Quantifying the detection likelihood of faint peaks in interferometric data through jackknifing Test application on finding z > 10 galaxy candidates

(article starts on next page)

Astronomy & Astrophysics

Quantifying the detection likelihood of faint peaks in interferometric data through jackknifing

Test application on finding $z > 10$ galaxy candidates

Joshiwa van Marrewijk[1,2],⋆ , Melanie Kaasinen[1] , Gergö Popping[1] , Luca Di Mascolo[3,4] ,
Tony Mroczkowski[1] , Leindert Boogaard[2,5] , Francesco Valentino[1] , Tom Bakx[6] , and Ilsang Yoon[7]

[1] European Southern Observatory, Karl-Schwarzschild-Str. 2, 85748 Garching, Germany
[2] Leiden Observatory, Leiden University, PO Box 9513, 2300 RA Leiden, The Netherlands
[3] Laboratoire Lagrange, Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, Blvd de l'Observatoire, CS 34229, 06304 Nice cedex 4, France
[4] Kapteyn Astronomical Institute, University of Groningen, Landleven 12, 9747 AD Groningen, The Netherlands
[5] Max Planck Institute for Astronomy, Königstuhl 17, 69117 Heidelberg, Germany
[6] Department of Space, Earth, & Environment, Chalmers University of Technology, Chalmersplatsen 4, 412 96 Gothenburg, Sweden
[7] National Radio Astronomy Observatory, 520 Edgemont Road, Charlottesville, VA 22903, USA

ABSTRACT

Context. False-positive emission-line detections bias our understanding of astronomical sources; for example, falsely identifying $z \sim 3$–4 passive galaxies as $z > 10$ galaxies leads to incorrect number counts and flawed tests of cosmology.

Aims. In this work, we provide a novel but simple tool to better quantify the detection of faint lines in interferometric data sets and properly characterize the underlying noise distribution. We demonstrate the method on three sets of archival observations of $z > 10$ galaxy candidates, taken with the Atacama Large Millimeter/Submillimeter Array (ALMA).

Methods. By jackknifing the visibilities using our tool, jackknify, we create observation-specific noise realizations of the interferometric measurement set. We apply a line-finding algorithm to both the noise cubes and the real data and determine the likelihood that any given positive peak is a real signal by taking the ratio of the two sampled probability distributions.

Results. We show that the previously reported, tentative emission-line detections of these $z > 10$ galaxy candidates are consistent with noise. We further expand upon the technique and demonstrate how to properly incorporate prior information on the redshift of the candidate from auxiliary data, such as from the James Webb Space Telescope.

Conclusions. Our work highlights the need to achieve a significance of $\gtrsim 5\sigma$ to confirm an emission line when searching in broad 30 GHz bandwidths. Using our publicly available method enables the quantification of false detection likelihoods, which are crucial for accurately interpreting line detections.

Key words. methods: data analysis – techniques: interferometric – galaxies: high-redshift – galaxies: individual: S5-z17-1 – galaxies: individual: Glass-z12 – galaxies: individual: Glass-z10

1. Introduction

Assessing the reliability of faint astronomical signals is essential for all areas of astrophysics and cosmology. To understand the entirety of a population, the astronomical community is always attempting to characterize sources at the limits of what can be observed with the current generation of facilities. Interferometers such as the Atacama Large Millimeter/Submillimeter Array (ALMA) and the Very Large Array (VLA) have enabled the efficient study of faint, distant sources in the mm/cm-wave regime. However, images generated from interferometric data are highly complex; their underlying noise distribution is challenging to quantify, leading to potential biases at low S/N and difficulties in interpreting the data. This can have profound consequences, for example, when determining the existence, redshift, and physical properties of faint galaxies.

Correctly determining the likelihood of faint peaks in interferometric data has become relevant in the context of studies of the interstellar medium in the first galaxies ($z > 10$). Over the past two years, six ALMA director's discretionary time (DDT) programs[1] were approved and executed to follow up $z > 10$ galaxy candidates. These galaxies were initially identified from optical and near-infrared photometry, mostly including data from the JWST. However, modeling the spectral line energy distribution of these photometric data resulted in a wide redshift probability distribution, including several potential solutions. Capitalizing on the broad spectral coverage of ALMA, these DDT programs aimed to confirm the redshifts of the galaxy candidates by targeting the [O III] 88 μm line. But, obtaining robust line detections with an accurate S/N estimate has proven challenging.

⋆ Corresponding author; marrewijk@strw.leidenuniv.nl

---

[1] Additionally, one 80+ hour NOEMA observation on GN-z11 (which resulted in an upper limit; Fudamoto et al. 2024).

So far, ALMA DDT observations of $z > 10$ candidates have resulted in a handful of potential upper limits on the targeted line emission (Popping 2023; Kaasinen et al. 2023) and marginal, low S/N (3–4σ) detections of [O III] 88 µm (see, Harikane et al. 2022; Yoon et al. 2023; Bakx et al. 2023; Fujimoto et al. 2023), with the exception of one source, JADES-GS-z14, for which a 6σ detection of the [O III] 88 µm line was recently reported (Carniani et al. 2024a,b; Schouws et al. 2024). Some of these marginal detections have since been proven not to correspond to actual emission lines, with follow-up JWST spectra revealing at least two of these galaxies to be at a different redshift to that implied by the initial false positive [O III] 88 µm detection (Zavala et al. 2025; Harikane et al. 2025). This highlights a broader issue: false positive identifications of emission lines lead to biases in the derived physical properties. In the case of incorrect redshift solutions, this has profound implications for the number counts of the earliest galaxies, thereby impacting our understanding of both cosmology and galaxy evolution.

In this work, we present a straightforward and effective technique to quantify the detection level and underlying noise distribution in interferometric data sets. We build upon the approach used in Kaasinen et al. (2023); by differencing the visibilities, we generate various noise realizations of the observation-specific ALMA measurement set. Then, by applying line-finding algorithms on both the real and source-free image cubes, we sample the likelihood of a marginal detection being real without needing to assume an underlying noise distribution. We make our technique readily available to the community in the form of the public tool `jackknify`, along with clear, step-by-step tutorials[2]. Although we implement and optimize the tool for ALMA data, `jackknify` is compatible with any type of interferometric observations that use the Common Astronomy Software Applications for Radio Astronomy (CASA, CASA Team 2022). For example, this jackknifing technique may be useful for characterizing the faint HI emission from $z > 1$ galaxies using Squared Kilometer Array (SKA; Dewdney et al. 2009) and its precursors (see, e.g., Baker et al. 2024)[3], or quantifying the faint molecular line emission from quiescent galaxies with ALMA. In this work, we apply `jackknify` to three archival DDT observations targeting $z > 10$ galaxy candidates. We will test two scenarios: one in which no prior knowledge regarding the source's redshift is available and one in which we can use a prior from JWST on the frequency location of the line.

The paper is outlined as follows. Section 2 describes which interferometric data sets we use, Section 3 provides the overview of how `jackknify` is built, Section 4 describes the performance of the tool via simulated observations, Section 5 contains the results of applying `jackknify` to the ALMA observations of $z > 10$ galaxy candidates, and Section 6 provides an overview and summary of this work. For all calculations, the assumed cosmology is based on Planck Collaboration XVI (2014), a spatially flat, Λ cold dark matter (ΛCDM) model with $H_0 = 67.7 \ \mathrm{km \, s^{-1} \, Mpc^{-1}}$ and $\Omega_{\mathrm{m}} = 0.307$.

2. The archival ALMA observations

We tested and validated our jackknifing tool, `jackknify`, on three sets of ALMA DDT observations targeting $z > 10$ galaxy candidates pre-identified from photometry (as presented in Castellano et al. 2022; Naidu et al. 2022; Harikane et al. 2023).

All of these DDT observations (Bakx et al. 2023; Yoon et al. 2023; Fujimoto et al. 2023) aimed to detect the [O III] 88 µm emission line over a wide redshift probability range. To maximize the redshift coverage, each spectral setup consisted of four adjacent tunings, covering a large, ≈30 GHz bandwidth. A comprehensive overview of the literature results and observational specifics is presented in Table 1.

Calibrated measurement sets were provided by the European ALMA regional center network (Hatziminaoglou et al. 2015) through the calMS service (Petry et al. 2020) using CASA v5.4.0. To reduce the data size, we performed two additional operations on the calibrated data, applying a time averaging of 30 seconds and spectral averaging of the visibilities in ≈30 MHz bins. The frequency binning was conducted in a manner that avoided interpolation over the frequency axis. We used `uvcontsub` to manually continuum subtract a first-order polynomial from the measurement sets to remove any potential continuum emitters. We omitted the frequency channels where tentative features might be present in the data based on the previously reported detections. Finally, we imaged all measurement sets using the TCLEAN task with natural weighting.

3. Methodology

3.1. Jackknifing

In the narrow-field approximation, interferometers measure the 2D Fourier transform of the intensity distribution of an astronomical source over a sparse collection of Fourier modes. Such measurement modes, called visibilities, have units of wavenumbers and are expressed in terms of uv-distances, where u and v are the orthogonal vector bases of the Fourier space. Assuming proper calibration[4], noise in the Fourier domain, unlike in the image plane, is additive and Gaussian. This occurs because the individual pulses from electrons are so numerous that they blend into an indistinguishable, continuous Gaussian random process with a mean of zero (Thompson et al. 2017).

