators, as these objects are often located in massive haloes and associated with large overdensities. The pruning fields and data in this work explicitly address this concern by removing CO lines coinciding with the redshifts of the calibrators. A detailed discussion is provided in Sect. 3.4.1.

Data cubes – consisting of two spatial dimensions and one spectral dimension – were created by combining multiple observations. The uv observations were concatenated into a single file to re-calibrate the uv weights using CASA’s `statwt` (Bean 2022). The choices made to optimise the imaging process can be summarised as follows. The beam size samples 3 pixels, and the image size is circular, with a ratio $1.8 \times$ that of the primary beam. The channel width is 31.2 MHz, a value that balances spectral resolution, S/N and data volume. Imaging was performed with CASA’s `tclean`, using linear interpolation, natural weighting, and $0.5''$ uv tapering. We took special care examining the shape of the synthesised beam when combining data from different arrays. By comparing the PSF shape to a 2D Gaussian fit, we found minimal deviations. As a result, the contributions to the residual map are negligible, ensuring that the flux measurements remain accurate and free from significant beam distortions. The pruned sample contains 1508 data cubes covering 401 different calibrator fields. Most of the cubes are in ALMA bands 3 and 6, with a maximum integration time of ~ 7 hours and a mean sensitivity of $\sim 0.78 \text{ mJy}$ per channel (see Bollo et al. 2024 for further details).

3. Line search

Detecting emission lines in more than 10^3 cubes with significant noise variations is challenging. To address this, we used the source-finding algorithm SoFiA2 (Westmeier et al. 2021), to identify potential candidates. We assessed these candidates’ reliability based on the S/N and line width detected, and we estimated the completeness factor of the entire line search process. In this section, we describe the identification of CO candidates (Sect. 3.1), the estimation of the completeness (Sect. 3.2), reliability (Sect. 3.3), the determination of redshifts (Sect. 3.4), and the final compilation of a catalogue of candidates (Sect. 3.5).

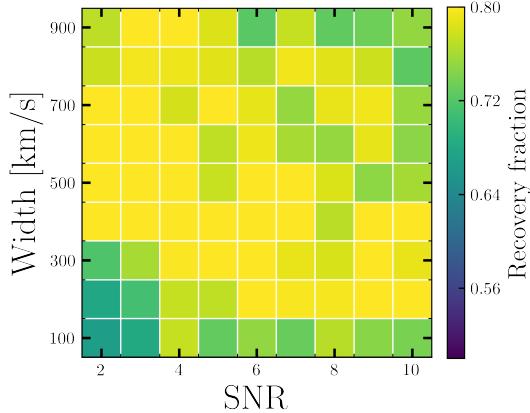


Fig. 1. Completeness fraction of mock sources. The grid shows several combinations of the injected peak signal-to-noise ratio (S/N) and line width. The heat map represents how successfully the algorithm, SoFiA2 (Westmeier et al. 2021), detects mock sources injected into data cubes, with the recovery fraction indicated by the colour. For sources with $S/N \gtrsim 4$ and line width $\gtrsim 300 \text{ km s}^{-1}$ we reach a completeness above $\sim 75\%$.

3.1. Source searching algorithm

SoFiA2 is a fully automated pipeline designed to find 3D sources in the WALLABY survey conducted on the Australian SKA Pathfinder (ASKAP). Based on the open access source finder SoFiA 1 (Serra et al. 2015), SoFiA2 has been re-written in C with multi-threading for speed, being considerably faster and more efficient than the previous version. It allows for searching on the spectral and spatial axes, offering two different algorithms: a simple threshold finder and the smooth and clip (S+C) finder. The threshold finder applies a flux threshold to the data, which is useful mainly for data adjusted for noise variations. The default S+C finder, described in Serra et al. 2012, iteratively smooths the data cube across various spatial and spectral scales to detect significant emission signals above a set threshold.

We input the processed ALMA data cubes after cleaning, calibrating, and based on multiple observations combined in uv space. We performed the search on data cubes that remained uncorrected for the shape of the primary beam response function. Before the primary beam correction, the data preserves its original sensitivity pattern to apply a uniform detection threshold to the emission line search. We flagged spectral channels for which the noise deviated from the median by more than 5σ . In SoFiA2, we applied spatial and spectral smoothing based on the median absolute deviation (MAD) with a 3σ threshold. This helps to automatically flag corrupted data, like channels with radio-frequency interference or pixels with residual continuum emission. SoFiA2 measures and adjusts for the local noise level using a running window of 25 pixels and 15 channels in the spatial and spectral domain.

We chose the smooth and clip (S+C) algorithm to identify sources, while smoothing to suppress noise and outliers. The algorithm iteratively rejects data points that deviate beyond the 3σ threshold, so extreme values do not affect the smoothing process. Then, the ‘linker’ combines detected pixels in the binary mask into coherent detections using a friends-of-friends algorithm. It links all pixels within a merging radius of three, treating sources as three-dimensional collections of pixels. After the source identification, SoFiA2 delivers the properties of the sources, such as their position, and size. Three co-authors (VB, CP, MZ) visually inspected the spectra and moment maps

of the candidates to ensure that they were not confused with noise peaks. The final catalogue contains 87 emission lines (see Sect. 3.5 for further details about the selection).

3.2. Completeness

We performed a completeness test to assess the quality of the cubes and the efficiency of the source-finding algorithm in retrieving artificial sources. To estimate the completeness factor, we injected mock sources into each data cube. We added artificial sources to each data cube and ran the source finder with the same parameters used to look for real emissions to quantify the recovery fraction. Iterating this process helps us to determine the quality of the cubes and algorithm’s robustness under different conditions.

A set of mock sources was generated with known flux densities, positions, and sizes. The S/N values of the synthetic sources range from 2 to 18, and the line widths vary between 100 and 1600 km s^{-1} . These ranges were chosen to explore the parameter space broadly. In each cube, 20 mock sources were randomly injected with varying properties. We ran SoFiA2 on the data cube containing both the real and injected sources. We estimated the percentage of recovered sources by comparing the position of detected sources with the injected mock sources.

We determined whether SoFiA successfully recovers the injected sources and records both true positives (correctly identified sources) and false negatives (missing sources). The process was repeated ten times per cube to build up statistically significant samples, with each iteration using different mock sources to simulate varying conditions.

We analysed the detection rate, or completeness factor (c_i), based on the S/N and line width of the mock sources across two categories: high-completeness and low-completeness cubes. Some cubes consistently showed a recovery fraction of zero across all iterations. Further investigation revealed that these cubes were either extremely narrow (significantly less than 4 GHz) or exhibited structured noise. Since real detections in these cubes would likely be missed and their missing data cannot be corrected, we excluded them from the sample. However, cubes with low but non-zero completeness are kept, as their incomplete information can be accounted for using the completeness factor, and they still contribute to the statistical inference in the analysis. Ultimately, we kept 1107 cubes, removing 401 cubes with zero completeness.

Fig. 1 displays the completeness grid for our sample, indicating the fraction of successfully recovered injected sources as a function of the signal-to-noise ratio (S/N) and line width within feasible detection limits. We find that synthetic sources with $S/N \gtrsim 4$ and line widths of $\gtrsim 300 \text{ km s}^{-1}$ are successfully recovered in over 70% of cases. Although the recovery fraction becomes roughly homogeneous after these values, there is a lack of monotonic improvement towards the top right corner. To investigate why the highest bin in the grid does not reach full completeness, we examined several mock sources with high S/Ns and broad velocity components. In most cases, these sources fell into narrow spectral cubes, where the continuum level is insufficient for comparison, leading the algorithm to miss them.

3.3. Fidelity

We estimated the reliability of emission lines based on the significance of the detections compared to the noise distribution in the data cubes. We inverted the cubes and ran the source finder

(SoFiA2) with the same parameters, again searching for emission lines (which are, of course, absorption features in the physical cube). As we do not expect absorption features, any emission lines in the inverted cubes must represent random noise peaks.

We examined the spatial and spectral extent of the identified sources, estimating their S/N and line width. This distribution of noise peaks can be defined through the fidelity coefficient, given by:

$$F(S/N, \sigma) = 1 - \frac{N_{\text{negative}}(S/N, \sigma)}{N_{\text{positive}}(S/N, \sigma)}, \quad (1)$$

where N_{negative} and N_{positive} are the number of negative and positive detections for a certain combination of S/N and line width.

Fig. 2, top panel, shows the 2D histogram of the positive candidates. The x axis represents S/N, while the y axis is the number of detection channels. We note that over 400 positive detections are initially found, but not all are real (see Sect. 3.5). The bottom panel of Fig. 2 shows the fidelity grid, built as is indicated by Eq. (1). There are cases where a high fidelity value is found for low S/N and small kernel widths (bottom left corner). This may result from a limitation of the source-finding algorithm, which can favour narrow lines and lead to poorly constrained statistics in that range. However, these values do not affect our estimates, since the majority of our detections lie well above this region. Emission line candidates spanning at least seven channels with $S/N > 7$ reach the highest fidelity values, as is indicated beyond the grey lines in the top and bottom panels of Fig. 2. Most of the high-fidelity sources are outside the grey region.

3.4. Redshift estimation

In ALMACAL-22, identifying emission line candidates to determine redshifts poses significant challenges. Some calibrator fields are observed multiple times using different ALMA bands, but the archival calibrator data provides uneven spectral coverage across different fields. When an emission line was detected in fields with coverage in multiple ALMA bands, the likelihood of covering the spectral range for a second CO emission line increases. By evaluating the probability of each detected line being a specific CO transition, we checked if the spectral coverage includes another transition in any of our data cubes. This analysis revealed a total of 37 fields where a second transition could have been potentially covered. However, in most cases (29 of those 37 fields), potential detections fell outside the field of view due to the shrinking primary beam in higher ALMA bands. For the other eight sources, the sensitivity is too low to detect fainter transitions. Ultimately, we did not identify any potential transitions in these fields, so we are reliant on detections of one emission line.

We took a two-step approach to estimate the redshift. First, we employed a probabilistic method to estimate the redshift (Sect. 3.4.1). Second, we utilised known spectroscopic or photometric information (Sect. 3.4.2).

3.4.1. Probabilistic approach

We created a redshift probability distribution using the semi-analytical model, SHARK-2 (Lagos et al. 2024), in the same way as is presented in Hamanowicz et al. (2022) with a previous version, SHARK-1 (Lagos et al. 2018). SHARK-2 includes improvements in physical models tracking the properties of supermassive black holes and active galactic nuclei feedback, and the environmental effects of satellite galaxies.

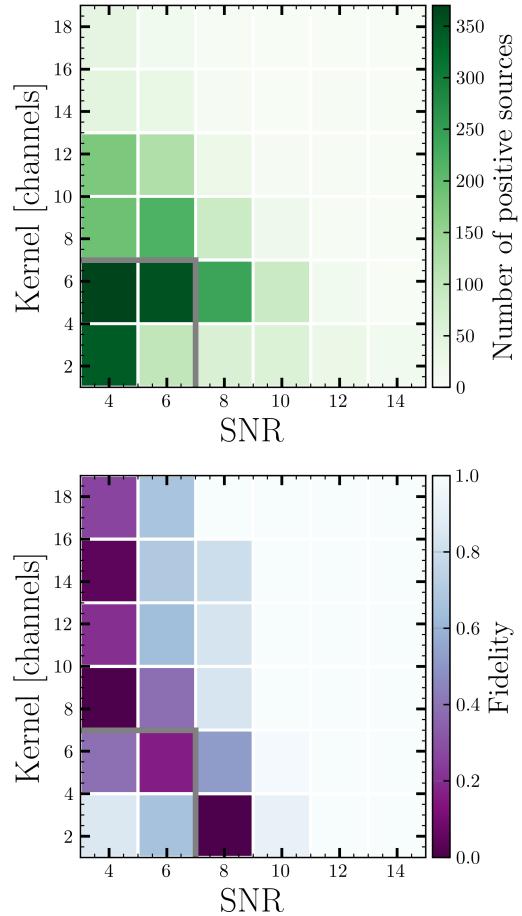


Fig. 2. *Top:* 2D histogram of the number of positive sources detected in all the data cubes when searching for emission lines using SoFiA2. The x axis represents the S/N of the line and the y axis shows the number of channels spanned by each detection. *Bottom:* fidelity grid derived from Eq. (1) considering the number of positive and negative sources found for a given combination of S/N and number of channels. Values close to unity represent the highest fidelity, while zero represents the least reliable parameters. Lines with S/N above seven and spanning about seven channels usually have high fidelity values, i.e. beyond the grey lines of both panels.

Based on the galaxy population of the SAM, we created a 2D histogram that maps CO transition flux against redshift. We then calculated a probability coefficient for each combination of flux and redshift by dividing the number of objects in each bin by the total number of galaxies in the simulated sample. This approach allowed us to determine the likelihood of a given CO flux originating at a specific redshift. High-level transitions ($J_{\text{up}} = 7, 8, 9, 10$) are rare since very high kinetic temperatures and densities are needed to excite the CO molecule sufficiently to populate these higher rotational levels. The most likely detections are transitions $J_{\text{up}} = 1, 2, 3, 4, 5$, and 6, placing our candidates between redshifts 0 and 6. Instead of simply selecting the J -transition with the highest probability for each detection, we adopted a more comprehensive approach. We included all plausible J transitions, each weighted according to the probabilities evaluated by SHARK-2. This method allows us to account for the uncertainty in identifying the exact J transition.

Fig. 3 shows the probability distribution of all the CO candidates in our sample. SHARK-2 assigns the CO(2–1) transition most frequently in our catalogue, and CO(1–0) as the least

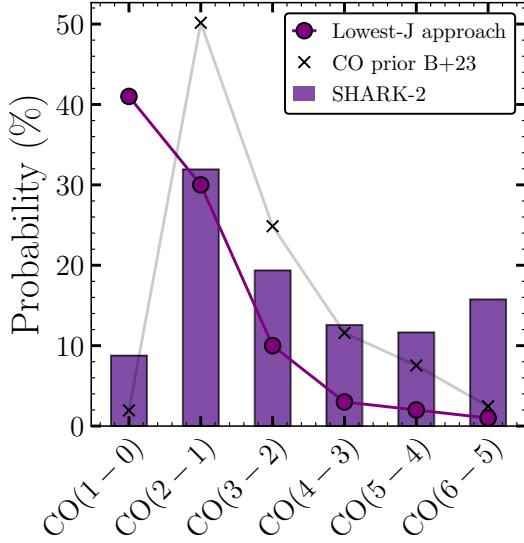


Fig. 3. Probability distribution of all the CO detections in our sample. Given the measured properties of each emission line, we estimated redshifts using a probabilistic approach, based on the semi-analytical model, SHARK-2 (Lagos et al. 2024) (see Sect. 3.4.1). Overall, the CO(2–1) transition is the most likely to be detected in our sample, while the CO(1–0) transition is the least likely. This probability is carried into the further analysis. For comparison, we also include the CO prior redshift distribution association of sources with *HST* counterparts reported in Boogaard et al. (2023), written as B+23, and the probability distribution obtained adopting the lowest- J approach explained in Sect. 4.5.

frequent. This distribution closely matches the shape of the CO line prior used for redshift associations in sources without counterparts seen in Boogaard et al. (2023). Two main factors influence the shape of this distribution. First, the CO molecule emits radiation at different frequencies depending on its rotational energy level (J level). Some transitions, like CO(2–1), are naturally more common, or stronger, while others are rarer or weaker. Second, the volume probed by our observations for each transition, especially CO(1–0), also plays a role, as lower-frequency transitions probe smaller volumes.

3.4.2. Existing spectroscopic and photometric data

We inspected NED, as well as MUSE and HST archives, to find the candidates’ optical/NIR counterparts. This search yielded seven MUSE cubes. From the MUSE cubes, we confirmed the spectroscopic redshift of three objects, all of them at $z < 1$. We fixed the known spectroscopic redshift for three sources and set the fidelity parameter to one. The other objects are not detected in any of the possible emission lines that MUSE could have covered. However, MUSE spectra do not cover any emission lines of sources at $1.4 \lesssim z \lesssim 2.8$. Another reason for not detecting these sources could be heavy dust obscuration, making a galaxy faint in the rest-frame UV range.

Photometric data are available for seven sources. We performed SED fitting with EAZY (Brammer et al. 2008) to estimate their photometric redshifts (photo- z). In three cases, the photo- z values match one of the highest probability redshifts given by SHARK-2; we adopted the photometric redshift for these sources. For the remaining objects, there are fewer than five photometric bands available, resulting in broad photo- z probability distributions that cannot reliably estimate the redshift or assign a redshift

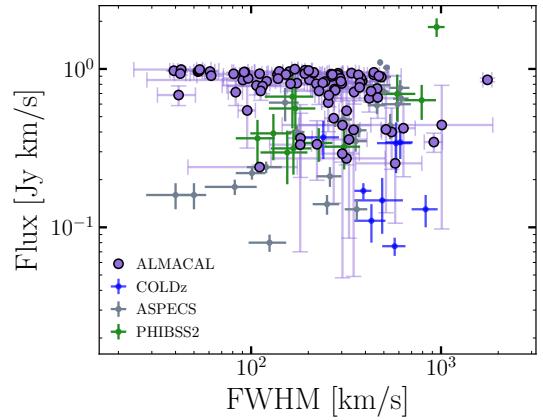


Fig. 4. Integrated flux and width of the CO emitters detected in ALMACAL-22 (purple). For comparison, we include detections from other surveys: ASPECS (Decarli et al. 2020), COLDz (Riechers et al. 2020a) and PHIBSS (Lenkić et al. 2020). Our candidates span a similar range of line widths as previous studies, and are generally brighter, as is expected from the difference in depth and volume.

probability greater than 10%. For these cases, we kept the redshift probability described earlier (Sect. 3.4.1).

3.5. Final catalogue

The final catalogue was created after a thorough visual inspection of all detections, along with an analysis of the spectra and the integrated emission-line map across the spectral range of the detected lines. We selected candidates with $S/N > 4$ and a clear peak in the moment map compared to the background noise. Duplicate detections – for example, those offset by only a few pixels but originating from the same region – were excluded. The final catalogue of CO emitters includes 87 sources.

We extracted the spectra using the aperture from SoFiA’s output catalogue. To ensure the accuracy of the number of channels spanned by the emission line, we re-calculated the S/N estimate by dividing the peak flux by the RMS of the moment map. The moment map was created by integrating across the frequency range of the emission and collapsing it into a single image. To measure the emission line flux, we used the function `scipy.optimize.curve_fit` to fit a Gaussian profile to the emission lines. We repeated this procedure 100 times for each candidate, perturbing the flux using the RMS of the cube. From the best fit, we use the mean values of the standard deviation and amplitude to estimate the flux and the full width at half maximum (FWHM). Our final catalogue of CO sources contains the flux and FWHM with their uncertainties, as well as the completeness and fidelity factors estimated from the S/N and line width of each object, reported in Table A.1 available at the CDS. Generally, we see that in fields where a lower rms noise level is reached, more detections are found. However, in some cases, fields with lower sensitivity still exhibit multiple detections, which can be attributed to the stochastic nature of source distribution within the field of view. Also, there are a few sources with a $FWHM \gtrsim 1000$ km/s, such as J0334–4008.4, most likely explained by blended sources for which the achieved resolution is insufficient to resolve them.

Fig. 4 shows the measured CO fluxes against the FWHM of each emission-line candidate. The FWHM values measured in ALMACAL-22 are comparable with previous surveys, while the fluxes are generally higher. This trend is expected due to the

differences in the sensitivity and volume probed by different surveys. The ALMACAL-22 survey benefits from a larger observational volume, but it has an uneven sensitivity distribution among data cubes. As a result, we can identify and measure the flux of more luminous sources, while detecting fainter systems is more challenging.

As an additional verification of the final catalogue, we estimate the probability of detecting high-redshift interlopers with bright [C II] emission, at 1900.53 GHz, instead of a CO transition. The [C II] emission line can be detected in one of the ALMA bands for redshifts $z > 4$, where it enters ALMA band 7. We estimate the volume covered by the [C II] line in band 7 between redshifts $z = 4$ and $z = 6$, to be 5858.16 cMpc^3 , using the same methodology employed to estimate the probed volume for CO transitions (see Sect. 4.1). We calculated the expected number of sources within this volume based on the [C II] LF in this redshift range (Yan et al. 2020; Casavecchia et al. 2024). We found that a maximum of two sources can be expected in a survey covering the volume probed by the [C II] line. We further discuss the effect of removing two band 7 detections in Sect. 4.6.

To confirm that our detections are independent of the calibrator, we verify that the redshifts of our sources differ from those of the calibrator. The calibrator redshift catalogue (z_{cal}) in ALMACAL-22 is based on a database from Bonato et al. (2018), which is extended through cross-matching with optical catalogues (NED, SIMBAD and Mahony et al. 2011) and supplemented by 70 VLT/X-shooter spectra (ID 111.253L.001, PI: S. Weng and ID 0101.A-0528, PI: E. Mahony). Our sample includes 87 CO emitters spread across 46 quasar fields, with redshift information available for 37 fields (Weng et al., in prep.). For each emission line, we compare the calibrator's redshift against all probable redshifts. Additionally, we verify that the estimated photometric redshift (Sect. 3.4.2) is distinct from that of the calibrator. Four candidates were excluded from our sample due to probable redshifts within 2000 km s^{-1} of the quasar redshift. Thus, we assert that our candidates are independent of the calibrator. Despite using data centred on calibrators, this approach does not introduce bias into our survey.

4. CO luminosity function and molecular gas mass density

A LF describes the statistical distribution of the number density of sources over a range of luminosities. In particular, the CO LF provides insight into the abundance of molecular gas in galaxies by revealing the variation in the number density of CO emitters. This section describes the methodology used to derive the CO LF and the results. First, we estimate the volume covered by the ALMACAL-22 survey in Sect. 4.1. We build the CO LF in Sect. 4.2 and fit Schechter functions in Sect. 4.3. We present the molecular gas mass density estimates at different redshift bins in Sect. 4.4. Finally, we explore the lowest- J approach in Sect. 4.5 and the uncertainties associated with our measurements in Sect. 4.6.

4.1. Volume estimation

We estimated the co-moving volume covered in our sample using the same methodology described in Hamanowicz et al. (2022). Here, we briefly summarise the steps for this calculation.

First, we defined the sky area covered by each ALMA cube, which depends on the frequency range and the field of view of the observations. For each CO transition ladder (from $J = 6$ to

$J = 0$), we calculated the co-moving volume by integrating over both the redshift element and the solid angle subtended by the detectable emission lines.

Second, we integrated the sensitivity variation across ALMA data cubes. The sensitivity decreases from the centre of the primary beam, following a pattern that resembles a Gaussian function. We estimated the lower limits of detectable volume, modelling an emission line with CO luminosities ranging from 10^5 to $10^{13} \text{ K km s}^{-1} \text{ pc}^2$. When the frequency coverage of the cube constrained the volume, we integrated the volume elements through concentric rings, each defined by the observed frequency. The resulting co-moving volumes for each CO transition are detailed in Table 1.

4.2. CO LF

We calculated the CO luminosity from CO fluxes using the following equation, from Solomon et al. (1997),

$$L'_{\text{CO}} = 3.25 \times 10^7 \frac{S_{\text{CO}} \Delta V}{(1+z)^3} \left(\frac{D_L}{\nu_{\text{obs}}} \right)^2 [\text{K km s}^{-1} \text{ pc}^2], \quad (2)$$

where L'_{CO} is in units of $\text{K km s}^{-1} \text{ pc}^2$, ν_{obs} is the observed frequency of the CO line in GHz, D_L is the luminosity distance of the galaxy in Mpc, z is the redshift, and $S_{\text{CO}} \Delta V$ is the integrated flux in Jy km s^{-1} . We converted the luminosity measured for mid- and high- J CO into CO(1–0), scaling by the empirical conversion factors from Boogaard et al. (2020): $r_{j \rightarrow 1} = \{3.33, 5.20, 4.76, 2.70, 0.53\}$ at $z < 2$ and $r_{j \rightarrow 1} = \{4.09, 8.24, 12.21, 14.68, 13.86\}$ at $z > 2$, for $J = 2, 3, 4, 5, 6$.

To construct the CO LF, we sampled over the uncertainties of the integrated flux and the conversion factor used to derive the CO(1–0) flux, the completeness and fidelity factor, and the redshift probability. The completeness factor depends on whether the emission lines were found in a cube with a high or low completeness factor (Sect. 3.2). The fidelity depends on the S/N and the line width of the emission lines. As was noted previously, if the line has been confirmed by a counterpart with a matching spectroscopic or photometric redshift, we assume the fidelity is equal to one.

We performed 1000 realisations of the CO LF over five independent bins, shifted by 0.2 dex, where each bin spans one dex in luminosity. We adopted the same methodology presented by Decarli et al. (2016), using independent shifted bins. This approach allows us to evaluate the dependence of the reconstructed CO LF on the bin definition. In each iteration, the redshift of each source will depend on the redshift probability calculator described earlier (Sect. 3.4). Hence, the same source will have different redshift values according to the weight of their probability in each realisation.

We binned galaxies by their CO luminosities, counted the number in each bin, and then normalised these counts by the co-moving volume. This gave us the number density of galaxies at different CO luminosities. We built the normalised LF, $\Phi(L)$, also known as the number density of objects per unit luminosity, per unit co-moving volume, using the following expression:

$$\Phi(\log L') \cdot d(\log L') = \sum_i \frac{F_i}{V_i \cdot c_i} [\text{Mpc}^{-1} \text{ dex}^{-1}], \quad (3)$$

where $\Phi(L')$ is the luminosity density per cubic megaparsec per dex, $d(\log L')$ is the luminosity bin, V_i is the comoving volume accessible for each transition, F_i is the fidelity and c_i is the completeness factor estimated in Sect. 3.2 and Eq. (1).

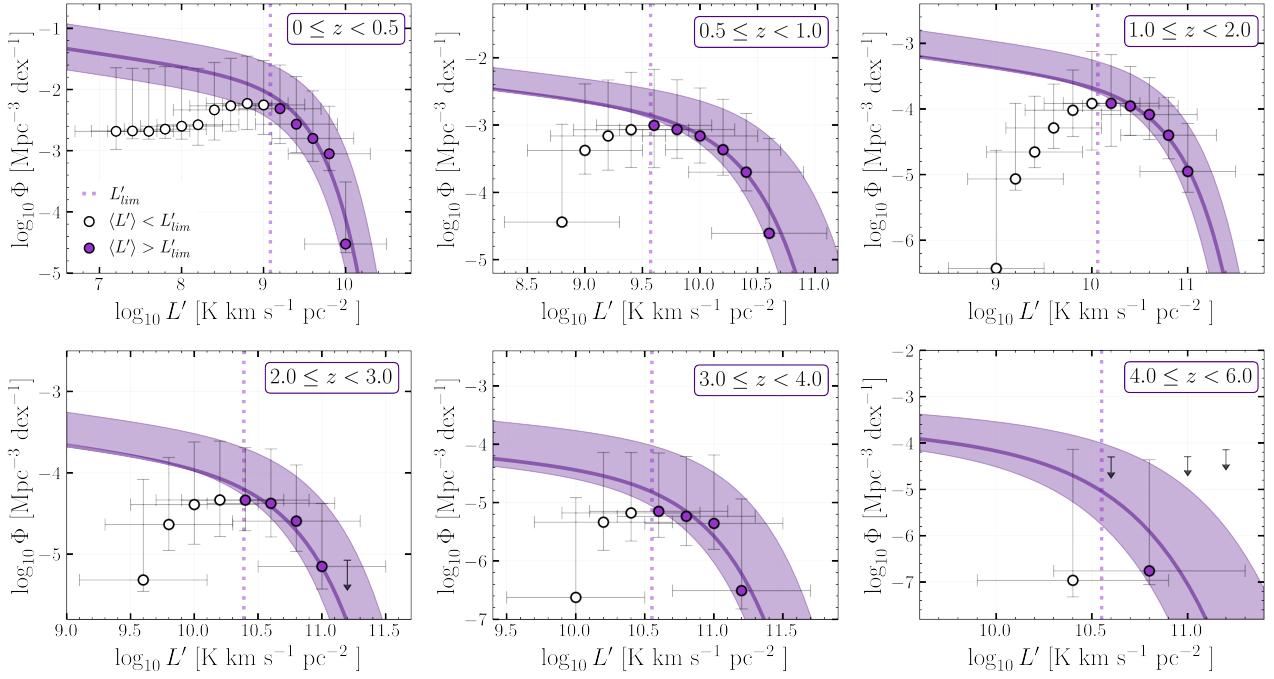


Fig. 5. CO LF across redshift bins: $z = 0\text{--}0.5$, $z = 0.5\text{--}1$, $z = 1\text{--}2$, $z = 2\text{--}3$, $z = 3\text{--}4$ and $z = 4\text{--}6$. Detections are shown in circles and non-detection as arrows. The vertical dotted line in each panel indicates the detection limit, representing the faintest luminosity detectable at a 50% reduction in primary beam sensitivity. The CO LFs are fitted using only the filled points that lie above this luminosity threshold. The best-fit Schechter function (Schechter 1976) is plotted, with Φ^* and L'^* allowed to vary freely, while the faint-end slope is fixed at $\alpha = -0.2$. The shaded region represents 1σ confidence intervals. Overall, the CO LF reveals a decreasing normalisation factor, while the ‘knee’ of the function shifts to higher luminosities as redshift increases.