In this work, we use jackknifing to characterize the noise properties of the data. Jackknifing is a common technique used to retrieve the noise properties from bolometric measurements taken with millimeter-wave single-dish facilities. It is most commonly employed on cosmic microwave background (CMB)/single dish experiment observations (e.g., Weiß et al. 2009; Romero et al. 2018; Naess et al. 2020); however, it has also been used extensively to characterize radio/millimeter-wave interferometric data (e.g., Padin et al. 2001; Readhead et al. 2004; Sharp et al. 2010; Keating et al. 2015). Since interferometric data has uncorrelated, additive, and Gaussian noise with a mean of zero, we can retrieve the uncorrelated noise distribution, n, from the data, d, by jackknifing – by randomly multiplying half of the real and imaginary amplitudes of the visibilities by –1 and then rebinning, thereby averaging out the signal, s (see, for instance, Kitayama et al. 2020; González-López et al. 2020; Kaasinen et al. 2023; Di Mascolo et al. 2023, for an effective implementation).

---

[2] See https://joshiwavm.github.io/jackknify/

[3] However, we do note that our tools are validated only on data where the w-term was neglected.

[4] In the case of incorrect calibration, such as when one of the 44 online ALMA antennas has an erroneous gain solution, the noise in the visibility plane becomes multiplicative rather than additive to the signal. The effects of jackknifing in such a situation are not thoroughly tested. However, the assumption of well-calibrated data without antenna calibration errors is reasonable, as it is common practice to check the calibration accuracy before imaging to ensure there are no spurious signals that could result in imaging artifacts.

Table 1. Archival ALMA observations of $z > 10$ galaxy candidates.

ID	GLASS-z12	GLASS-z10	S5-z17-1
RA (deg)	3.4990	3.5119	339.0160
Dec (deg)	−30.3248	−30.3719	33.9046
ALMA source name	GLASS-z13 [GLASS-z12]	GLz11	S5-z17-1
PI	T. Bakx [J. Zavala]	I. Yoon	S. Fujimoto
ALMA Project Code	2021.A.00020.S [2023.A.00017.S]	2021.A.00023.S	2021.A.00031.S
ALMA Band	6 [6]	7	7
Beam ($''$)	$0\rlap{.}''32 \times 0\rlap{.}''29, 46°$ ($0\rlap{.}''45 \times 0\rlap{.}''40, 73°$)	$0\rlap{.}''77 \times 0\rlap{.}''59, 81°$	$0\rlap{.}''78 \times 0\rlap{.}''45, 3°$
Channel width (MHz)	31 [31]	31	31
On source time (h)	8.1 [1.6]	6.9	0.5
Reported tentative detections			
Ref. for initial identification	Castellano et al. (2022), Naidu et al. (2022)	Naidu et al. (2022)	Harikane et al. (2023)
Reference for ALMA data	Bakx et al. (2023), Zavala et al. (2024, 2025)[†]	Yoon et al. (2023)	Fujimoto et al. (2023)
z_{phot}	12.2 ± 0.2	10.4 ± 0.5	$16.7^{+1.9}_{-0.3}$
Δz covered	11.9–13.5	10.1–11.1	16.1–17.3
S/N ([O III] 88 μm)[*]	5.8, 3	4.4	5.1
Offset	$0\rlap{.}''5$, [$0\rlap{.}''0$]	$0\rlap{.}''17$	–
z_{spec}^{ALMA}	12.117 ± 0.001, [12.3327 ± 0.0035]	10.38	16.01
z_{spec}^{JWST}	12.33 ± 0.02	9.875 ± 0.008[**]	–
Freq (GHz)	258.7, [254.487 ± 0.019]	298.25	338.726
$\Delta \nu$ (km s^{-1})	400 ± 70, [186 ± 58]	225	118

Notes. Candidate coordinates and reported detection significance of the previously reported line searches. [†]The reported details after precise redshift information became available from the James Webb Space Telescope (JWST) observations and deeper ALMA Cycle 11 observations and shown within brackets. [*]In the case of S5-z17-1, we quote the redshift of the plausible [O III] 52 μm line. The beams are given in terms of major and minor beam axes and position angle. [**]The redshift solution for GLASS-z10 was reported by Napolitano et al. (2025).

When visibilities are imaged, the *uv*-data is automatically rebinned and averaged since all imaging and deconvolution tools, including `tclean`, grid the *uv*-coordinates before Fourier transforming. Any artifacts caused by the imaging, such as spatially correlated noise, are captured by performing the jackknifing in the visibility plane and performing the inference in the image plane. Jackknifing in the visibility plane is thus essential, rather than splitting in the image plane, to avoid removing any correlated noise induced by the Fourier transform present in the data.

To generate various noise realizations, we change the seeding of the randomization process. This allows us to sample the noise distribution, n, until convergence. Once n is adequately sampled, we use the distribution to infer the likelihood of false detection. Figure 1 shows schematically how this type of detection inference works. Given that the data, d, is a linear combination of the noise distribution, n, plus the signal, s, we can define the likelihood of detection $\mathcal{L}_D(\gamma)$ by integrating from an arbitrary S/N threshold, γ, onward, as indicated with the shaded area in Figure 1. By nullifying s through jackknifing of d, we recover n in the visibility plane and thus are able to compute the likelihood of a false positive detection $\mathcal{L}_{FD}(\gamma)$. The ratio of the two likelihoods provides the significance of detection.

Resampling the noise distributions requires that the cubes are re-imaged at every iteration. We ensure that `jackknify` interfaces with CASA and `tclean` to retain similar metadata for the spectral cubes, such that it is compatible with most line-finding algorithms. Because of this, `jackknify` can become computationally intensive, particularly for large data sets. As we will show in this paper, the analysis of ALMA spectral scans is computationally feasible. Large mosaics, however, might be more problematic. Therefore, we have created another implementation named `jaxknify`, which utilizes the `jax` implementation of
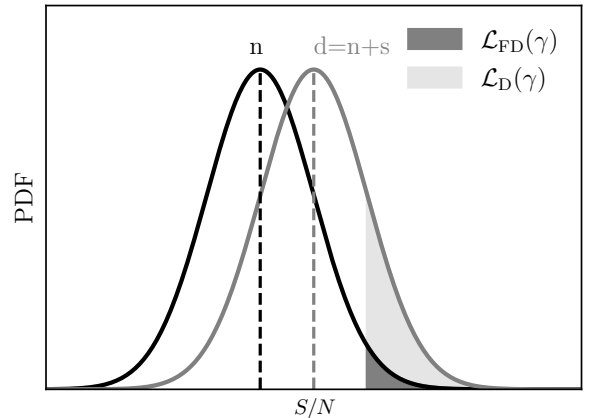

Fig. 1. Schematic of the detection inference. Given the data, d, which is the linear combination of the noise distribution, n, plus the signal, s, we can define the likelihood of detection $\mathcal{L}_D(\gamma)$ by setting an arbitrary S/N threshold γ and integrating the probability density function of the S/N for d from γ to ∞, as indicated with the shaded area. By nullifying s through jackknifing of d, we recover the ideal thermal noise, n, and thus compute the likelihood of a false positive detection $\mathcal{L}_{FD}(\gamma)$ as we do for computing $\mathcal{L}_D(\gamma)$. The ratio of the two likelihoods provides the significance of detection. This figure is inspired by Fig. 1 of Vio & Andreani (2016).

the iFFT as a substitute for `tclean`. The output of the Fourier Transform is, however, not compatible with existing line-find methods. Hence, for the remainder of this work, we used the CASA-compatible version of `jackknify`[5].

---
[5] `jaxknify` was only used to compute the many realizations that are visualized in Figure 5 as described in Section 4.

3.2. Line finding

3.2.1. Traditional empirical methods

To date, most approaches to finding emission-line sources in (sub)millimeter interferometric data are based on determining the significance of an emission line (a positive peak) via the line "fidelity" or "reliability" (e.g., Walter et al. 2016; Pavesi et al. 2018; González-López et al. 2019; Westmeier et al. 2021). The fidelity (or reliability) function is defined as

$$\text{fidelity}(\gamma) = 1 - \frac{N_{\text{neg}}(\gamma)}{N_{\text{pos}}(\gamma)}, \qquad (1)$$

where $N_{\text{neg}}(\gamma)$ and $N_{\text{pos}}(\gamma)$ are the number of negative and positive peaks, respectively, above a given detection threshold γ. The value of γ is usually a function of the S/N but can be arbitrarily chosen.