The final CO LF was divided into six redshift bins: $0 \leq z < 0.5$, $0.5 \leq z < 1$, $1 \leq z < 2$, $2 \leq z < 3$, $3 \leq z < 4$, and $4 \leq z < 6$. Fig. 5 shows the median values with their uncertainties. The error bars show the natural spread of data derived from 1000 realisations. We calculated the 16th and 84th percentiles and the 1σ Poisson confidence intervals for each bin, which accounts for the statistical fluctuations, following the method described in Gehrels (1986). When a particular luminosity bin contains fewer than one source on average, we provide a 1σ upper limit to reflect the lower reliability of the data in that bin.

4.3. Schechter fits

We fitted the CO LF using an analytical Schechter function (Schechter 1976) in logarithmic scale as follows:

$$\log \Phi(L') = \log \Phi^* + \alpha \log \left(\frac{L'}{L'^*} \right) - \left(\frac{L'}{\ln(10)L'^*} \right) + \log(\ln 10), \quad (4)$$

where $\Phi(L')$ is the number of galaxies per co-moving volume with a CO luminosity between $\log L'$ and $\log L' + d(\log L')$. Φ^* is the normalisation factor, α is the faint-end slope, and L'^* indicates the luminosity at which the LF changes from a power law to an exponential function, also known as the ‘knee’ of the LF.

Fig. 5 shows the best-fit Schechter functions, with the shaded regions representing 1σ confidence intervals. Vertical lines indicate the luminosity limits, which were determined by considering the noise level of each ALMA cube and a 50% drop-off across the Gaussian-shaped primary beam. To ensure consistency, we adopted the limits from the shallowest cubes in the sample for each redshift bin and CO transition. The highest redshift panel contains only one data point above the luminosity limit with significant statistical weight. Nevertheless, we display

Table 2. Schechter best-fit parameters of the CO LF.

Redshift	$\log_{10} \Phi^*$ [Mpc $^{-3}$ dex $^{-1}$]	$\log_{10} L'^*$ [K km s $^{-1}$ pc $^{-2}$]
0.0–0.5	$-1.51^{+0.32}_{-0.37}$	$9.31^{+0.18}_{-0.09}$
0.5–1.0	$-2.48^{+0.51}_{-0.31}$	$10.12^{+0.17}_{-0.51}$
1.0–2.0	$-3.32^{+0.50}_{-0.17}$	$10.58^{+0.26}_{-0.18}$
2.0–3.0	$-3.61^{+0.59}_{-0.08}$	$10.59^{+0.30}_{-0.17}$
3.0–4.0	$-4.10^{+0.91}_{-0.29}$	$10.62^{+0.50}_{-0.18}$
4.0–6.0	$-3.54^{+0.40}_{-0.16}$	$10.16^{+0.33}_{-0.17}$

Notes. Fixed $\alpha = -0.2$.

the fit for completeness. To derive the Schechter function fits, we performed 500 realisations, sampling the CO LF within its uncertainties. We only included bins where the median luminosity exceeds the calculated luminosity limit (indicated by filled points in Fig. 5). We allowed the parameters Φ^* and L'^* to vary freely during the realisations. However, due to the limited number of data points at the faint end, we adopted a fixed slope at $\alpha = -0.2$, following the approach used by Boogaard et al. (2023). The median values and the uncertainties obtained from the best fits are reported in Table 2 for different redshift bins.

4.4. Molecular gas mass density

The molecular gas mass density refers to the total mass of molecular gas per unit volume in a given region of the Universe. It is often denoted as ρ_{H_2} or Ω_{H_2} when normalised to the critical

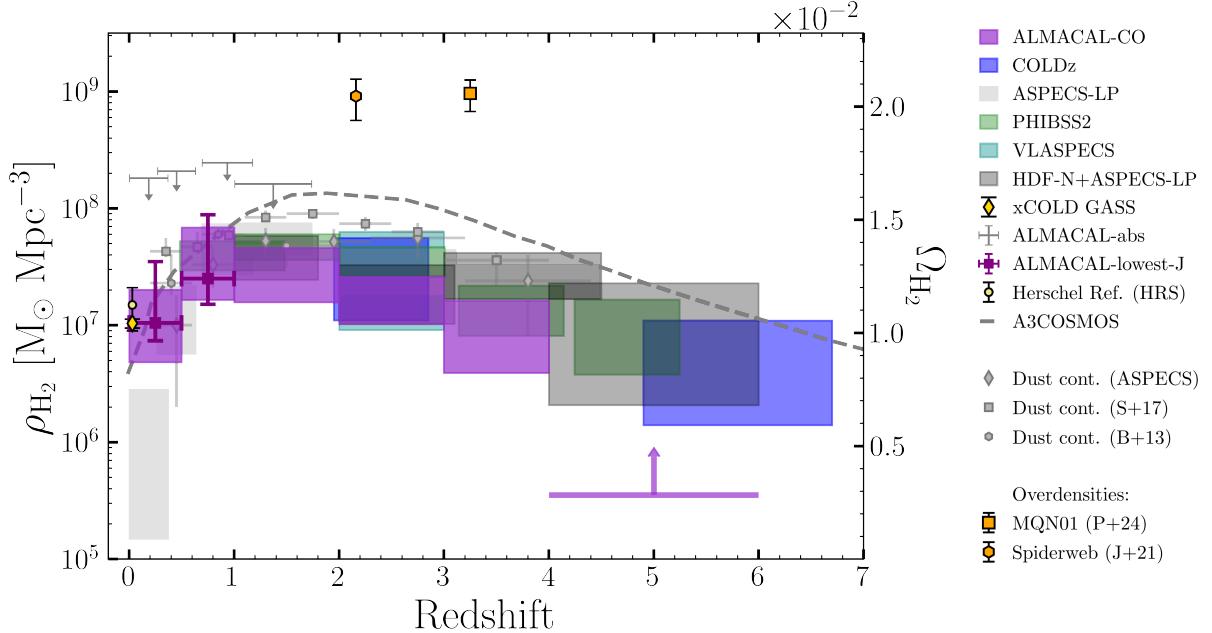


Fig. 6. Cosmic molecular gas mass density evolution measured from ALMACAL-22 in purple. The right y axis represents the unitless density parameter for molecular gas, $\Omega_{\text{H}_2} = \rho_{\text{H}_2}/\rho_{\text{crit},z=0}$. For comparison, we include estimates from other surveys: COLDz (Riechers et al. 2020a), ASPECS CO (Decarli et al. 2020) and dust measurements (Magnelli et al. 2020), PHIBSS (Lenkić et al. 2020), VLASPECS (Riechers et al. 2020a), HDF-N+ASPECS-LP (Boogaard et al. 2023), xCOLD GASS (Saintonge et al. 2017; Fletcher et al. 2021), ALMACAL in absorption (Klitsch et al. 2019b), ALMACAL-22 adopting the lowest- J approach (this work, see Sect. 4.5), from the K -band of the Herschel Reference Survey (HRS) (Andreani et al. 2020), A^3COSMOS (Liu et al. 2019), other estimates from dust continuum (Scoville et al. 2017; Berta et al. 2013), and from recent works reporting over-densities at $z \sim 2$ in the Spiderweb galaxy protocluster (Jin et al. 2021) and at $z \sim 3$ in the MUSE Quasar Nebula 01 (MQN01) field (Pensabene et al. 2024). We plot consistent estimates scaled to the same cosmology (H_0) used in this work and without helium contribution. We find overall consistency with the trend reported by other blind surveys, although slightly lower values are found at high redshift ($z > 1$).

density of the Universe. We derive the molecular gas mass of the CO emitters using the CO(1–0) luminosity and a conversion factor, α_{CO} , as follows:

$$M_{\text{H}_2} = \alpha_{\text{CO}} \cdot L'_{\text{CO}(1-0)} [\text{M}_\odot], \quad (5)$$

where we used $\alpha_{\text{CO}} = 3.6 \text{ M}_\odot (\text{K km s}^{-1} \text{ pc}^2)$ for all detections. Adopting this value makes the comparison with previous studies straightforward, since they use the same α_{CO} (Decarli et al. 2019; Riechers et al. 2019; Lenkić et al. 2020; Riechers et al. 2020a). A different constant value affects our results linearly.

We derive the molecular gas mass density, $\rho(M_{\text{H}_2})$, by integrating $L'_{\text{CO}} \Phi(L'_{\text{CO}})$ up to the luminosity limit and applying the conversion factor, α_{CO} . We exclude the same data points as before when fitting Schechter functions. Fig. 6 shows the estimates from ALMACAL-22 for each redshift bin, where the right y axis shows the unitless density parameter for molecular gas, $\Omega_{\text{H}_2} = \rho_{\text{H}_2}/\rho_{0,\text{crit}}$. The associated uncertainties were estimated in the same way as for the CO LF, explained above.

4.5. Lowest- J possible

We now explore the conservative approach of assuming the lowest- J CO transition detectable for all sources. Given the frequency of each line, we classify the CO emission line at the lowest possible redshift. This method focuses on transitions from $J = 1$ to $J = 4$, which are expected to be the brightest and the most likely to be observed. We performed 1000 realisations using the same luminosity and redshift bins. In each realisation, we perturbed the luminosity and the empirical scaling factor

Table 3. Cosmic molecular gas density from ALMACAL

Redshift	$\log \rho_{\text{H}_2} [\text{M}_\odot \text{ Mpc}^{-3}]$
0.0–0.5	$6.92^{+0.35}_{-0.37}$
0.5–1.0	$7.44^{+0.35}_{-0.38}$
1.0–2.0	$7.35^{+0.21}_{-0.51}$
2.0–3.0	$7.18^{+0.13}_{-0.47}$
3.0–4.0	$6.84^{+0.34}_{-0.35}$

to convert high- J into CO(1–0) within linear uncertainties. We applied the same estimates of the volume and the correction factors (completeness and fidelity) described in Sect. 4.2.

We built the CO LF using Eq. (3) and derived the molecular gas mass density using Eq. (5). Our results are displayed in Fig. 6 as purple boxes. These estimates match the previous results using the redshift probability calculator built from the semi-analytical model, SHARK-2, described in Sect. 3.4. The main difference is the uncertainty, which is larger for the estimates derived using a probabilistic approach. Using the lowest- J approach, most of our detections would be classified as CO(1–0), considering that most were found in band 3. This likely skews the distribution of CO across cosmic epochs, since it does not consider the volume probed by each transition. As a result, we may be over-estimating the molecular gas content through this approach. Still, it is a conservative way to validate the use of

models to estimate redshifts based on more complex CO line ratios, as was previously noted by [Hamanowicz et al. \(2022\)](#).

4.6. Possible biases and uncertainties

We identified the potential effects of the measured uncertainties on interpreting our results. One significant factor is the Poisson error, particularly in bins with very few objects ($\lesssim 5$), as we are probing nearly one dex in luminosity at the bright end of the LF. Additionally, the redshift uncertainties, particularly in converting high- J CO transitions into CO(1–0), could lead to the misclassification of transitions. Misinterpreting a higher- J transition as CO(1–0) would imply lower gas masses.

Another uncertainty stems from the uneven sensitivity of the data cubes. While we treat sensitivity on average, the data combined multiple observations with varying configurations, leading to changes in the noise across different spectral channels. Despite the efforts to correct the presence of false positives or missing detections through the completeness and reliability factors, we still allow for the possibility that some uncertainties are not fully understood. In some cases, we had to exclude data cubes where no mock sources were retrieved, resulting in a zero completeness factor. Low completeness and low reliability, especially for sources with low signal-to-noise ratio and narrow line width, add further uncertainty, particularly for the brightest sources.

We also explored the possibility that we are detecting other lines than CO, particularly the bright [C II] emission line, but the probabilities are quite low. This line is detectable in ALMA band 7 at high redshifts ($z > 4$), where the observed volume suggests a maximum of two detections. We observe five detections in band 7, and even if two of them in the high-redshift bin were removed, it would not affect the overall results, as the values in that bin are lower limits.

5. Discussion

This section presents our findings in the context of cosmic evolution. In Sect. 5.1 and Sect. 5.2, we compare our CO LF and the molecular gas mass density estimates with previous works using observations and simulations. In Sect. 5.3, we delve into the effects of the field-to-field variance of ALMACAL-22 compared with previous surveys. In Sect. 5.4, we investigate the evolution of baryons associated with galaxies averaged over cosmic time and space, including the new estimates from ALMACAL-22.