The fidelity function empirically estimates the likelihood ratio of the false detection, $\mathcal{L}_{\text{FD}}(\gamma)$, over the detection, $\mathcal{L}_{\text{D}}(\gamma)$ by taking the integral over the probability density functions (PDFs) of their respective probability distributions, $\mathcal{P}_{\text{FD}}(x)$ and $\mathcal{P}_{\text{D}}(x)$. $\mathcal{P}_{\text{FD}}(x)$ and $\mathcal{P}_{\text{D}}(x)$ are sampled by searching for negative and positive peaks, respectively. Therefore, the following relation of the likelihood ratio between detection and false detection holds:

$$\Lambda(\gamma) = \frac{N_{\text{pos}}(\gamma)}{N_{\text{neg}}(\gamma)} = \frac{\mathcal{L}_{\text{D}}(\gamma)}{\mathcal{L}_{\text{FD}}(\gamma)} = \frac{\int_\gamma^\infty \mathcal{P}_{\text{D}}(x|\Delta v)\, dx}{\int_\gamma^\infty \mathcal{P}_{\text{FD}}(x|\Delta v)\, dx}. \qquad (2)$$

This relationship is a function of the kernel width used to determine the linewidth, Δv. For example, to characterize the line emission in the ALMA Spectroscopic Survey of the Hubble Ultra Deep Field (ASPECS) González-López et al. (2019) chose the detection threshold γ such that fidelity$(\gamma) = 0.9$. This resulted in $\gamma \simeq 6.4\sigma$ for the ASPECS sample, assuming that Eq. (1) is well described by an error function (see also, Decarli et al. 2020).

The other assumption made while employing the fidelity function is that the flux distribution of negative peak values is a proper estimator for $\mathcal{P}_{\text{FD}}(x)$. However, this does not hold when absorption with respect to the CMB or a bright background source leads to negative line fluxes, contaminating the noise statistics. Additionally, since the method is applied in the image domain, any bright line or continuum flux will also create a negative signal through the dirty beam patterns. Bright sources can be "cleaned" (if they fall within the imaged part of the sky), but the fainter ones contribute, albeit marginally, to spatially correlated noise (Tsukui et al. 2023). Furthermore, ALMA has a limited data volume along the spectral axis; hence, $\mathcal{P}_{\text{FD}}(x)$ is not fully sampled over the δv space, making these empirical approaches inaccurate.

3.2.2. Line detection inference through jackknifing

Here, we describe how we perform the line detection throughout this work by utilizing the jackknifed data sets. To determine the likelihood of detection, we first quantify the likelihood of false detection, for which we need the underlying PDF. The distribution of false positive detections is set by the distribution of peak values in a data cube that only contains noise (Section 3.2.1). The distribution of positive or negative peaks is a subset of the pixel-value distribution in the data cube (Vio & Andreani 2016). The number, location, and amplitude of peak values depend on the realization of the total pixel-value distribution, which for interferometric data in the image plane, can be approximated as a
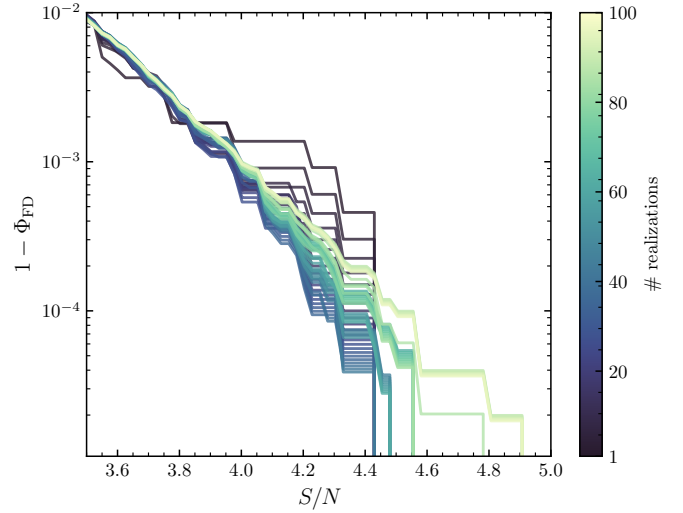

Fig. 2. How an increasing number of jackknife realizations leads to a smoother $\Phi(x)$ that better samples the high-S/N of the probability of false detection, $\mathcal{P}_{\text{FD}}(x)$. The probability functions are shown as one minus the mean cumulative distributions $\Phi(x)$ of $\mathcal{P}_{\text{FD}}(x)$ as a function of the peak S/N for various amounts of jackknife realizations.

smoothed random Gaussian field. By jackknifing the visibilities, we effectively shuffle the noise distribution, altering the noise realization and creating another subsample of the pixel-value distribution, namely another peak distribution. This allows us to increase the sample statistics of $\mathcal{P}_{\text{FD}}(x)$ while using the same data set.

Figure 2 shows that for an increasing number of realizations, we better probe the tail of the distribution at high S/N. The cumulative distribution function (CDF) Φ_{FD} of $\mathcal{P}_{\text{FD}}(x)$ is smooth till $S/N \approx 4.5$ and reaches a $S/N = 5$ within a hundred realizations. Specifically, each CDF incrementally includes data from additional jackknife realizations. For example, the CDF for 35× jackknife realizations contains all the data from the 34× realization set. Furthermore, the jackknifed data only includes peaks within a $0.''6$ aperture of the center of the data cube (chosen so that this test is consistent with the discussion in Section 5). Figure 2 extends to a hundred realizations in total. Since there is a trade-off between sampling more realizations and computation time and from ≈ 50 realizations onward we recover the high-S/N tail of the distribution; we decided to use 50 jackknife realizations throughout this work.

The difference between the sampled PDF of the false positives, $\mathcal{P}_{\text{FD}}(x)$, coming from the negative peak distribution (e.g., one noise realization) in the real data versus that from 50 jackknife realizations is shown in Figure 3. For the jackknifed realizations, we show the inner 95% confidence interval per bin. Again, for both the real and jackknifed data, we only consider peaks within a $0.''6$ aperture of the center of the data cube. Since we do not expect a negative signal in this data set, the two distributions should be equivalent. However, our method provides better sampling (as also shown in Figure 2) and thus better statistics, particularly at higher S/N. Furthermore, the scatter in the jackknife realizations within the 95% confidence interval is consistent with the Poisson uncertainty in the real data for each bin, except for a potential outlier at $S/N \approx 4$. This indicates that the various realizations are consistent with each other.

As shown in Figures 2 and 3, jackknifing allows for a more complete sampling of the noise distribution without the need for complex models or computationally intensive four-dimensional
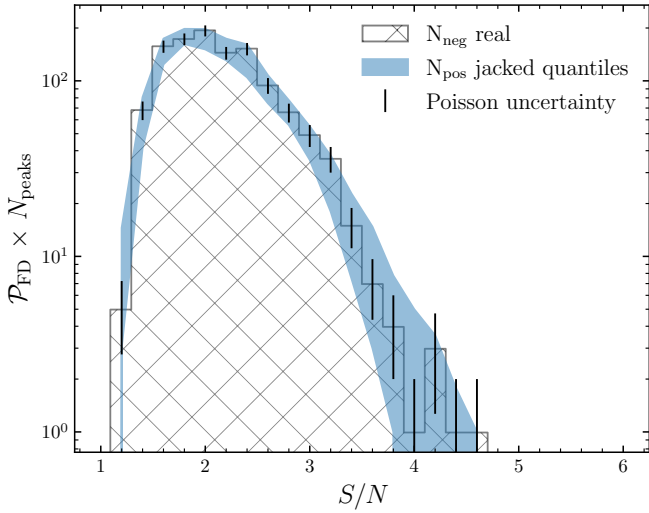
Fig. 3. Sampled noise distributions, $\mathcal{P}_{\mathrm{FD}}(x)$, as a function of the peak S/N. The gray hatched histogram shows the results from sampling the number of negative peak values of the original data set (shown with corresponding Poisson uncertainty). The blue-filled region represents the 95% confidence interval of the positive peak values from the jack-knife observations.

covariance estimates (see, Vio & Andreani 2016; Tsukui et al. 2023, for a discussion on the autocorrelation function implementations used for linefinding). By leveraging `jackknify`, we can obtain a more reliable and efficient measure of the probability of false detection, $\mathcal{P}_{\mathrm{FD}}(x)$, and take the likelihood ratio, $\Lambda(\gamma)$ with $\mathcal{L}_{\mathrm{D}}(\gamma)$ according to Eq. (2) to estimate a feature being real or not. Then,

- if $\Lambda(\gamma) > 1$, the observed data is more likely to fall under the hypothesis that a signal is present, whereas
- if $\Lambda(\gamma) \leq 1$, the observed data falls under the null hypothesis and is thus better described as being drawn from the underlying noise distribution.

Finally, an excess in the ratio is considered significant if the likelihood ratio exceeds a critical value k, such that:

- if $\Lambda(\gamma) \geq k$, one can reject the null hypothesis that no signal is present in the data, whereas
- if $\Lambda(\gamma) < k$, one cannot reject the null hypothesis.

4. Application to simulated ALMA data

To validate our method, we simulated mock ALMA data using CASA's `simobserve` task. We simulated a one-hour-long execution block centered at an RA and Dec of zero degrees. The simulated data contained six channels centered at a frequency of $\nu = 279$ GHz with a channel width of 31 MHz. We created the simulated uv-coverage using configuration 4 from cycle 7, leading to a resolution of $0\rlap.{''}6$ when `cleaning` using natural weighting. The setup of the simulations was designed to be similar to the observational setup of GLASS-z12 (Table 1). Using this setup, we simulated several datasets containing continuum and line emission from two separate sources with various integrated S/N values. We incremented the flux of the sources, maintaining the same uv-coverage for every observation. The line flux was simulated using a 3D Gaussian model. One dimension corresponds to the spectral flux distribution; its flux is constrained as described above – so that the integral over the Gaussian equals the respective S/N – and has a full-width half-maximum (FWHM) equal to ~73 MHz, corresponding to a line

with an FWHM of 80 km s^{-1}. The second and third dimensions of the Gaussian correspond to the spatial dimensions and were used to generate a blob in each channel map, with its amplitude set by the first Gaussian and an FWHM of 2.355 pixels (1 pixel = $0\rlap.{''}01$). Hence, the observation is considered unresolved. The same angular size on the sky is used for the continuum source.