5.1. CO LF

5.1.1. Comparison with other observations

Fig. 7 shows the redshift evolution of the CO LF derived from $z = 0$ to $z = 6$ from the ALMACAL-22 survey. Our estimates are based on the CO(1–0) luminosity. For comparison, we also include measurements from previous surveys at similar redshifts. Findings from the COLDz survey ([Riechers et al. 2019](#)) ASPECS survey are also based on the $L'_{\text{CO}(1-0)}$. Estimates from PHIBSS2 ([Lenkić et al. 2023](#)), ASPECS ([Decarli et al. 2019, 2020](#)), and HDFN surveys ([Boogaard et al. 2023](#)) are derived from higher- J CO transitions, scaled to the CO(1–0) consistently with our measurements as is detailed above (Sect. 4.2).

At low redshift, $z \sim 0$ –1.0, our estimates indicate higher values of the number density of sources at a given luminosity, compared to findings from ASPECS and PHIBSS. At interme-

diate redshift, $z \sim 1$ –2, we find values slightly lower than the results from PHIBSS but consistent with ASPECS and HDFN. At higher redshift, $z \sim 2$ –6, our results are in good agreement with COLDz and HDFN, but slightly lower than ASPECS. At these redshifts, the estimates from PHIBSS are derived from high- J CO transitions, and the uncertainties converting them CO(1–0) may be larger, so they were included just for reference.

Another interesting way to explore the evolution of the CO LF is through their Schechter parameters. In Fig. 8, we show the evolution of the Schechter parameters derived in Sect. 4.3. We compare our results with the surveys mentioned above, including the parameters of the well-known CO LF in the local Universe reported by xCOLD ([Fletcher et al. 2021](#)). The normalisation factor (top panel), Φ^* , shows a decreasing trend towards higher redshift, from $z \sim 0$ to $z \sim 6$ by a factor of $\sim 3\times$. The characteristic luminosity (bottom panel), L'^* , increases from redshift $z \sim 0$ to $z \sim 2$ by a factor of $\sim 1.15\times$ and has a constant value towards higher redshift. This evolution suggests the CO luminosity in high-redshift galaxies is dominated by galaxies of $\log L'_{\text{CO}} \lesssim 10.5$ up to cosmic noon, after which brighter systems become more common. Overall, there is a good agreement between the parameters estimated from the ALMACAL-22 survey and previous studies. The largest difference lies in the low-redshift bins, for which the assumption of a fixed faint slope at $\alpha = -0.2$ may not be appropriate, as low- z studies have reported a different value ($\alpha \sim 1.2$, [Fletcher et al. 2021](#)). Further, examination of the evolution of the Schechter parameters of the CO LF is needed, particularly in the low-mass regime ($\log L'_{\text{CO}} \lesssim 9$), where the number density values are more difficult to constrain due to observational limitations.

5.1.2. Comparison with simulations

Several approaches have been used to model the physics of the CO emission line through hydrodynamical simulations and semi-analytical models. Hydrodynamical simulations simultaneously model the physics of gas, dark matter and stars within galaxies, providing detailed predictions about the formation of molecular gas like CO and H₂ (e.g. [Pelupessy & Papadopoulos 2009](#); [Obreschkow et al. 2009](#); [Narayanan et al. 2012](#); [Katz et al. 2017](#); [Vallini et al. 2018](#); [Casavecchia et al. 2024](#)). These simulations capture the small-scale physical processes, but their high computational cost limits the number of galaxies that can be simulated, especially over a wide range of properties. Also, their resolution is often insufficient to fully resolve small-scale processes like star formation ([Lagos et al. 2012](#)). On the other hand, semi-analytical models (SAMs) use simplified assumptions to describe the evolution of galaxies, such as symmetries in galaxies and halos. These models are usually run on halo catalogues which are made beforehand from an available dark-matter-only simulation (e.g. SURFS [Elahi et al. 2018](#)), and then simplified models for gas physics are applied. The main advantage of SAMs is their computational efficiency, allowing them to explore a wide range of values in the parameter space. SAMs have been successful in modelling the interstellar medium (ISM), star formation, and CO emission across different transitions, as well as considering factors like gas metallicity and radiation fields (e.g. UV and X-ray [Lagos et al. 2012](#); [Popping et al. 2014, 2016](#)). While hydrodynamical simulations provide detailed insights into physical processes, SAMs offer broader statistical power, complementing both approaches in understanding the physics of CO emission and galaxy evolution.

We display the predictions of the CO LF from simulations in Fig. 9. We compare the results from ALMACAL-22 with the

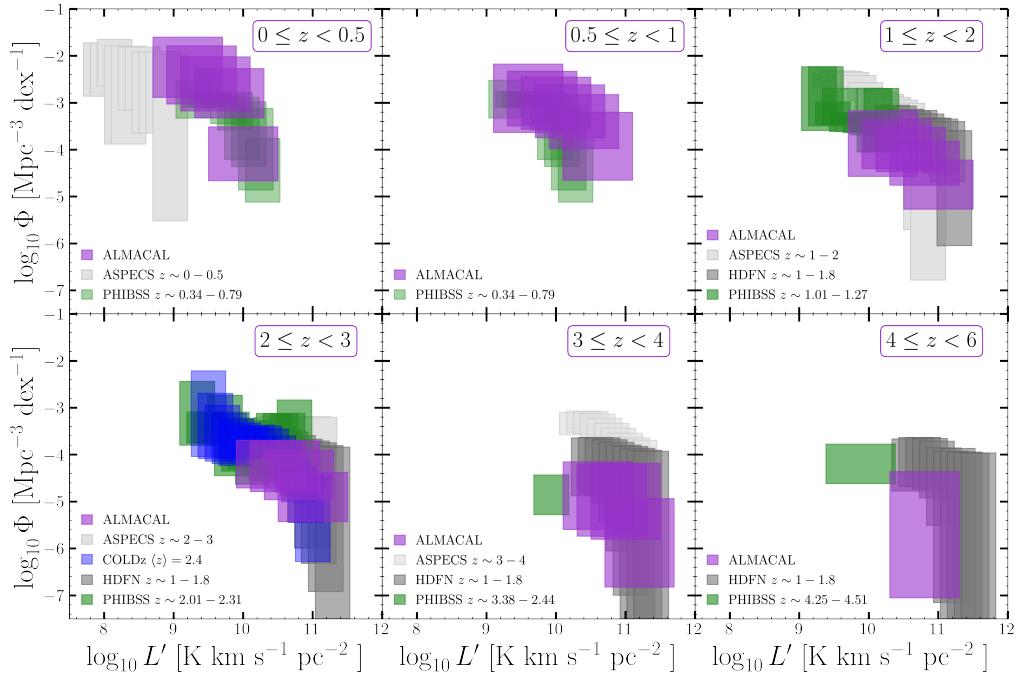


Fig. 7. CO LF in different redshift bins, from $z = 0$ to $z = 6$, as is indicated at the top right of each panel. The measurements from ALMACAL-22 are shown in purple, including the uncertainty of each bin given by the extension of the boxes. We derived the CO LF based on the CO(1–0) luminosity, calculated as is explained in Sect. 4.2. For comparison, we include observational constraints found by previous surveys at similar redshift ranges: ASPECS (Decarli et al. 2019, 2020), PHIBSS2 (Lenkić et al. 2023), COLDz (Riechers et al. 2019), HDFN (Boogaard et al. 2023). In some cases, the estimates from these surveys were derived from different CO transitions, here we converted the luminosities $L'_{\text{CO}(1-0)}$ consistently to our methodology. The CO LF derived from the ALMACAL survey aligns well with findings from other surveys, while probing a wider redshift range.

findings from the semi-analytical model SHARK-1 (Lagos et al. 2018), SHARK-2 (Lagos et al. 2024) and the phenomenological simulation SPRITZ (Bisigello et al. 2022), which is based on empirical and theoretical relations. At low redshift, $z \lesssim 1$, SPRITZ produces values that are very much consistent with ALMACAL, particularly at the bright end of the LF. In this range, both SHARK versions produce values slightly above our estimates. In the rest of the panels, simulations are in good agreement with our findings, particularly with SHARK-2, which effectively constrains the bright end of the CO LF. Although we use the redshift probability distribution derived from SHARK-2 as a prior in our analysis, the final results on the molecular gas mass density and luminosity function are decoupled from the simulation. This independence is demonstrated by the differences observed in the CO luminosity functions and the results from the lowest- J approach, explained in Sect. 4.5 and further discussed in Sect. 5.2.1.

SHARK-2 implements a new AGN feedback model, compared to SHARK-1, which reduces gas content and enhances quenching in massive galaxies, especially at lower redshifts, driving differences seen at higher luminosities. At lower masses, the updated model accounts for additional environmental processes like ram pressure and tidal stripping, affecting gas content and causing variations in the faint-end slope of the CO LF, particularly at lower redshifts where satellite galaxies are more common. We recall the need for hydrodynamical simulations to expand the simulated galaxy population towards the high-mass regime and probe fainter regimes of the LF. So far, the CO LF, particularly at high redshift, is still not available in a cosmological context; in other words, big boxes (Bisigello et al. 2022; Garcia et al. 2023).

5.2. Molecular gas mass density evolution

5.2.1. Comparison with other observations

Fig. 6 compares the molecular gas mass density estimates, ρ_{H_2} , from ALMACAL-22 with previous works. We include the estimates from ASPECS (Decarli et al. 2019, 2020), COLDz (Riechers et al. 2020a), PHIBSS (Lenkić et al. 2020), VLASPECS (Riechers et al. 2020a), ALMACAL in absorption (Klitsch et al. 2019a), dust measurement from Berta et al. (2013), Scoville et al. (2017), Magnelli et al. (2020), estimates in the local Universe from xCOLD GASS (Saintonge et al. 2017; Fletcher et al. 2021) and Andreani et al. (2020), estimates at redshift $z \sim 2.2$ and $z \sim 3.2$ for over-dense systems from (Jin et al. 2021; Pensabene et al. 2024). We also included the estimates from the A³COSMOS dataset (Liu et al. 2019), which uses different calibration methods to derive the molecular gas mass, including CO lines, SED-fitted dust mass and Rayleigh-Jeans(RJ)-tail dust continuum.

In the lower redshift bins ($z \sim 0$ –1), our estimates are consistent with the findings from xCOLD GASS and Andreani et al. (2020), and slightly above the estimate by ASPECS. The lowest- J approach, explained in Sect. 4.5, shows excellent agreement with the main methodology, which relies on redshift estimated through SHARK-2, demonstrating that our results remain independent. At $z \sim 1$ –2, we agree with the results from PHIBSS, but are below HDFN and ASPECS. At $z \sim 2$ our estimates are consistent with VLASPECS, while PHIBSS, while HDFN and ASPECS report higher values. At $z \sim 3$ –4 we find consistency with PHIBSS, but the other surveys show higher values. In the highest redshift panel, the integration of the CO LF does not provide fully consistent measurement, as none of the individual

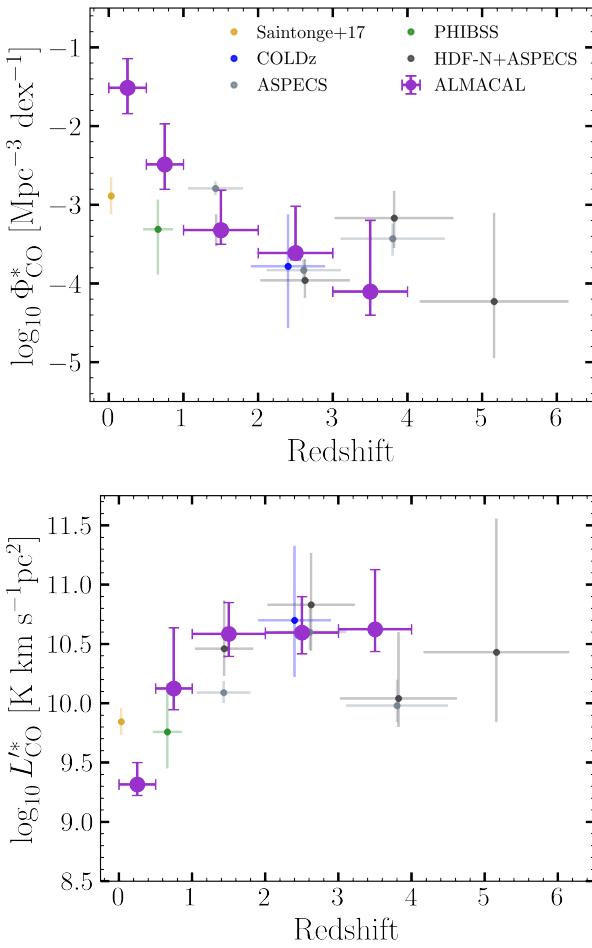


Fig. 8. Evolution of the Schechter best-fit parameters for the CO LF across $z \sim 0$ to $z \sim 6$. *Top*: evolution of Φ^* shows a decrease with redshift from $z = 0$. *Bottom*: evolution of L_{CO}^* shows a consistent increase from $z \sim 0$ to $z \sim 2$, and remains roughly constant at higher redshifts. For the fits shown in Fig. 5, the faint-end slope was fixed at $\alpha = -0.2$.

measurements is completely above the luminosity limit, so we provide a lower limit. Estimates of the molecular gas mass density from dust continuum Berta et al. (2013), Scoville et al. (2017), Magnelli et al. (2020) are usually slightly above our estimates but still follow the same evolutionary trend. High-redshift estimates of regions classified as over-densities are reported by Jin et al. (2021) and Pensabene et al. (2024). These estimates are ~ 1.5 dex above the trend followed by blind surveys, including ALMACAL, as one might expect. In particular, Pensabene et al. (2024) provides estimates, based on ALMA bands 3 and 6, of a sky region comparable in size to the ASPECS large programme ($\sim 4 \text{ arcmin}^2$ vs. 4.2 arcmin^2). Their detection of over-densities at such high redshifts further highlights the significant impact of cosmic variance on cosmic measurements. This emphasises the critical need for surveys that cover multiple sky regions, rather than the common practice of focusing on a single contiguous area.