Using the simulated data, we checked how well `jackknify` removes the signal. Figure 4 shows the moment-0 maps of the $S/N = 5$ observation (left), the corresponding noise map (Middle), and the map resulting from jackknifing the visibilities (right). We created the noise-only map by subtracting the noise-free dirty map (which is an output of `simobserve`) from the noisy output. Thus, the noise-only map is representative of the noise in a "perfectly `cleaned`" map. The simulated noise map and jackknifed realization exhibit the same noise structure, validating our jackknifing approach.

We tested how well jackknifing performs as a function of the sources' S/N, as the randomization process might not fully remove the signal, leaving residuals in the generated noise realization. We initially noticed a difference between the standard deviation from the single cleaned map and the mean of the 50 jackknife realizations (see Appendix A). This bias depended on the seeding used in `simobserve`, which is the result of the inherent variance of `simobserve`, combined with the fact that the standard deviation is estimated from maps with a finite size. Due to the limited representation of the chosen image domain, large- and intermediate-scale modes can induce non-negligible biases in the measured noise variance compared to the underlying truth distribution. To explicitly test for this, we create jackknifed realizations over a large set of `simobserve` runs without a source, each assuming a different input random seed (Fig. 5). All the resulting marginalized jackknife distributions are found to be consistent with the same normal distribution, which is characteristic of the noise probability function. The `simobserve` variance estimates are scattered according to the same distribution, implying that the `simobserve` output itself shows the same statistical properties of the result of an individual jackknifing cycle. Therefore, jackknifing can be used to better describe the underlying distribution function independent of the noise realization used in `simobserve`.

We tested how the measured standard deviations from the jackknife realizations evolve with S/N in Figure 6. To this end, we simulated the sources as Gaussians that are off-centered from the phase reference point. This setup allows the signal to leak from the real to the imaginary components, testing the effectiveness of the jackknife routines without assuming the source to be at the phase reference – an assumption often made in interferometric data analysis tools (such as `uvcontsub`). We compare the mean and standard error of the standard deviation values measured for the jackknifed realizations versus the observations simulated with `simobserve`.

We find that our jackknifing approach is perfectly reliable for $S/N \lesssim 50$. That is, the jackknife realizations of simulated observations with sources at $S/N \lesssim 50$ are perfectly consistent with the cleaned noise maps. The observations simulated with a peak $S/N \gtrsim 50$, however, indicate a stronger deviation from the cleaned noise estimate. We note that this does not manifest as a residual of the source but as an overall larger standard deviation throughout the map. We therefore advise that before applying `jackknify` to the measurement set, one should remove extremely bright continuum and line-emitting sources (e.g., $S/N > 50$) from the data set in the visibility plane, as is common practice with other line-finding methods (e.g., Walter et al. 2016). Overall though, the tests we have performed with
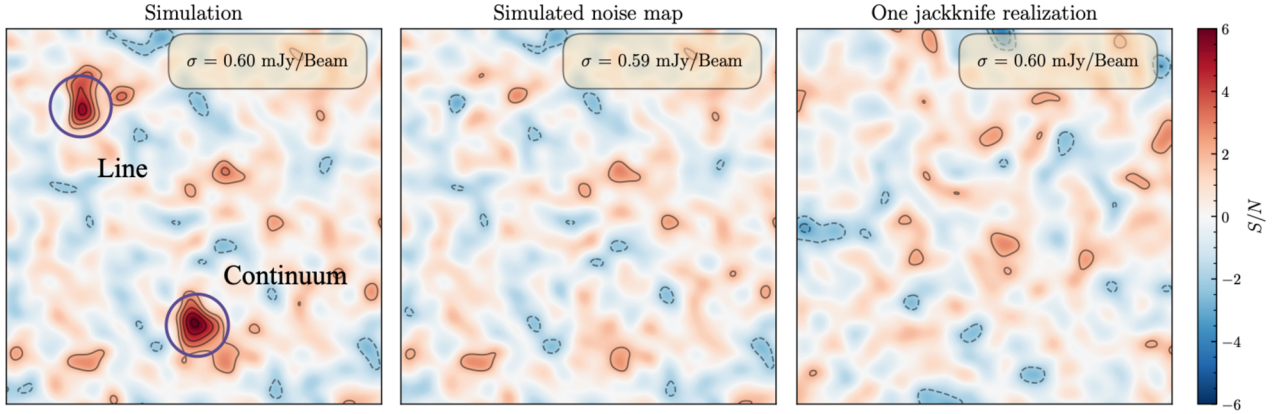
Fig. 4. Moment-0 maps of simulated ALMA data, generated using CASA `simobserve`. Left: output of the simulated observation, which contains both a continuum and a line-emitting source with $S/N = 5$. Middle: the corresponding noise map, referred to as the "cleaned" map, obtained by subtracting the noiseless output (i.e., filtered sky model) of `simobserve` from the noisy one. Right: the moment-0 map from a single jackknife realization. All contours are drawn at $[\pm 2, \pm 3, \pm 4, \pm 5, \pm 6]\,\sigma$.
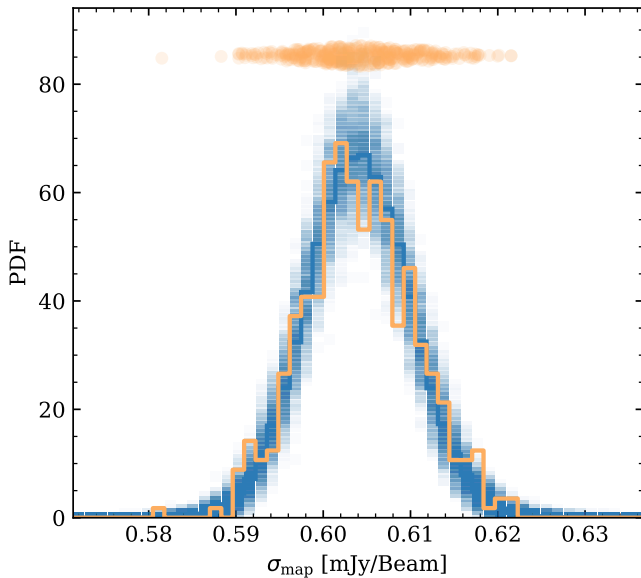

Fig. 5. Comparison of the standard deviation σ_{map} measured from different `simobserve` realizations using 200 different random seeds (orange points and corresponding histogram) and the σ_{map} of the jackknifing for each corresponding mock observation (shaded blue squares). The thick blue line traces the total distribution of the derived standard deviations, σ_{map} when averaging over all the jackknife realizations for all the different `simobserve` seeds. Through jackknifing, we thus recover the true noise distribution of the observations.

Fig. 6. Standard deviation for all maps as a function of the peak S/N of the sources. The blue-shaded region indicates the standard deviation over the various noise estimates from the 50 jackknife realizations for each S/N. The orange-shaded region is the standard deviation over the orange data points in Fig. 5. By jackknifing the observations, we retrieve the noise properties of the cleaned map accurately when the sources have $S/N < 50$.

simulated data (Figures 5 and 6), show that jackknifing allows for an accurate estimation of the underlying noise distribution[6].

The final validation we performed with the simulated data was to test the accuracy of the line-finding method. Since jackknifed data cubes have an output identical to the original dataset, we apply the same tools to the real and noise realizations. Therefore, we use the simulation for which the sources have an integrated $S/N = 5$. We sample the distribution of false positive detections, $\mathcal{P}_{\mathrm{FD}}(x)$, by applying the line finding algorithm to the jackknifed data and use the `FindClump` algorithm[7] to search for lines in the jackknifed data cubes (see Section 5.1 for a description on how `FindClump` is setup). The average distribution of the peak values in the 50 jackknife realizations is shown by the blue histogram in Figure 7. This histogram reveals a smooth peak distribution, declining from $S/N = 2.5$ to $S/N \approx 5.5$[8]. Given the number of peaks `FindClump` detected in a single realization ($N_{\mathrm{peaks}} = 70$), and a likelihood of $\mathcal{L}_{\mathrm{FD}}(\gamma = 4.5) = 0.0021$, we expect to find on average 0.1 ± 0.4 peaks above $S/N = 4.5$ in the simulated observation. By applying the line-finding algorithm to

---

[6] We note that we did not evaluate the performance of jackknifing the visibilities in cases of large extended bright emission such as in molecular clouds within low-redshift galaxies.
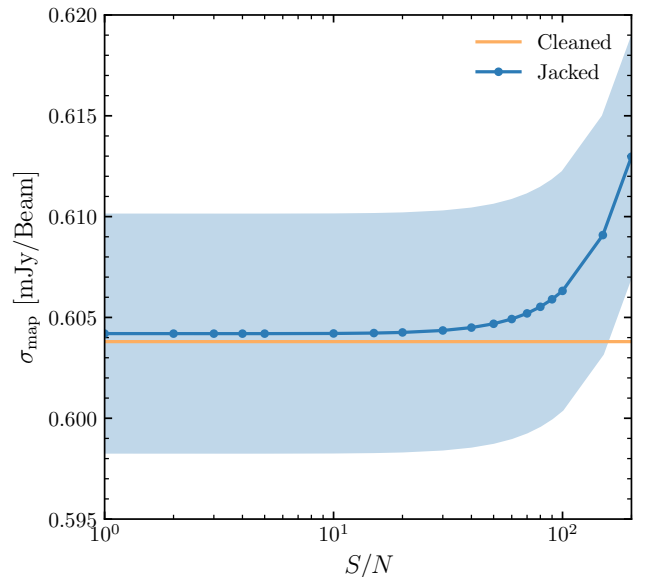
[7] We note that this is an arbitrary choice and other algorithms may be better suited.