5.2.2. Comparison with simulations

Cosmological hydrodynamical simulations have also predicted the evolution of the molecular gas mass density, ρ_{H_2} . These simulations infer the molecular gas mass density by post-processing

the outputs of cosmological boxes. First, they simulate the large-scale distribution of galaxies and gas in cosmological boxes. Then, molecular gas content is estimated a posteriori by models of density or temperature that link the simulated gas properties to the presence of molecular gas. We compare our estimates with predictions from simulations in the top panel of Fig. 10. We include the predictions from the semi-analytical model SHARK-1 (Lagos et al. 2018), the cosmological simulation TNG (Popping et al. 2019a), EAGLE (Lagos et al. 2015), COLDSim (Maio et al. 2022; Casavecchia et al. 2024), and UniverseMachine Guo et al. (2023). Among these, COLDSim stands out by uniquely integrating time-dependent non-equilibrium chemistry with cosmological hydrodynamics, enabling direct modelling of molecular gas. We find a relatively good agreement between our estimate and the predictions by SHARK, EAGLE and COLDSim, while TNG and UniverseMachine have higher values.

We followed the model presented in Walter et al. (2020) to fit the molecular gas mass density evolution with redshift, compiling the current estimates. We included all the CO observations compiled in that work together with the new estimates from ALMACAL-22. We performed 500 realisations of the following functional fit proposed by Madau & Dickinson (2014):

$$\rho_{\text{H}_2} = \frac{a \cdot (1+z)^b}{1 + [(1+z)/c]^d}. \quad (6)$$

We find $a = (5.2^{+2.6}_{-1.8}) \times 10^6$, $b = 3.5^{+0.01}_{-0.07}$, $c = 2.2^{+0.2}_{-0.1}$, and $d = 5.9^{+0.4}_{-0.5}$. The bottom panel of Fig. 10 shows our results with a solid purple line, and the 1σ uncertainty is shown in the shaded region. The contribution of ALMACAL-22 to the molecular gas mass density estimation indicates lower values at high redshift. If we extrapolate our model to high redshift, we find an excellent agreement with the estimation of ρ_{H_2} at redshift $z \sim 7$ measured by the REBELS survey (Bouwens et al. 2022) from the [C II] emission line reported by Aravena et al. (2024), shown in the dark red point in the bottom panel of Fig. 10.

5.3. Cosmic variance

We explored the effect of the field-to-field variance in ALMACAL-22 in comparison with the previous surveys. The public simulation, SIDES (Béthermin et al. 2022), provides an estimate of the evolution of variance in the molecular gas mass density as a function of survey size, assuming $J = 2$ and a redshift range $z = 0.5$ –6. The estimate of the cosmic variance is affected by the redshift slice, since high-redshift bins are less prone to cosmic variance effects. We compare the effect of cosmic variance only considering the area covered by ALMACAL-22 and previous surveys in the top panel of Fig. 11.

Aside from the total area covered, the number of independent sight lines is key to estimating cosmic variance. We adapted Equation (4) from Driver & Robotham (2010) for a conical survey, replacing the transverse lengths A and B by πR as follows:

$$\zeta_{\text{CV}}(\%) = [219.7 - 52.4 \log_{10}(\pi R^2 \cdot 291.0) + 3.21(\log_{10}(\pi R^2 \cdot 291.0))^2] / \sqrt{N \cdot C/291}, \quad (7)$$

where R is the transverse length at the median redshift, C is the radial depth, and N is the number of independent sight lines, all in units of $h_{0.7}^{-1}$. We performed this calculation for each CO transition. We considered unique calibrator fields and data cubes where each line could be detected to account for the number of independent sight lines. Table 1 provides the estimates of the

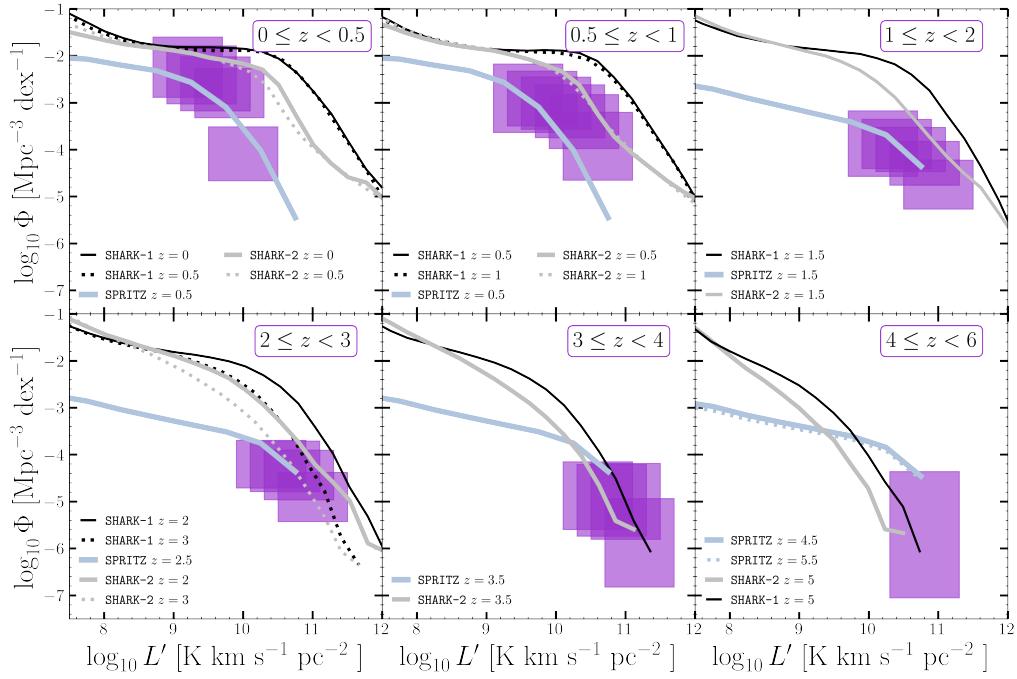


Fig. 9. CO LF from redshift $z = 0$ to $z = 6$ derived from ALMACAL-22 in comparison with simulations. We include the predictions from SHARK-1 (Lagos et al. 2018) at $z = 0$ –3, from SHARK-2 (Lagos et al. 2024) at $z = 0$ –5.5, and from SPRITZ (Bisigello et al. 2022) at $z = 0.5$ –5.5. We report consistency in most of the redshift ranges, and we recall the need for simulations to expand their range towards the bright end of the CO LF.

redshift range for each CO transition, the number of independent sight lines (N_{ind}), and the percentage of cosmic variance.

The bottom panel of Fig. 11 compares the cosmic variance from different surveys. We use the values of cosmic variance compiled in Lenkić et al. (2020) (Table 1 of their paper). We see that for all CO transitions, ALMACAL-22 has the lowest cosmic variance seen to date by at least a factor $\sim 5\times$, reaching values lower than 5%. The main reason that ALMACAL-22 has such low values of cosmic variance is that the survey comprises many independent calibrator fields instead of one contiguous area. Previous studies have acknowledged this issue, particularly for the ASPECS survey. Popping et al. (2019b) estimated the H_2 cosmic density in boxes covering the same volume covered in ASPECS and simulated the field-to-field variations. They found that cosmic variance can lead to variations up to a factor of $3\times$. This variance is particularly significant due to the small survey area of ASPECS, 4.6 arcmin^2 .

The formula provided by Driver & Robotham (2010) for estimating cosmic variance was originally intended for surveys measuring the variance in galaxy counts in fairly uniform galaxy populations. However, it has been noted by Keenan et al. (2020) that it does not take into account the variance in luminosity moments since they use galaxies in a restricted range of magnitude at $z \sim 0$. This issue becomes important for CO line surveys or molecular gas surveys, particularly for biased populations like bright CO emitters. A new method to estimate cosmic variance was proposed by Keenan et al. (2020) using simulated data. This new prescription includes the influence of galaxy number counts and how galaxies are distributed and clustered. We use the formula presented in Appendix A of Keenan et al. (2020) for mean brightness temperature to calculate the fractional uncertainty in the cosmic variance of different area and redshift intervals. We use the mean redshift, the redshift interval and the area covered by each field for a given CO transition to estimate the fractional uncertainty. We then combine the uncertainty of all the fields

as independent samples dividing by the square root of the number of independent sight lines ($\sqrt{N_{\text{ind}}}$). We obtain values consistent with the estimates derived from the formula proposed by Driver & Robotham (2010), as is shown in the bottom panel of Fig. 11.

Gkogkou et al. (2022) used the Uchuu cosmological simulation to analyse the impact of survey size on the variance of the luminosity function. They modelled the errors using both Poisson statistics and clustering contributions. The total variance is decomposed into Poisson and clustering components. They reported the field-to-field variance of simulations that replicate the redshift range and sizes of observational data from ASPECS and COLDz. The field-to-field variance introduced in the molecular gas density ranges from 77% to 81% in an ASPECS-like survey and from 57% to 73% for a COLDz-like survey. They also found a shift between the mean ρ_{H_2} of their model and the one from Keenan et al. (2020), which could be mainly explained by differences in the integration limits, the cosmological simulation used, the scaling relations, and the scatter to assign CO luminosities to galaxies. The predictions from Keenan et al. (2020) may be under-estimated by an order of magnitude according to the predictions from Gkogkou et al. (2022) at redshifts $z \sim 2$ –4. However, both approaches agree we need survey sizes of at least $\sim 70 \text{ arcmin}^2$ to prove the evolution at redshifts, $z \gtrsim 3$.

Overall, cosmic variance plays a significant role in the measurement of galaxy number density and LF, with uncertainties reaching as high as 70% for the smallest fields and decreasing to about 25% for the largest ones probed until now. Accounting for cosmic variance is essential to improve the accuracy of observational constraints on the cosmic star formation rate density and to deepen our understanding of the large-scale structure of the Universe. Figure 6 shows that the ALMACAL results are consistent with previous surveys; however, the uncertainties due to cosmic variance are not displayed. In Figure 11, we compare the impact of cosmic variance and we observe that this issue

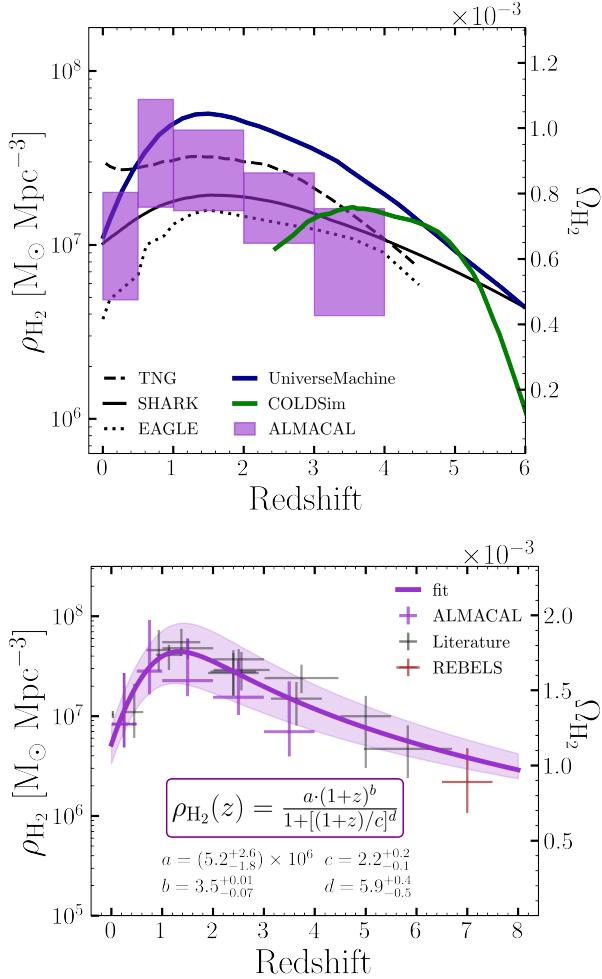


Fig. 10. *Top:* evolution of the molecular gas mass density obtained from ALMACAL-22 (purple) in comparison with predictions from simulations. We include the results from Illustris TNG (Popping et al. 2019b), EAGLE (Lagos et al. 2015), SHARK-2 (Lagos et al. 2018), UniverseMachine (Guo et al. 2023), and COLDsim (Maio et al. 2022). *Bottom:* we compiled several measurements from previous CO surveys, as is shown by the grey points, including ALMACAL-22 (see text). We present a new fit of the functional form used in Walter et al. 2020, including the ALMACAL-22 estimate, shown by the purple line, with the 1σ uncertainty shown by the shaded region. For reference, we add the estimates of ρ_{H_2} at the highest redshift measured up to date by the REBELS survey as the dark red point (Aravena et al. 2024), which is in excellent agreement with the extrapolation of our fit.

dominates other surveys. In contrast, due to its survey strategy, ALMACAL has a minimum value. To optimise survey strategies, it is beneficial to utilise multiple small fields instead of a single contiguous area, as this approach effectively mitigates cosmic variance. Future surveys are encouraged to consider the impact of cosmic variance when reporting uncertainties, moving beyond the traditional focus on Poisson uncertainties. The literature has shown that relying solely on Poisson uncertainties fails to capture the true levels of uncertainty (Popping et al. 2019b; Keenan et al. 2020; Gkogkou et al. 2022).