[8] We note that the drop-off at lower S/N is due to the cropping of peaks and the way `Source Extractor` catalogs its findings.
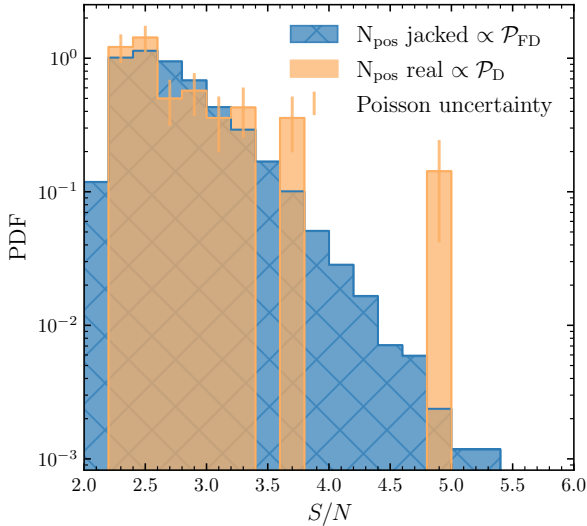
Fig. 7. Output of the line finding done on the simulated ALMA observation that had an off-centered line and continuum source, which was initialized to have an integrated S/N of 5 in the integrated velocity map (see Fig. 4). In blue, we show the underlying PDF of a false detection $\mathcal{P}_{FD}(x)$ derived from counting peak values in the jackknifed realizations. In orange, we show the results on the simulated data, which shows a clear detection of the two sources at a $S/N = 5$. The peak at $S/N \approx 3.6$ is consistent with being noise (as explained in the main body of the text).

the real data (orange histogram), we recover the simulated continuum and lines sources, both with a $S/N = 4.9$. Compared to the underlying noise distribution (in blue), the likelihood ratio tests yield $\Lambda(\gamma = 4.0) = 13$. We also find a small excess at $S/N \approx 3.6$ and check if our likelihood ratio test indicates this to be real or not (we know it is not). We find that the number of peaks, expected from jackknifing, at the S/N range of $3 < \gamma < 4$ ($3.5 < \gamma < 4$) is 14.6 ± 3.8 (2.1 ± 1.5), while we find 16 ± 4 (5.0 ± 2.2) in the simulated data, resulting in a $\Lambda(3 < \gamma < 4) = 1.1$ ($\Lambda(3.5 < \gamma < 4) = 2.4$). Thus, our statistical tests indeed find these peaks to be consistent with being noise, meaning we recover the simulated sources while the noise is correctly characterized as such.

5. Application to high-z galaxy candidates

Our tool, `jackknify`, can be applied to various science cases, including ones involving blind line searches in deep fields (i.e., without known counterparts), targeted searches around candidates identified at other wavelengths (e.g., following up $z > 10$ candidates) and searches for companions around known sources. Here, we apply our method to three sets of ALMA observations targeting the $z > 10$ galaxy candidates GLASS-z12, GLASS-z10, and S5-z17-1 (Bakx et al. 2023, Yoon et al. 2023, and Fujimoto et al. 2023, respectively). There was also a DDT program targeting the $z \sim 13$ candidate HD1, but we exclude these observations here as more recent work by Harikane et al. (2025) spectroscopically confirmed the object to be a passive galaxy at $z = 4.0$ and a similar jackknifing analysis was already performed in Kaasinen et al. (2023). Furthermore, we note that during the referee process, two additional ALMA observations targeting $z > 10$ galaxies were reported. Specifically, these include observations on GL-z14 (Project Code: 2023.A.00037.S, PI: S. Schouws; Carniani et al. 2024a; Schouws et al. 2024) and a deeper observation targeting GLASS-z12 (Project Code: 2023.A.00017.S, PI:

J. Zavala; Zavala et al. 2024). These observations are not included in this work but underscore how rapidly the field is advancing.

In Section 5.1, we describe the experimental setups, in Section 5.2, we detail the results of the blind line search across the entire spectral axis and, in Section 5.3, we summarize the line finding results for GLASS-z12, incorporating an additional redshift prior based on spectroscopic JWST/Mid-Infrared Instrument (MIRI) measurements taken with the Medium Resolution Spectrograph (MRS) of the Hα line (Zavala et al. 2025).

5.1. Specific experiment setup

To quantify the detection significance of a line along the full spectral axis, we adopt the following procedure.

1. We jackknife each observation set 50 times, creating 50 different noise realizations. Each jackknifing iteration includes differing visibilities and imaging of the full spectral scan and takes ≈ 9 minutes on a 2019 MacBook Pro with a 2.6 GHz 6-Core Intel Core i7 processor. A tenfold improvement in performance speed was found when using a MacBook Pro with the M3 chip.
2. We use `FindClump` (Walter et al. 2016), as implemented in the `interferopy` package (Boogaard et al. 2021), to sample the distribution of positive peak values from both the real and jackknifed data.
3. We mask out features outside a 0.″6 radius of the optically derived center. This corresponds to a physical radius of ≈ 2 kpc at $z = 12$ and is $\approx 2\times$ the size of the beam for all four datasets. We chose a radius of $r = 0.″6$ to include the tentative detection previously reported in our analysis (the feature was found at an offset of 0.″51 from the JWST-derived centroid; Bakx et al. 2023), but be able to exclude serendipitous detections of interlopers.
4. We calculate the likelihood ratio of the respective distributions to infer if there is an excess in the peak distribution in the real data compared to the sampled noise distribution.
5. We consider a detection to be significant if the excess of peaks in the real data with respect to the noise has at least a significance of 2σ based on Poisson statistics. Therefore, we require $\Lambda(\gamma)k$, with $k = 3$.

In step 2, we configure `FindClump` to search for emission lines with widths between 100 and 500 km s^{-1}, corresponding to 3–11 times the channel width. The velocity width varies slightly across different observation sets as we are examining galaxies at various redshifts. With `FindClump`, we identify peaks that exceed a signal-to-noise ratio (S/N) of 0, and we crop "identical" peaks if they are found within 0.2″ of each other spatially (approximately 1/3 of the beam width) and within 0.2 GHz spectrally (roughly 4–5 channels apart). Additionally, we have modified the default settings in `SExtractor` (Bertin & Arnouts 1996), specifically the analysis and detection thresholds, which we both set to 2 to better detect low S/N clumps.

5.2. Results on blind line searches

We check for any real signal by comparing the distribution of positive peaks in the real and jackknifed data (Fig. 8). According to Eq. (2), this comparison translates to the ratio of the probability distribution of a detection, $\mathcal{P}_D(x)$, versus a false detection, $\mathcal{P}_{FD}(x)$. Thus, any excess of positive features in the real data relative to the noise distribution indicates a true positive line detection. Table 2 provides an overview of the likelihood of previously reported tentative detections being real. In the following section, we will discuss the results for each source, one by one.
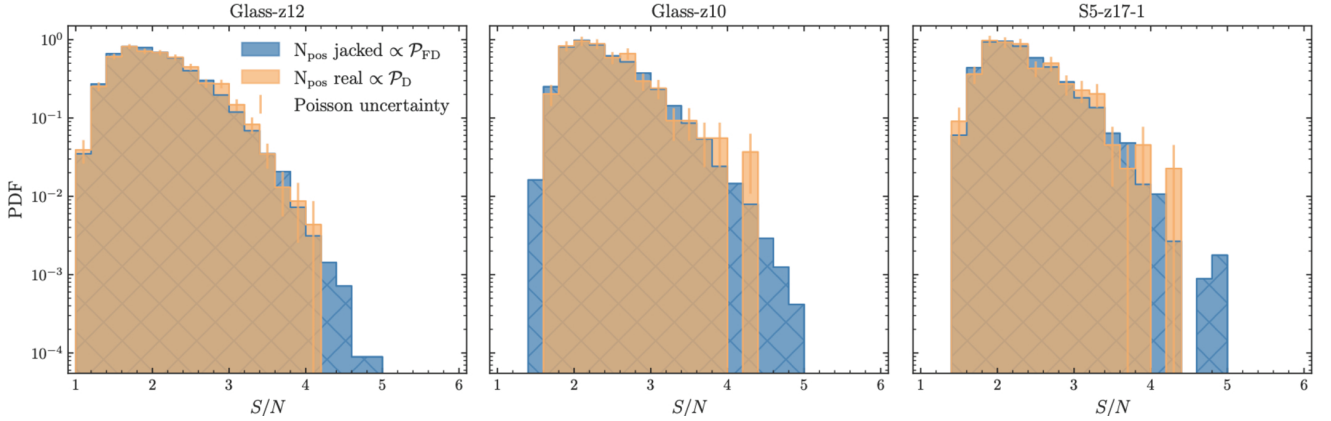
Fig. 8. Probability distribution of false detections, $\mathcal{P}_{FD}(x)$ noted as N_{pos} jacked, and the peak value distribution of the observations, $\mathcal{P}_D(x)$ noted as N_{pos} real. Both distributions are obtained by running `FindClump` on the jackknifed cubes and the real data cube, respectively. From left to right, we show the three data sets of GLASS-z12, GLASS-z10, and S5-z17-1 (see Table 1). An excess of $\mathcal{P}_D(x)$ over $\mathcal{P}_{FD}(x)$ is indicative of line detection. However, no such excess is detected.