5.4. Census of the baryon cycle

In this section, we use the new ρ_{H_2} estimates provided by ALMACAL to revisit the cosmic baryon cycle. Specifically, we

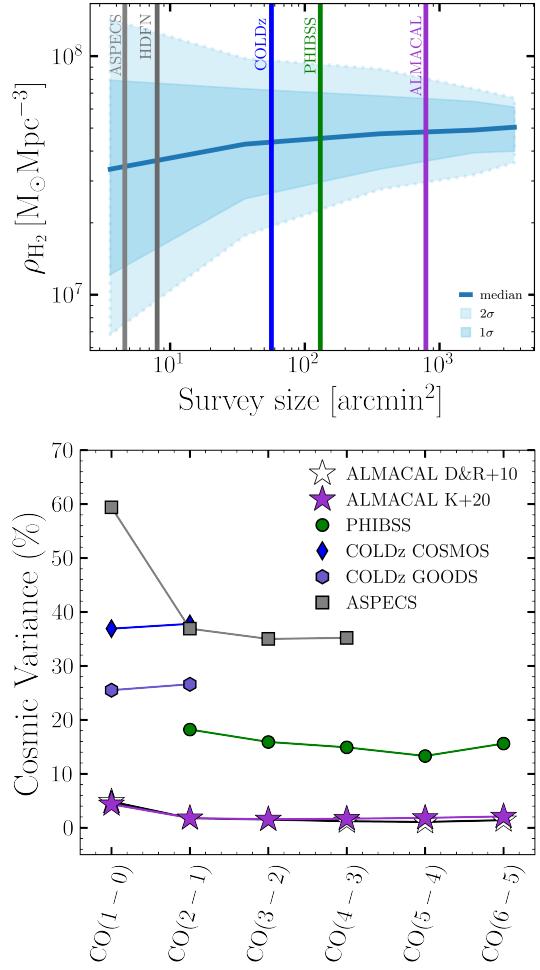


Fig. 11. *Top:* effect of cosmic variance based purely on the survey size. We show the results of the molecular gas mass density variation derived from the SIDES simulation (Béthermin et al. 2022; Gkogkou et al. 2022), publicly available for $J = 2$ at redshifts $z = 0.5$ –6. The solid blue line indicates the median value of ρ_{H_2} , and the dark and light areas represent the 1σ and 2σ variation due to the cosmic variance. We plot the area covered by ALMACAL-22 along with other CO surveys that have measured the molecular gas mass density. Overall, ALMACAL-22 is the survey less affected by cosmic variance to date. *Bottom:* cosmic variance estimation of ALMACAL-22 survey in purple, based on Driver & Robotham 2010 (D&R+10), for different CO transitions (Table 1). We also plot the variance derived from the formula presented by Keenan et al. (2020) (K+20), which agrees with the previous prescription. The figure shows estimates from ASPECS (Decarli et al. 2019, 2020), PHIBSS (Lenkić et al. 2020), COLDz GOODS and COSMOS (Pavesi et al. 2018; Riechers et al. 2019), previously reported in Lenkić et al. (2020). For every CO transition, ALMACAL-22 has the lowest effect of field-to-field variance ($\lesssim 5\%$).

investigate how the cosmic baryon density has evolved over time, focusing on the contributions from stars, atomic hydrogen (H I), and molecular gas (H_2).

Madau & Dickinson (2014) compiled data of two key observables: the cosmic star formation rate density, ψ_* , and the stellar mass density, ρ_* . Péroux & Howk (2020) used a Chabrier IMF (Chabrier 2003) in their review, and a return fraction of $R = 0.41$, which is the fraction of the stellar mass that is immediately returned to the gas when massive stars explode, whereas Walter et al. (2020) assumed a Salpeter IMF (Salpeter 1955) and $R = 0.27$. For the atomic gas mass density,

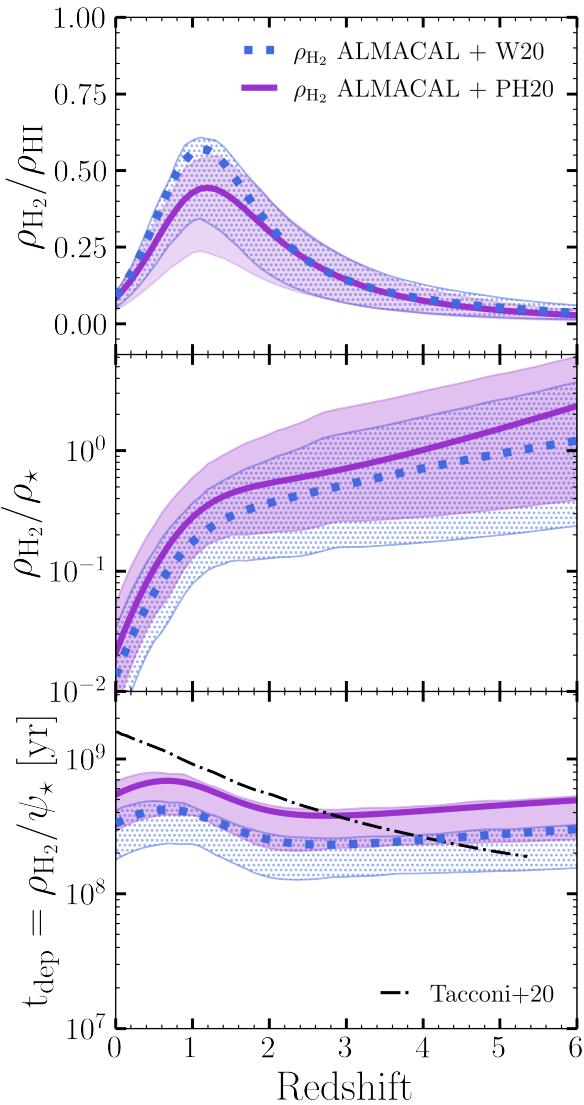


Fig. 12. Redshift evolution of baryonic component in the Universe. *Top*: ratio of cosmic molecular-to-atomic gas density as a function of redshift. *Middle*: ratio of molecular gas-to-stellar mass density as a function of redshift. *Bottom*: cosmic gas depletion timescale, is defined as the density of molecular gas over the cosmic star formation rate density. We include the molecular gas depletion times commonly derived for individual galaxies as the ratio $M_{\text{H}_2}/\text{SFR}$, as is reported in Tacconi et al. (2020). We use the fit of ρ_* presented by Madau & Dickinson (2014). We used the values of ρ_{HI} and ψ_* fitted in the compilation done by Péroux & Howk (2020) in dotted blue and Walter et al. (2020) in purple. For the star formation rate density and stellar mass function, the former assumes a Chabrier IMF (Chabrier 2003) and a return fraction $R = 0.41$, while the latter assumes a Salpeter IMF (Salpeter 1955) and $R = 0.27$. The ρ_{H_2} corresponds to the fit presented in Sect. 5.2, which includes the results from previous surveys that have measured CO along with ALMACAL.

Péroux & Howk (2020) modelled the evolution of neutral gas with redshift using a two-parameter power law, which uses fewer free parameters. On the other hand, Walter et al. (2020) adopted a hyperbolic tangent function, originally suggested by Prochaska & Neeleman (2018), to fit the atomic gas. In summary, we adopt the best-fit models for $\psi_*(z)$, $\rho_*(z)$ and $\rho_{\text{HI}}(z)$ from both Péroux & Howk (2020) and Walter et al. (2020). For

$\rho_{\text{H}_2}(z)$, we use the updated fit described in Sect. 5.2, which includes estimates from ALMACAL-22.

We analysed the gas density ratios of these three components (stars, atomic gas and molecular gas) in Fig. 12. In each panel, the solid purple lines indicate the ratios derived using $\rho_{\text{H}_2}(z)$ from this work, and the fits for ρ_{HI} , ψ_* and ρ_* provided in Walter et al. (2020). The dotted blue lines show the ratios using $\rho_{\text{H}_2}(z)$ including ALMACAL-22 and the fits for ψ_* , ρ_* and ρ_{HI} presented in Péroux & Howk (2020). For both models, the 1σ uncertainty is shown by the shaded regions.

The top panel shows the molecular-to-atomic gas density ratio as a function of redshift ($\rho_{\text{H}_2}/\rho_{\text{HI}}$). It peaks at around $z \sim 1.5$, indicating a period of high star formation activity, characterised by the conversion of HI to H_2 being particularly efficient. This efficiency is slightly lower than the one predicted by Walter et al. (2020) but still consistent with their predicted scenario of cosmic twilight, where the star formation activity in galaxies declines, likely due to the shutdown of the inflow and the accretion of gas.

The middle panel shows the ratio of molecular gas-to-stellar mass density (ρ_{H_2}/ρ_*) as a function of redshift. This quantity evolves significantly from $z \sim 1.5$ down to the local universe ($z = 0$), indicating that the molecular gas for star formation becomes more limited relative to the existing stellar mass. At high redshift ($z > 1.5$), our fits indicate a lower molecular gas, which means a lower capacity for star formation compared to that observed by Walter et al. (2020). The Chabrier IMF predicts about 30% lower total stellar mass for a given population than the Salpeter IMF, since fewer low-mass stars are included. This would suggest that galaxies still have significant molecular gas reservoirs available for star formation, particularly at high redshifts, which may not be true at low redshifts.

The bottom panel shows the cosmic gas-depletion timescale, defined as the density of molecular gas over the cosmic star formation rate density (ρ_{H_2}/ψ_*). The dot-dashed line shows the functional fit to the molecular depletion timescale based on $M_{\text{H}_2}/\text{SFR}$ ratio for a sample of massive galaxies (Tacconi et al. 2020), which roughly agrees with the global estimates from Péroux & Howk (2020). At redshifts above $z \sim 2$, we find a relatively constant depletion timescale, which indicates that molecular gas is consumed by star formation steadily during this period. The depletion timescale increases from $z \sim 2$ to $z = 0$, so the consumption rate is slowing down relative to the available gas, leading to the quenching of star formation. A Salpeter IMF would imply that galaxies deplete their gas more quickly, while a higher ratio (Chabrier IMF) would suggest galaxies have longer depletion timescales, consistent with galaxies at high redshift (Scoville et al. 2017). This trend is consistent with the findings from Péroux & Howk (2020) and their claim regarding the need to continuously replenish molecular gas to sustain star formation.

These findings align with predictions from the gas regulator model, also known as the ‘bathtub model’, as is described in studies like Bouché et al. (2010), Davé et al. (2011), Lilly et al. (2013), Peng & Maiolino (2014). This model uses continuity equations to track the flow of baryons into and out of galaxies and picture them as systems in a dynamic equilibrium between gas inflow, outflows and star formation. As has previously been noted by Péroux & Howk (2020), this model provides an effective framework for understanding how the continuous cycling of baryons drives galaxy evolution. In the early stages, gas accumulates, and star formation is primarily constrained by the available cold gas reservoir. As galaxies evolve, they reach a steady state where the net gas accretion rate regulates the overall star formation rate.

6. Conclusions

In this paper, we have presented the ALMACAL-CO project, built upon the experience of the previous pilot survey presented by Hamanowicz et al. (2022). We searched for serendipitous detections of CO emission lines in the fields of ALMA calibrators. We used the highest available quality of data, selected as part of the ALMACAL-22 data release presented in Bollo et al. (2024). We performed a search for CO emitters using the source-finding algorithm, SoFiA-2 (Westmeier et al. 2021). We detected 87 CO emitters over 299 calibrator fields. We confirmed the redshift of three sources with optical spectroscopic counterparts and estimated the photometric redshift of three sources with optical and near-IR counterparts. We determined redshift probabilities for the remaining 81 candidates based on the semi-analytical model, SHARK-2 (Lagos et al. 2024). We built the CO LF through realisations that sample the properties of the CO emitter, such as their completeness, fidelity and redshift probability. We constrained the CO LF and calculated the molecular gas mass density up to redshift, $z \sim 6$. We summarise our main findings as follows:

1. We have probed the bright end of the CO LF, from $z \sim 0$ to $z \sim 6$. We find good agreement with existing observations, but with slightly higher values of the number density of sources in the low redshift bins ($z \lesssim 1$) and slightly lower values in the higher redshift bins ($z \gtrsim 1$). Simulations of the CO LF (SHARK, SPRITZ, Lagos et al. 2018; Bisigello et al. 2022) agree with our findings, but we stress the need to expand their prediction to the bright end (Figs. 7 and 9).
2. We found an evolution of the Schechter parameters of the CO LF with redshift. The normalisation factor, Φ_{CO}^* , decreases with redshift. The characteristic luminosity, defined as the knee of the luminosity function, shows an increasing trend from the local Universe to $z \sim 2$ (Fig. 8).
3. The molecular gas mass density reported by ALMACAL-22 shows higher values than previous observations in the lower redshift bins ($z \lesssim 1$). It presents a slight tension at higher redshift ($z \gtrsim 1$), where ALMACAL-22 finds lower values (Fig. 6). We found a good agreement with the simulations predicting ρ_{H_2} from SHARK, EAGLE and COLDSim (Fig. 10, top panel).
4. We present a new analytical molecular gas mass density fit, compiling previous estimates based on CO surveys and the new ALMACAL-22 results. Irrespective of what CO line ρ_{H_2} is based on, the effect of cosmic variance is less than 5% (Fig. 10, bottom panel).
5. ALMACAL-22 is the survey that best addresses the uncertainty introduced by cosmic variance at the brighter end (Fig. 11).
6. We re-visit the redshift evolution of the baryon components, updating the molecular gas mass density with the values provided by ALMACAL. We use previous fits of the atomic gas mass density, stellar mass function and the cosmic star formation rate density from Péroux & Howk (2020) and Walter et al. (2020). Overall, we report consistency with their claims regarding the evolution of the density ratios, though with smaller molecular gas content (Fig. 12).

Our findings serve as a valuable stepping stone for future work that conducts surveys less prone to cosmic variance. We emphasise the need for additional telescope time to confirm the redshift of our sources, particularly those that may be obscured at optical wavelengths, where sub-millimetre observations would be crucial.

Probing the molecular gas density across different environments is a collective effort by the scientific community. The more data we gather, the clearer the picture becomes. Each survey brings unique strengths and limitations, and to fully understand the evolution of molecular gas in the Universe, we also need simulations with diverse prescriptions. Together, these approaches will help unravel the complexities of molecular gas evolution over cosmic time.

Data Availability

Table A.1 is available at the CDS via anonymous ftp to cdsarc.cds.unistra.fr (130.79.128.5) or via <https://cdsarc.cds.unistra.fr/viz-bin/cat/J/A+A/695/A163>. The ALMA Calibrator Source Catalogue is available at <https://almascience.eso.org/alma-data/calibrator-catalogue>. ALMACAL raw data and processed data cubes are available upon reasonable request. Please contact almacal@eso.org.

Acknowledgements. We thank L. Boogaard and I. Smail for their contribution to the analysis and discussion. This research was supported by the International Space Science Institute (ISSI) in Bern, through ISSI International Team project #564 (The Cosmic Baryon Cycle from Space).