Table 2. Significances of detected peaks.

	GLASS-z12	GLASS-z10	S5-z17-1
Reported significance	5.8σ	4.4σ	5.1σ
Found highest significance[†]	4.2σ	4.3σ	4.3σ
$\mathcal{L}_{FD}(\gamma = 4)$	0.0011	0.0054	0.0032
$\mathcal{L}_{FD}(\gamma = 4) \times N_{peaks}$	1.3 ± 1.1	1.5 ± 1.2	0.7 ± 0.8
N_{peaks} found	1	2	1
$\Lambda(\gamma = 4)$	0.80	1.4	1.4

Notes. In this table, we quote the individual likelihoods of false detections for the three DDT-ALMA data sets on $z > 10$ galaxies. Multiplied with the number of peaks in the real dataset (N_{peaks}), this leads to the expected number of peaks above $S/N > 4$ in a purely noisy dataset. We compare this with the found number of peaks in the real data and the likelihood ratio test as defined in Eq. (2).[†] We note that the peaks with the highest significance are not the same peaks as the ones that correspond to the reported significance (top row). See Section 5.2 for more detail.

For GLASS-z12 (left panel, Fig. 8), we find no excess of positive peaks in the real data compared to the underlying noise distribution. We integrated the probability of a false detection from the detection threshold of $\gamma = 4\sigma$ onward. Considering the total number of peaks in the real data ($N_{peaks} = 1150$), we estimated the expected number of peaks due to noise fluctuations. We found $\mathcal{L}_{FD}(\gamma = 4\sigma) \times N_{peaks} = 0.0011 \times 1150 = 1.3 \pm 1.1$. In the real data, we recover one peak above $S/N > 4$, which is consistent with being noise. This is also indicated by the likelihood ratio test (Eq. (2)), which results in $\Lambda(4\sigma) = \mathcal{L}_D(4\sigma)/\mathcal{L}_{FD}(4\sigma) = 0.80$. We note that the $S/N = 4.2$ peak discussed here is not the same peak found by Bakx et al. (2023). The peak we have recovered is at a frequency of $\nu = 246.6106$ GHz and coordinates RA, Dec = (3.49895, −30.32469), which differs from their reported 5.8σ peak over 400 km s^{-1}, offset 0″.5 from the JWST position (Table 1). We recover a peak at the same location but at a lower significance of 2.9σ for a linewidth of 280 km s^{-1} (the maximum S/N peak over any linewidth at this position). The difference in S/N and linewidth is due to the difference in how the data were imaged. Whereas we imaged these data using natural weighting and using a channel width of 46 km s^{-1}, Bakx et al. (2023) tapered the data to 0″.3 and used a channel width of 150 km s^{-1}. This highlights the importance of imaging the jackknifed data in the same way as the real data.

For GLASS-z10 (middle panel, Fig. 8), we find a small excess of peaks in the real versus the jackknifed (noise-only) data at $S/N = 4.1$–4.3. Given the noise distribution and the total number of peaks in the real data ($N_{peaks} = 271$), we expect to find 1.5 ± 1.2 peaks at $\gamma > 4\sigma$. In the real data, we find two peaks at $\gamma > 4\sigma$. The likelihood ratio resulted in $\Lambda(\gamma = 4\sigma) = 1.4$. This does not exceed our detection threshold of $k = 3$, meaning that the two tentative features[9] are consistent with being noise at a likelihood of $\mathcal{L}_{FD}(\gamma = 4\sigma) = 0.0054$. Neither peak was previously reported by Yoon et al. (2023). We also recover the previously reported detection at 2.7σ with a linewidth of 235 km s^{-1}. This line has a false detection likelihood of $\mathcal{L}_{FD}(\gamma = 2.7\sigma) = 0.19$ and a likelihood ratio test resulting in $\Lambda(\gamma = 2.7\sigma) = 0.9$[10].

For S5-z17-1 (right panel, Fig. 8), the expected number of peaks is $\mathcal{L}_{FD}(\gamma = 4\sigma) \times N_{peaks} = 0.0032 \times 221 = 0.7 \pm 0.8$. Given that we find one peak above $S/N > 4$ and since $\Lambda(\gamma = 4\sigma) = 1.4$, we conclude that these data are also consistent with being noise. We recover the peak reported in Fujimoto et al. (2023), albeit with a significance of 3.9σ instead of the reported 5.1σ. This line has a false detection likelihood of $\mathcal{L}_{FD}(\gamma = 3.9\sigma) = 0.0060$ and a likelihood ratio test resulting in $\Lambda(\gamma = 3.9\sigma) = 2.2$.

In conclusion, our analysis of the GLASS-z12, GLASS-z10, and S5-z17-1 datasets reveals no significant detections of any of these (candidate) $z > 10$ galaxies in the ALMA data. All of the peaks identified in the real data are consistent with noise, as indicated by the likelihood ratio tests and comparison with the jackknifed noise distributions. Given the number of peaks per data set, the detection significance does not follow a Gaussian distribution. Since a 4σ detection should reflect that only one in fifteen thousand random draws should be a false detection, we reach a 5σ fluke already within a sample size of $\approx$200–1000 peaks. This highlights the importance of rigorous noise analysis in confirming potential astronomical signals.

---

[9] Both features have a $S/N = 4.3$, a spatial offset of 0″.35 and 0″.39 from the optical counterpart, and are found at $\nu = 291.58$ and $\nu = 296.12$ GHz. Both have a linewidth of 170 km s^{-1}.

[10] Our findings on a non-detection of GLASS-z10 in the submillimeter regime are in line with that of Napolitano et al. (2025). They reported a spectroscopic redshift of GLASS-z10 of $z = 9.875$ through JWST observations targeting the Lyman-break, which falls outside of the redshift space ALMA probed.

5.3. Quantifying detection significance with known redshifts

So far, we have presented the jackknifing plus line-finding approach for the case of a blind detection experiment and the case where the approximate position of the source is known but where we have no strong prior on the central frequency. We now extend the analysis for cases where the redshift and, hence, the line frequency are known from other observations. To provide an example, Zavala et al. (2025) detected bright nebular emission lines of GLASS-z12 with JWST, constraining the redshift to be $z_{spec} = 12.33 \pm 0.02$. With this prior knowledge, we will reanalyze the data reported in Bakx et al. (2023) at the expected location (spatially and spectrally) of the [O III] 88 µm line. Notably, this feature is spectrally and spatially offset from the earlier reported tentative detection of Bakx et al. (2023).

The redshift uncertainty ($\Delta z = \pm 0.02$) derived from the JWST/MRS spectrum of GLASS-z12 translates to a frequency width of $\Delta \nu_{obs} = \pm 0.4$ GHz at the expected [O III] 88 µm frequency of $\nu_{obs} = 254.54$ GHz. We compute the frequency uncertainty from,

$$\left| \frac{\Delta z}{1 + z} \right| = \left| \frac{\Delta \nu_{obs}}{\nu_{obs}} \right| = \left| \frac{\Delta v}{c} \right|, \qquad (3)$$

where Δz is the redshift range probed, $\Delta \nu_{obs}$ is the corresponding frequency width centered on the frequency ν_{obs}, Δv is the corresponding velocity width, and c is the speed of light. This frequency uncertainty is roughly equivalent to the maximum correlated bandwidth of a single ALMA spectral window. Even with better redshift constraints, astrophysical processes like outflows and differing dust geometries could lead to both spatial and spectral offsets, as observed in the [C II] 158 µm and [O III] 88 µm lines reported by Fujimoto et al. (2024). Therefore, the likelihood of false detections, $\mathcal{L}_{FD}(\gamma)$, is still determined by the probability density function similar to the blind search scenario, albeit with the addition of a relatively wide prior of ± 0.4 GHz from the JWST/MRS observations.

To incorporate the redshift prior from auxiliary data sets into our line-finding routine, we adopt a Bayesian approach using Bayes theorem,

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}, \qquad (4)$$

where $P(B|A)$ is the general expression for a probability function – in our case, the probability of a detection, $\mathcal{P}_D(x)$, and false detection, $\mathcal{P}_{FD}(x)$ –, $P(A)$ is the prior, and $P(B)$ is the Bayesian evidence. For the detection inference of GLASS-z12, we adopt a Gaussian prior on the redshift, $\mathcal{G}(\mu = 12.33, \sigma = 0.02)$. We also use a uniform prior on the radial distance, $\mathcal{U}(0, 0.6'')$, and a uniform prior on the S/N of the peak, $\mathcal{U}(2.5, \infty)$, thereby only counting peaks with $S/N > 2.5$. After multiplying the probability distribution by the priors to obtain the posterior distribution, we normalize it such that the integral over the bins of S/N, frequency, and radial distance equals one.