References

- Andreani, P., Miyamoto, Y., Kaneko, H., et al. 2020, [A&A](https://doi.org/10.1051/0004-6361/201936140), 643, L11
 Aravena, M., Heintz, K., Dessauges-Zavadsky, M., et al. 2024, [A&A](https://doi.org/10.1051/0004-6361/202336140), 682, A24
 Audibert, A., Dasyra, K. M., Papachristou, M., et al. 2022, [A&A](https://doi.org/10.1051/0004-6361/202142000), 668, A67
 Berta, S., Lutz, D., Nordon, R., et al. 2013, [A&A](https://doi.org/10.1051/0004-6361/201220400), 555, L8
 Béthermin, M., Gkogkou, A., Cuyck, M. V., et al. 2022, [A&A](https://doi.org/10.1051/0004-6361/202142000), 667, A156
 Bisigello, M., Vallini, L., Gruppioni, C., et al. 2022, [A&A](https://doi.org/10.1051/0004-6361/202142000), 666, A193
 Bolatto, A. D., Wolfire, M., & Leroy, A. K. 2013, [AARA&A](https://doi.org/10.1088/2041-8205/51/2/207), 51, 207
 Bollo, V., Zwaan, M., Péroux, C., et al. 2024, [A&A](https://doi.org/10.1051/0004-6361/202336140), 690, A258
 Bonato, M., Liuzzo, E., Giannetti, A., et al. 2018, [MNRAS](https://doi.org/10.1093/mnras/478.1512), 478, 1512
 Boogaard, L. A., van der Werf, P., Weiss, A., et al. 2020, [ApJ](https://doi.org/10.3847/1538-4357/ab9021), 902, 109
 Boogaard, L. A., Decarli, R., Walter, F., et al. 2023, [ApJ](https://doi.org/10.3847/1538-4357/ab9451), 945, 111
 Bouché, N., Dekel, A., Genzel, R., et al. 2010, [ApJ](https://doi.org/10.1088/0004-637X/718/2/1001), 718, 1001
 Bouwens, R. J., Smit, R., Schouws, S., et al. 2022, [ApJ](https://doi.org/10.3847/1538-4357/ab9311), 931, 160
 Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2008, [ApJ](https://doi.org/10.1088/0004-637X/686/2/1503), 686, 1503
 Carilli, C. L., & Walter, F. 2013, [AARA&A](https://doi.org/10.1088/2041-8205/51/2/207), 51, 105
 Bean, B. 2022, [Publications of the Astronomical Society of the Pacific](https://doi.org/10.1088/0004-6285/134/11/114501), 134, 114501
 Casavecchia, B., Maio, U., Péroux, C., & Ciardi, B. 2024, [A&A](https://doi.org/10.1051/0004-6361/202336140), 689, A106
 Chabrier, G. 2003, [Publications of the Astronomical Society of the Pacific](https://doi.org/10.1088/0004-6285/115/3/763), 115, 763
 Davé, R., Oppenheimer, B. D., & Finlator, K. 2011, [MNRAS](https://doi.org/10.1093/mnras/415/1/11), 415, 11
 Decarli, R., Walter, F., Aravena, M., et al. 2016, [ApJ](https://doi.org/10.3847/1538-4357/833/2/69), 833, 69
 Decarli, R., Walter, F., Gómez-López, J., et al. 2019, [ApJ](https://doi.org/10.3847/1538-4357/882/2/138), 882, 138
 Decarli, R., Aravena, M., Boogaard, L., et al. 2020, [ApJ](https://doi.org/10.3847/1538-4357/902/2/110), 902, 110
 Dessauges-Zavadsky, M., Ginolfi, M., Pozzi, F., et al. 2020, [A&A](https://doi.org/10.1051/0004-6361/201936140), 643, A5
 Driver, S. P., & Robotham, A. S. G. 2010, [MNRAS](https://doi.org/10.1093/mnras/407.2.2131), 407, 2131
 Elahi, P. J., Welker, C., Power, C., et al. 2018, [MNRAS](https://doi.org/10.1093/mnras/475.5.5338), 475, 5338
 Fletcher, T. J., Saintonge, A., Soares, P. S., & Pontzen, A. 2021, [MNRAS](https://doi.org/10.1093/mnras/501.4.411), 501, 411
 Garcia, K., Narayanan, D., Popping, G., et al. 2023, [ApJ](https://doi.org/10.3847/1538-4357/974/2/14), 974, 14
 Gehrels, N. 1986, [ApJ](https://doi.org/10.3847/0004-637X/303/2/336), 303, 336
 Gkogkou, A., Béthermin, M., Lagache, G., et al. 2022, Arxiv e-prints [[arxiv:2212.02235](https://arxiv.org/abs/2212.02235)]
 Guilloteau, S., Delanney, J., Downes, D., et al. 1992, [A&A](https://doi.org/10.1051/0004-6361/1992162062401), 262, 624
 Guo, H., Wang, J., Jones, M. G., & Behroozi, P. 2023, [ApJ](https://doi.org/10.3847/1538-4357/955/2/57), 955, 57
 Hamanowicz, A., Zwaan, M. A., Péroux, C., et al. 2022, [MNRAS](https://doi.org/10.1093/mnras/519.3.34), 519, 34
 Hodge, J. A., & da Cunha, E. 2020, [R. Soc. Open Sci.](https://doi.org/10.1088/2041-8205/200556p), 7, 200556P
 Jin, S., Dannerbauer, H., Emonts, B., et al. 2021, [A&A](https://doi.org/10.1051/0004-6361/202036140), 652, A11
 Jones, M. G., Haynes, M. P., Giovanelli, R., & Moorman, C. 2018, [MNRAS](https://doi.org/10.1093/mnras/477.2.127), 477, 2
 Katz, H., Kimm, T., Sijacki, D., & Haehnelt, M. G. 2017, [MNRAS](https://doi.org/10.1093/mnras/468.4.4831), 468, 4831
 Keenan, R. P., Marrone, D. P., & Keating, G. K. 2020, [ApJ](https://doi.org/10.3847/1538-4357/904/2/127), 904, 127
 Kennicutt, R. C., & Evans, N. J. 2012, [ARA&A](https://doi.org/10.1088/2041-8205/50/2/531), 50, 531
 Klitsch, A., Péroux, C., Zwaan, M. A., et al. 2019a, [MNRAS](https://doi.org/10.1093/mnras/490.1220), 490, 1220
 Klitsch, A., Zwaan, M. A., Péroux, C., et al. 2019b, [MNRAS](https://doi.org/10.1093/mnras/482.1.L65), 482, L65

- Lagos, C. D. P., Lacey, C. G., & Baugh, C. M. 2012, arXiv e-prints [arXiv:1210.4974]
- Lagos, C. D. P., Crain, R. A., Schaye, J., et al. 2015, *MNRAS*, **452**, 3815
- Lagos, C. D. P., Tobar, R. J., Robotham, A. S. G., et al. 2018, *MNRAS*, **481**, 3573
- Lagos, C. D. P., Bravo, M., Tobar, R., et al. 2024, *MNRAS*, **531**, 3551
- Lenkić, L., Bolatto, A. D., Förster Schreiber, N. M., et al. 2020, *ApJ*, **159**, 190
- Lenkić, L., Bolatto, A. D., Fisher, D. B., et al. 2023, *ApJ*, **945**, 17
- Lilly, S. J., Carollo, C. M., Pipino, A., Renzini, A., & Peng, Y. 2013, *ApJ*, **772**, 119
- Liu, D., Schinnerer, E., Groves, B., et al. 2019, *ApJ*, **887**, 235
- Madau, P., & Dickinson, M. 2014, *AARA&A*, **52**, 415
- Magnelli, B., Boogaard, L., Decarli, R., et al. 2020, *ApJ*, **892**, 66
- Mahony, E. K., Sadler, E. M., Croom, S. M., et al. 2011, *MNRAS*, **417**, 2651
- Maio, U., Péroux, C., & Ciardi, B. 2022, *A&A*, **657**, A47
- Moster, B. P., Somerville, R. S., Newman, J. A., & Rix, H.-W. 2011, *ApJ*, **731**, 113
- Narayanan, D., Krumholz, M. R., Ostriker, E. C., & Hernquist, L. 2012, *MNRAS*, **421**, 3127
- Obreschkow, D., Croton, D., Lucia, G. D., Khochfar, S., & Rawlings, S. 2009, *ApJ*, **698**, 1467
- Oteo, I., Zwaan, M., Ivison, R., Smail, I., & Biggs, A. 2016, *The Messenger*, **164**, 41
- Pavesi, R., Sharon, C. E., Riechers, D. A., et al. 2018, *ApJ*, **864**, 49
- Pelupessy, F. I., & Papadopoulos, P. P. 2009, *ApJ*, **707**, 954
- Peng, Y.-J., & Maiolino, R. 2014, *MNRAS*, **443**, 3643
- Pensabene, A., Cantalupo, S., Cicone, C., et al. 2024, *A&A*, **684**, A119
- Péroux, C., & Howk, J. C. 2020, *ARA&A*, **58**, 363
- Péroux, C., McMahon, R. G., Storrie-Lombardi, L. J., & Irwin, M. J. 2003, *MNRAS*, **346**, 1103
- Pillepich, A., Nelson, D., Hernquist, L., et al. 2018, *MNRAS*, **475**, 648
- Popping, G., Somerville, R. S., & Trager, S. C. 2014, *MNRAS*, **442**, 2398
- Popping, G., van Kampen, E., Decarli, R., et al. 2016, *MNRAS*, **461**, 93
- Popping, G., Narayanan, D., Somerville, R. S., Faisst, A. L., & Krumholz, M. R. 2019a, *MNRAS*, **482**, 4906
- Popping, G., Pillepich, A., Somerville, R. S., et al. 2019b, *ApJ*, **882**, 137
- Prochaska, J. X., & Neeleman, M. 2018, *MNRAS*, **474**, 318
- Ragone-Figueroa, C., Granato, G. L., Parente, M., et al. 2024, *A&A*, **691**, A200
- Riechers, D. A., Boogaard, L. A., Decarli, R., et al. 2020a, *ApJ*, **896**, L21
- Riechers, D. A., Hodge, J. A., Pavesi, R., et al. 2020a, *ApJ*, **895**, 81
- Riechers, D. A., Pavesi, R., Sharon, C. E., et al. 2019, *ApJ*, **872**, 7
- Saintonge, A., & Catinella, B. 2022, *Annu. Rev. Astron. Astrophys.*, **60**, 319
- Saintonge, A., Catinella, B., Tacconi, L. J., et al. 2017, *ApJS*, **233**, 22
- Salpeter, E. E. 1955, *ApJ*, **121**, 161
- Schechter, P. 1976, *ApJ*, **203**, 297
- Scoville, N., Lee, N., Bout, P. V., et al. 2017, *ApJ*, **837**, 150
- Serra, P., Oosterloo, T., Morganti, R., et al. 2012, *MNRAS*, **422**, 1835
- Serra, P., Westmeier, T., Giese, N., et al. 2015, *MNRAS*, **448**, 1922
- Solomon, P. M., Downes, D., Radford, S. J. E., & Barrett, J. W. 1997, *ApJ*, **478**, 144
- Somerville, R. S., & Davé, R. 2015, *ARA&A*, **53**, 51
- Somerville, R. S., Primack, J. R., & Faber, S. M. 2001, *MNRAS*, **320**, 504
- Tacconi, L. J., Genzel, R., & Sternberg, A. 2020, *ARA&A*, **58**, 157
- Tumlinson, J., Peeples, M. S., & Werk, J. K. 2017, *Annu. Rev. Astron. Astrophys.*, **55**, 389
- Vallini, L., Pallottini, A., Ferrara, A., et al. 2018, *MNRAS*, **473**, 271
- Walter, F., Decarli, R., Aravena, M., et al. 2016, *ApJ*, **833**, 67
- Walter, F., Carilli, C., Neeleman, M., et al. 2020, *ApJ*, **902**, 111
- Westmeier, T., Kitaeff, S., Pallot, D., et al. 2021, *MNRAS*, **506**, 3962
- Yan, L., Sajina, A., Loiacono, F., et al. 2020, *ApJ*, **905**, 147
- Zanella, A., Daddi, E., Magdis, G., et al. 2018, *MNRAS*, **481**, 1976
- Zwaan, M., Ivison, R., Péroux, C., et al. 2022, *The Messenger*, **186**, 10
- Zwaan, M. A., Meyer, M. J., Staveley-Smith, L., & Webster, R. L. 2005, *MNRAS*, **359**, L30

Appendix A: Detections

We display the spectra and the moment map of the 87 detections found in the ALMACAL-22 survey, identified as CO emitters.