Unlike methods such as Markov chain Monte Carlo, which directly sample the posterior distribution, we first sampled the PDFs using `FindClumps` without applying any priors. We then applied Bayes' theorem in post-processing to determine which feature is most likely to correspond to GLASS-z12. This approach assumes no correlation between the S/N of a candidate (since the intrinsic flux of the source is unknown) and its location along the RA, Dec, and frequency axes. This is a reasonable assumption, as there should be no correlation
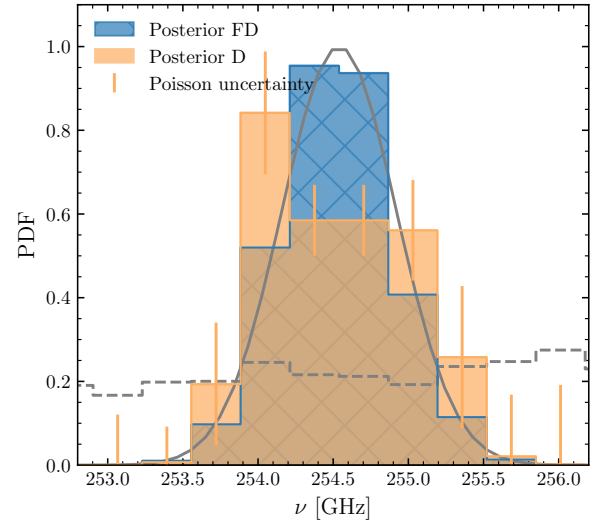

Fig. 9. Posterior distribution of peaks above $S/N > 2.5$ and within $0.''6$ of the JWST derived centroid as a function of frequency for both the real and jackknifed realization. The Gaussian prior derived from JWST/MRS Hα observation of Zavala et al. (2025) is shown as the gray line, and the initial PDF of the probability of false detections is the gray-dashed line.

between the S/N of the line detection and these parameters for non-primary-beam-corrected maps.

We visualize our approach to determining the likelihood of a positive peak being real, given the redshift prior, in Figure 9. The posterior distributions for both the real and jackknifed observations are shown in the frequency domain. We also show the prior from JWST/MRS observations and the initial PDF of the probability of false detections without applying the prior. The latter indicates that the PDF has a flat distribution with respect to frequency, thus supporting the assumption that there is no correlation between the S/N of a peak and its location along frequency axes. We find a slight excess of detection peaks at a frequency of $\nu = 254.0$ GHz, corresponding to two peaks located at RA, Dec, $\nu = 3.49892°, -30.32469°, 253.92$ GHz and RA, Dec, $\nu = 3.49918°, -30.32469°, 253.92$ GHz. These peaks have S/N values of 2.7 and 2.6, and linewidths of 230 km s^{-1} and 140 km s^{-1}, respectively. Given the probability of false detection and the number of peaks found in the real data, the likelihood ratio test results in $\Lambda(\gamma = 2.5\sigma) = 1.64$. This does not exceed our detection threshold of $k > 3$, set in Section 5.1.

For illustrative purposes, Figure 10 shows a moment-0 map with a tentative feature alongside three jackknife realizations. In one of these realizations, we observe a similar spatial flux distribution as in the real data. The moment-0 maps and the posterior distribution clearly indicate that there is no detection of GLASS-z12 in the ALMA data (Project code: 2021.A0020.S, PI: Bakx). Even including a prior from JWST, the ALMA data is statistically consistent with being noise. To reiterate, we did not include in our analysis the newer observations targeting GLASS-z12 that were recently published (Project Code: 2023.A.00017.S, PI: J. Zavala; Zavala et al. 2024, , see also Table 1). That work showed a ≈ 4.5–5.2σ detection of the [O III] 88 µm line when the new observations were combined with the older data.

6. Summary and implications

In this study, we introduce an easy-to-use framework for determining the likelihood of faint emission in interferometric data
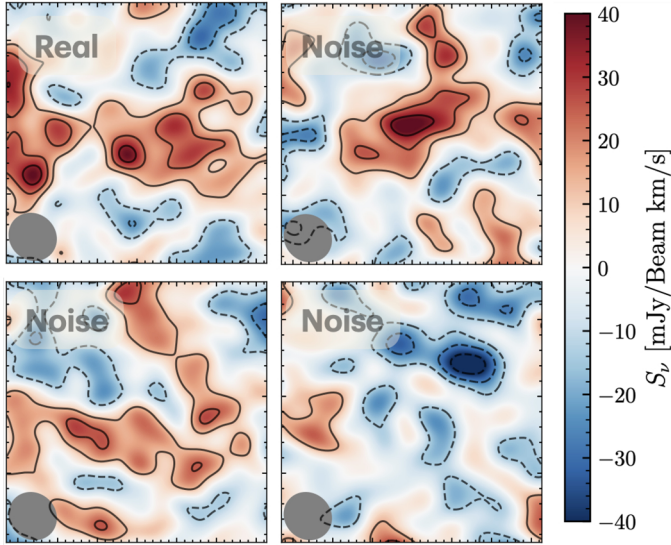
Fig. 10. Moment-0 maps centered at $\nu = 254.35$ GHz with a linewidth of 280 km s^{-1} for the real data set and three jackknife realizations, which are imaged identically. In the lower left, we show the beam size. The size of the images is 1''.7 by 1''.7. The contours are drawn at $\pm 1, \pm 2, \pm 3\sigma$.

bring real. By jackknifing the visibilities, we generate noise realizations of the measurement set. Line-finding tools can then be applied to both the noise and the original data set to quantify the level of false detections. In this work, we have tested our approach using `FindClumps` (Walter et al. 2016), but there are several other line-finding tools, which may be more appropriate for interferometric data, including `SoFiA` (Serra et al. 2014; Westmeier et al. 2021), `LineSeeker` (González-López et al. 2020) or `MF3D` (Pavesi et al. 2018). Our entire analysis procedure, modulo the user's choice of line-finding algorithm, is publicly available[11].

We applied this methodology to three ALMA-DDT observations targeting galaxy candidates at $z > 10$ to recover the likelihood of peaks at a certain S/N being real. Given the number of peaks in each dataset and their respective distribution of peak values, the likelihood ratio test using a minimum detection threshold of $\gamma = 4\sigma$ – as formalized in Eq. (2) – resulted in $\Lambda(\gamma = 4\sigma) = 0.80, 1.4, 1.4$, respectively (see Table 2). Since we set the detection threshold to $\Lambda(\gamma) > k$ with $k = 3$ (see Section 5.1), we conclude that none of the previously reported tentative detections could be distinguished from noise, even when incorporating additional prior information from JWST/MRS slit measurements in the inference.

Our analysis shows that, given the current data volumes ($N_{peaks} \approx 200$–1000), we expect to find approximately 3 ± 2 line features with $S/N \sim 4$–5 in broad ALMA line scans based on the underlying noise distribution. Ensuring a secure detection is therefore challenging, requiring $>5\sigma$ detections when performing a blind line search in a cube probing 30 GHz of bandwidth and targeting a single line. Detecting two lines at matching redshifts would strengthen the significance; however, even then, the likelihood of detecting two noise features within a reasonable spatial and frequency offset needs to be accounted for. Even when probing multiple lines in different bandwidths, spurious features can arise at a significance of $\approx 4\sigma$ at the right

frequency and realistic spatial offset in both data sets (e.g., Kaasinen et al. 2023).

While the elegance of `jackknify` lies in its straightforward implementation, new approaches are needed to optimize line searches computationally and to increase fidelity. There is currently no publicly available line-finding algorithm implemented in *uv*-space for extragalactic sources (see, for instance, Loomis et al. 2018 with the code `VISIBLE`, which they used for finding isotopes in protoplanetary disks in high spatial and spectral resolution observations). Operations in *uv*-space are not affected by side lobes or other sources of correlated noise to which analyses in the image plane are susceptible, but they come at the cost of computational efficiency. The flux of the source is not concentrated within a resolution element but rather spread over a large number of visibilities. Additionally, ongoing ALMA studies – albeit still in the image plane – are exploring the use of unsupervised machine learning to identify faint emission lines (see, e.g., Baronchelli et al. 2021, 2024). Although these methodologies are in the early stages of development, their implementation could drastically increase computation speed and efficiency.

Improving current line-finding methods is important in the context of several upcoming surveys and telescope upgrades. For example, the Wideband Sensitivity Upgrade of ALMA (Carpenter et al. 2023) will increase the bandwidth and sensitivity by a factor of 2–4. This upgrade will greatly increase the efficiency of line-finding experiments, especially at high frequencies, and is, therefore, critical for identifying $z > 7$ galaxies. Looking to the more distant future, significant advances are expected from major new submillimeter-to-centimeter facilities, such as the proposed 50 m single dish, named the Atacama Large Aperture Submillimeter Telescope (AtLAST; Booth et al. 2024; Mroczkowski et al. 2025) and SKA (Dewdney et al. 2009). AtLAST would enable large, unbiased surveys of cosmological volumes in multiple bands through the use of on-chip spectrometers (e.g., Endo et al. 2012) and its large 2° field of view, providing secure line identifications for large samples of high-z galaxies. Although splitting in the visibility plane will not be possible for single-dish facilities, a similar approach where the differencing is done in the time domain will still be useful (see, e.g., Weiß et al. 2009). At centimeter wavelengths, deep HI surveys are already revealing new HI detections up to $z \sim 0.4$ (e.g., Baker et al. 2024; Xi et al. 2024; Kazemi-Moridani et al. 2024), with the full SKA likely to push this to $z \sim 1$. With the increase in data volume associated with these new and upgraded facilities comes the increased potential for spurious line detections. Thus, it is crucial that the community accurately identifies the probability of false detections using statistically motivated approaches like the one we presented in this work.

Data availability

Supplementary material is available at: https://joshiwavm.github.io/jackknify/

---

[11] See https://joshiwavm.github.io/jackknify/, and the tool, `jackknify`, is installable using pip; https://pypi.org/project/jackknify/ (van Marrewijk & Di Mascolo 2024).

References

Baker, A., Blyth, S., Holwerda, B., et al. 2024, in American Astronomical Society Meeting Abstracts, 243, 346.07

Bakx, T. J. L. C., Zavala, J. A., Mitsuhashi, I., et al. 2023, MNRAS, 519, 5076

Baronchelli, I., Scarlata, C. M., Rodríguez-Muñoz, L., et al. 2021, ApJS, 257, 67

Baronchelli, I., Bonato, M., De Zotti, G., et al. 2024, Galaxies, 12, 26

Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393

Boogaard, L., Meyer, R. A., & Novak, M. 2021, https://doi.org/10.5281/zenodo.5775603

Booth, M., Klaassen, P., Cicone, C., et al. 2024, SPIE Conf. Ser., 13102, 1310206

Carniani, S., D'Eugenio, F., Ji, X., et al. 2024a, arXiv e-prints [arXiv:2409.20533]

Carniani, S., Hainline, K., D'Eugenio, F., et al. 2024b, Nature, 633, 318

Carpenter, J., Brogan, C., Iono, D., & Mroczkowski, T. 2023, in Physics and Chemistry of Star Formation: The Dynamical ISM Across Time and Spatial Scales, eds. V. Ossenkopf-Okada, R. Schaaf, I. Breloy, & J. Stutzki, 304

CASA Team (Bean, B., et al.) 2022, PASP, 134, 114501

Castellano, M., Fontana, A., Treu, T., et al. 2022, ApJ, 938, L15

Decarli, R., Aravena, M., Boogaard, L., et al. 2020, ApJ, 902, 110

Dewdney, P. E., Hall, P. J., Schilizzi, R. T., & Lazio, T. J. L. W. 2009, IEEE Proc., 97, 1482

Di Mascolo, L., Saro, A., Mroczkowski, T., et al. 2023, Nature, 615, 809

Endo, A., Werf, P., Janssen, R. M. J., et al. 2012, J. Low Temp. Phys., 167, 341

Fudamoto, Y., Oesch, P. A., Walter, F., et al. 2024, MNRAS, 530, 340

Fujimoto, S., Finkelstein, S. L., Burgarella, D., et al. 2023, ApJ, 955, 130

Fujimoto, S., Ouchi, M., Nakajima, K., et al. 2024, ApJ, 964, 146

González-López, J., Decarli, R., Pavesi, R., et al. 2019, ApJ, 882, 139

González-López, J., Novak, M., Decarli, R., et al. 2020, ApJ, 897, 91

Harikane, Y., Inoue, A. K., Mawatari, K., et al. 2022, ApJ, 929, 1

Harikane, Y., Ouchi, M., Oguri, M., et al. 2023, ApJS, 265, 5

Harikane, Y., Inoue, A. K., Ellis, R. S., et al. 2025, ApJ, 980, 138

Hatziminaoglou, E., Zwaan, M., Andreani, P., et al. 2015, The Messenger, 162, 24

Kaasinen, M., van Marrewijk, J., Popping, G., et al. 2023, A&A, 671, A29

Kazemi-Moridani, A., Baker, A. J., Verheijen, M., et al. 2024, arXiv e-prints [arXiv:2412.11426]

Keating, G. K., Bower, G. C., Marrone, D. P., et al. 2015, ApJ, 814, 140

Kitayama, T., Ueda, S., Akahori, T., et al. 2020, PASJ, 72, 33

Loomis, R. A., Öberg, K. I., Andrews, S. M., et al. 2018, AJ, 155, 182

Mroczkowski, T., Gallardo, P. A., Timpe, M., et al. 2025, A&A, 694, A142

Naess, S., Aiola, S., Austermann, J. E., et al. 2020, J. Cosmology Astropart. Phys., 2020, 046

Naidu, R. P., Oesch, P. A., van Dokkum, P., et al. 2022, ApJ, 940, L14

Napolitano, L., Castellano, M., Pentericci, L., et al. 2025, A&A, 693, A50

Padin, S., Cartwright, J. K., Mason, B. S., et al. 2001, ApJ, 549, L1

Pavesi, R., Sharon, C. E., Riechers, D. A., et al. 2018, ApJ, 864, 49

Petry, D., Stanke, T., Biggs, A., et al. 2020, The Messenger, 181, 16

Planck Collaboration XVI. 2014, A&A, 571, A16

Popping, G. 2023, A&A, 669, L8

Readhead, A. C. S., Myers, S. T., Pearson, T. J., et al. 2004, Science, 306, 836

Romero, C., McWilliam, M., Macías-Pérez, J. F., et al. 2018, A&A, 612, A39

Schouws, S., Bouwens, R. J., Ormerod, K., et al. 2024, arXiv e-prints [arXiv:2409.20549]

Serra, P., Westmeier, T., Giese, N., et al. 2014, SoFiA: Source Finding Application, Astrophysics Source Code Library [record ascl:1412.001]

Sharp, M. K., Marrone, D. P., Carlstrom, J. E., et al. 2010, ApJ, 713, 82

Thompson, A. R., Moran, J. M., & Swenson, George W., J. 2017, Interferometry and Synthesis in Radio Astronomy, 3rd edn.

Tsukui, T., Iguchi, S., Mitsuhashi, I., & Tadaki, K. 2023, J. Astron. Telesc. Instrum. Syst., 9, 018001

van Marrewijk, J., & Di Mascolo, L. 2024, https://doi.org/10.5281/zenodo.12516585

Vio, R., & Andreani, P. 2016, A&A, 589, A20

Walter, F., Decarli, R., Aravena, M., et al. 2016, ApJ, 833, 67

Weiß, A., Kovács, A., Coppin, K., et al. 2009, ApJ, 707, 1201

Westmeier, T., Kitaeff, S., Pallot, D., et al. 2021, MNRAS, 506, 3962

Xi, H., Peng, B., Staveley-Smith, L., et al. 2024, ApJ, 966, L36

Yoon, I., Carilli, C. L., Fujimoto, S., et al. 2023, ApJ, 950, 61

Zavala, J. A., Bakx, T., Mitsuhashi, I., et al. 2024, ApJ, 977, L9

Zavala, J. A., Castellano, M., Akins, H. B., et al. 2025, Nat. Astron., 9, 155

Appendix A: Further comparison between the simulated noise and jackknifed realizations

Figure A.1 shows the difference between the standard deviation from the cleaned map of the first, single simulated observation (i.e., one seed used for `simobserve` visualized with the orange point) and the distribution of standard deviations obtained from the various jackknife realizations (blue dotted line). We find that the jackknifed realizations follow the average distribution of standard deviations, obtained with various `simobserve` seeds (solid line, Fig. 5), but their median clearly differs from the first single `simobserve` estimate. This shows that jackknifing can be used to better describe the underlying distribution function than a single noise realization used in `simobserve`.
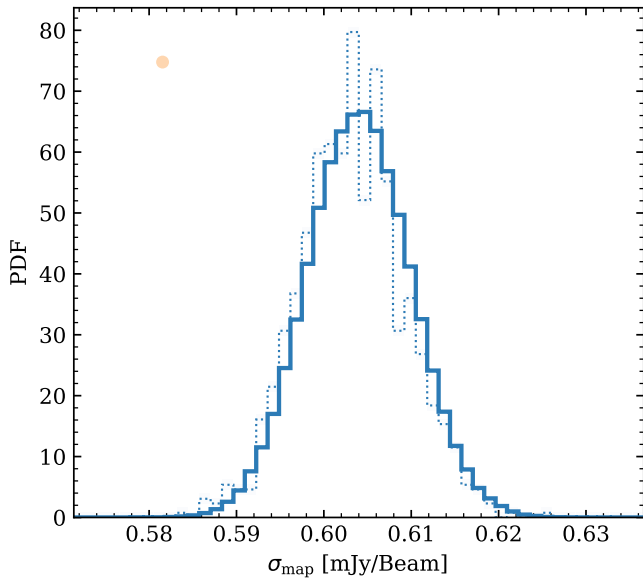

Fig. A.1. Comparison of the standard deviation σ_{map} measured from a single `simobserve` simulation (i.e., using one random seed; orange point also shown in Fig 5) and the output of jackknifing that single realization (blue dotted line). The latter clearly follows the total distribution of the derived standard deviations (blue solid line of Fig. 5) instead of being centered on the single `simobserve` realization.