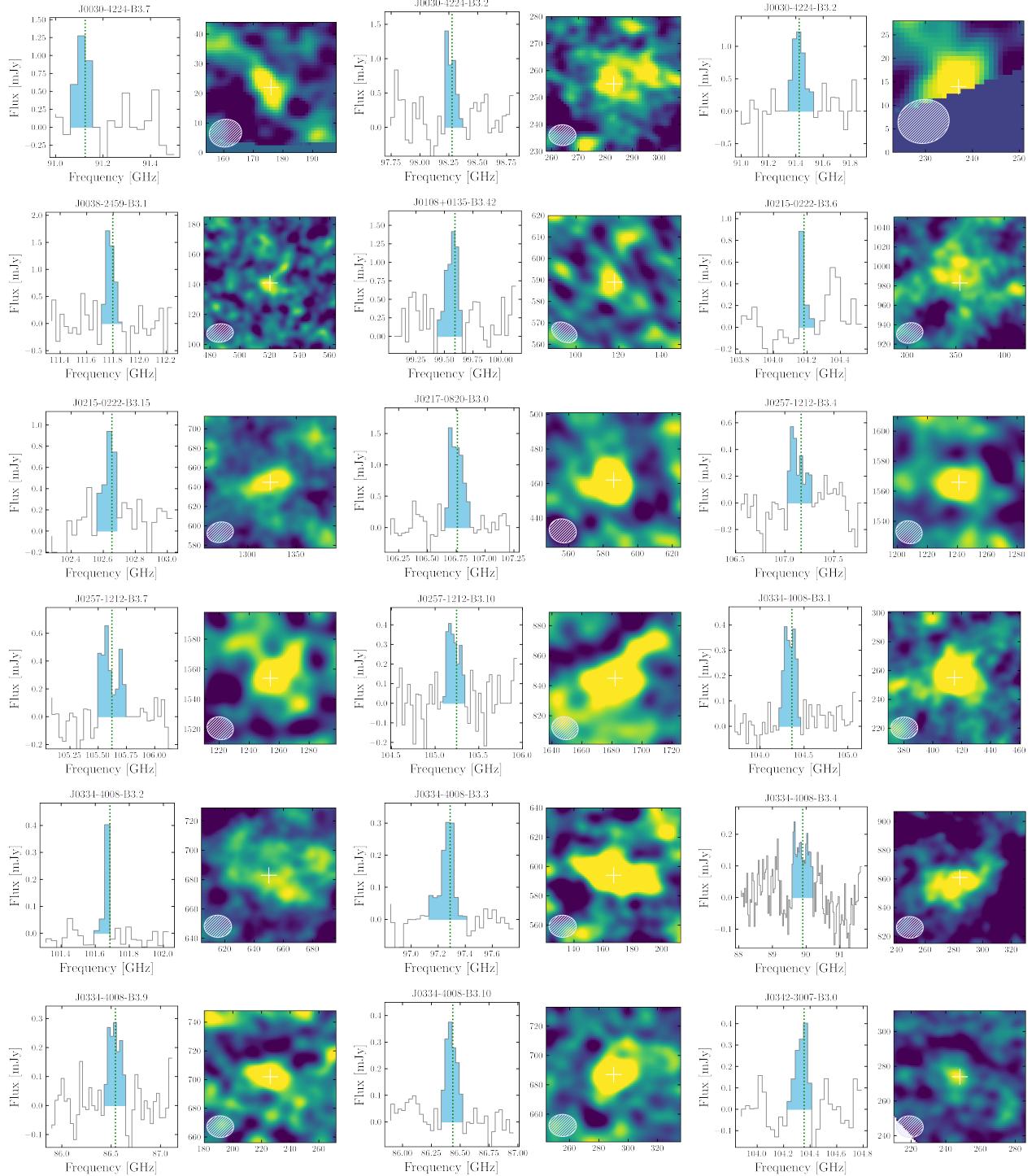


Fig. A.1. ALMACAL-22 emission lines detections. Left panel: Spectrum of the emission lines at the brightest pixel. Right panel: The emission line moment map centred on the detection range. The parameters of all candidates are summarised in Table A.1.

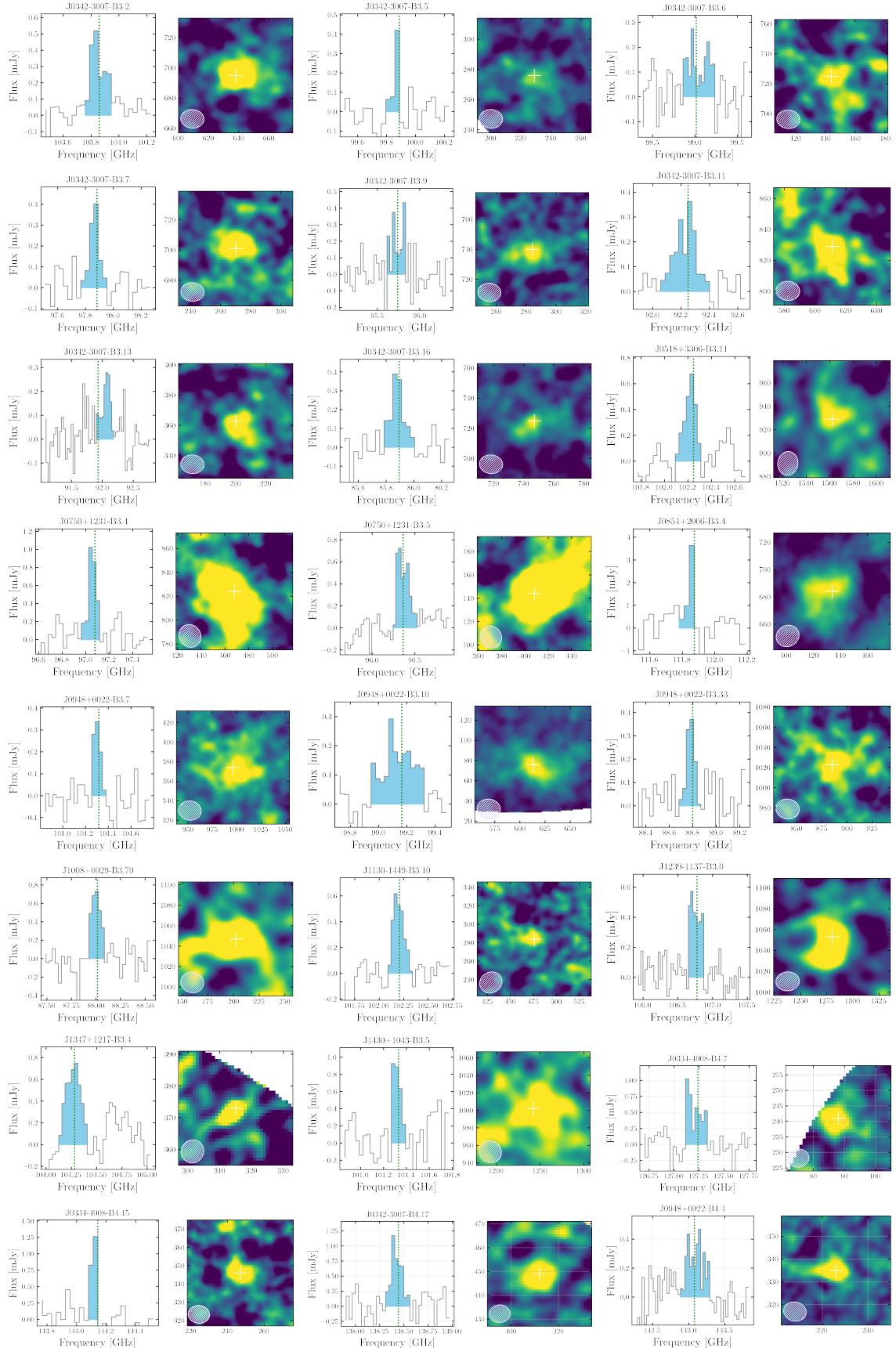
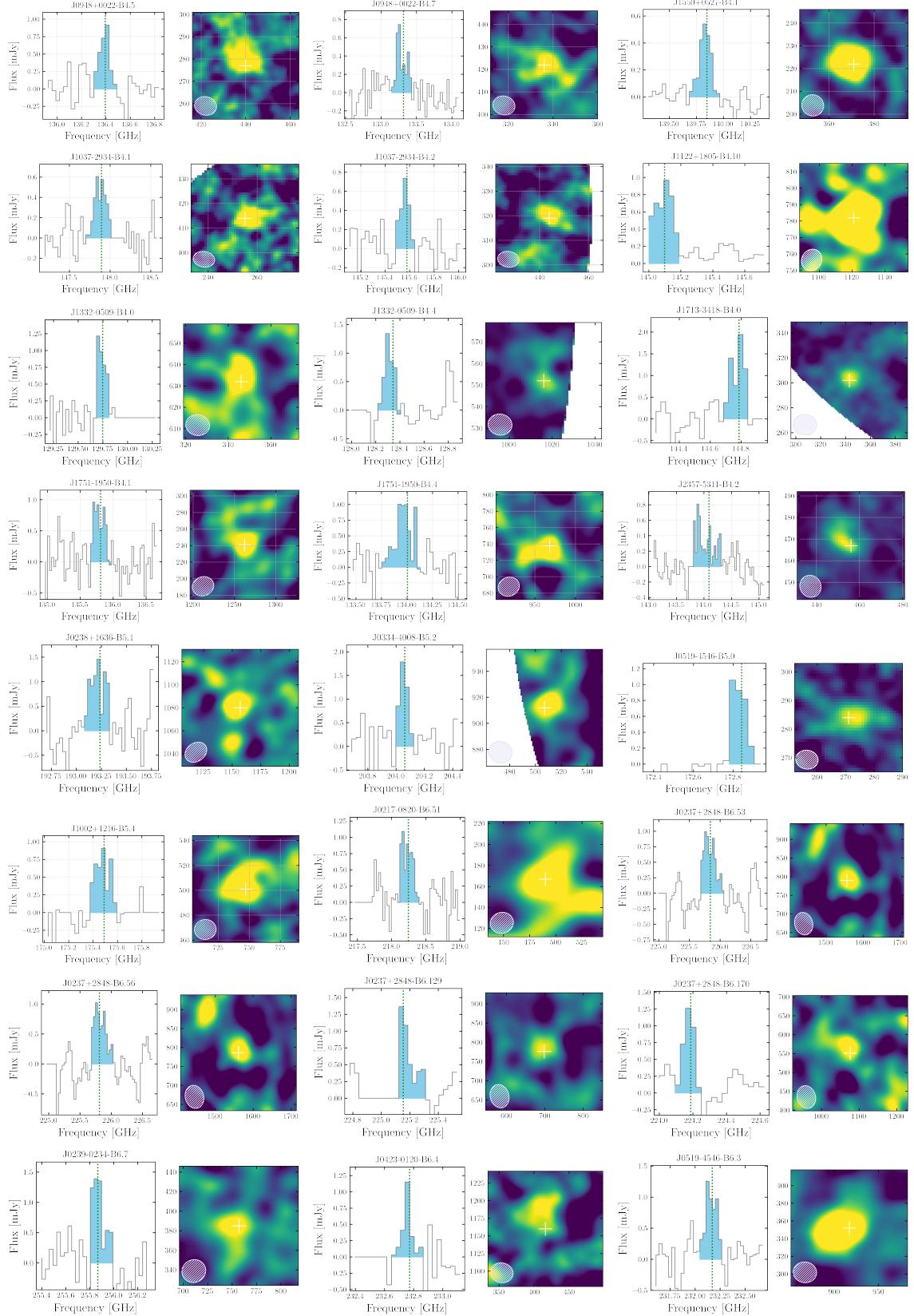


Fig. A.2. continued.

**Fig. A.3.** continued.

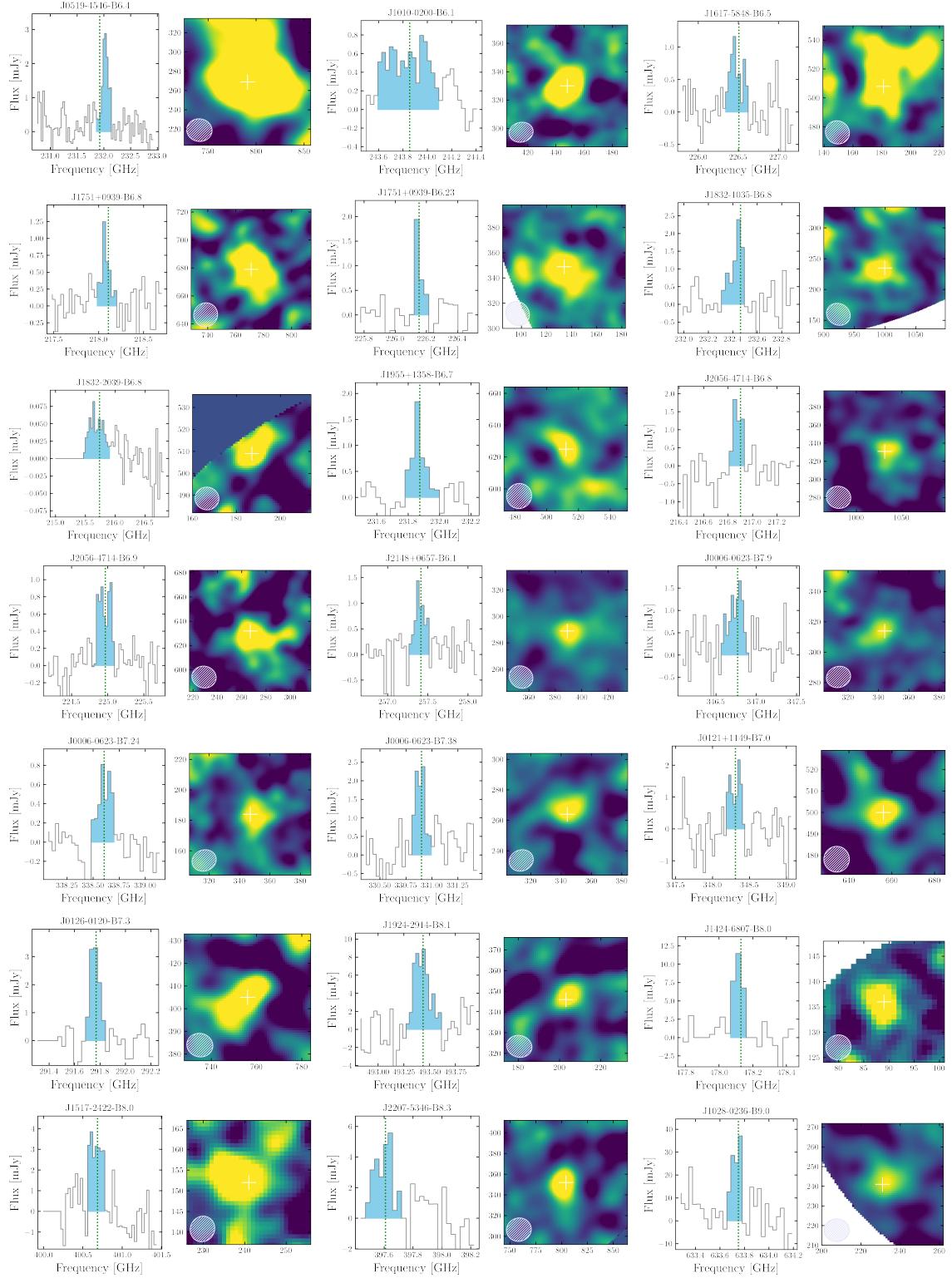


Fig. A.4. continued.