MNRAS **486,** 2695–2704 (2019)
Advance Access publication 2019 April 8
doi:10.1093/mnras/stz977

Radio galaxy detection in the visibility domain

A. Malyali,[1,2★] M. Rivi,[1,3] F. B. Abdalla[1,4] and J. D. McEwen[5]

[1]*Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK*
[2]*Max-Planck-Institut für extraterrestrische Physik, Giessenbachstrasse 1, D-85748 Garching, Germany*
[3]*INAF - Istituto di Radioastronomia, via Gobetti 101, I-40129 Bologna, Italy*
[4]*Department of Physics and Electronics, Rhodes University, PO Box 94, Grahamstown, 6140, South Africa*
[5]*Mullard Space Science Laboratory, University College London, Surrey RH5 6NT, UK*

Accepted 2019 March 28. Received 2019 March 24; in original form 2018 October 30

ABSTRACT

We explore a new Bayesian method of detecting galaxies from radio interferometric data of the faint sky. Working in the Fourier domain, we fit a single, parameterized galaxy model to simulated visibility data of star-forming galaxies. The resulting multimodal posterior distribution is then sampled using a multimodal nested sampling algorithm such as MULTINEST. For each galaxy, we construct parameter estimates for the position, flux, scale length, and ellipticities from the posterior samples. We first test our approach on simulated SKA1-MID visibility data of up to 100 galaxies in the field of view (FOV), considering a typical weak lensing survey regime (SNR $\geq$ 10) where 98 per cent of the input galaxies are detected with no spurious source detections. We then explore the low-SNR regime, finding our approach reliable in galaxy detection and providing in particular high accuracy in positional estimates down to SNR $\sim$ 5. The presented method does not require transformation of visibilities to the image domain, and requires no prior knowledge of the number of galaxies in the FOV, thus could become a useful tool for constructing accurate radio galaxy catalogues in the future.

Key words: methods: statistical – techniques: interferometric – radio continuum: galaxies.

1 INTRODUCTION

The detection of galaxies from radio interferometric data has traditionally relied on a set of image reconstruction techniques to transform measured visibilities to the image domain (Högbom 1974; Bhatnagar & Cornwell 2004; Cornwell 2008), on which image analysis tools are applied to measure source properties. The imaging process usually transforms visibilities into images, corrected for an approximated point spread function, by an iterative non-linear process. This can introduce artefacts and correlated noise distributions, with subsequent measurements of scientific parameters suffering from systematic errors that are difficult to accurately estimate. In addition, uncertainties in the image domain are typically not accessible, although recent work can provide uncertainty quantification, e.g. Cai, Pereyra & McEwen (2018a,b).

Until recently this has not been a major issue, but the increased sensitivities and size of the forthcoming generation of radio interferometers, such as the Square Kilometre Array (SKA)[1] (Dewdney et al. 2009), will allow new scientific measurements, such as radio weak lensing (Brown et al. 2015), that require more reliable and complete source catalogues, meaning higher accuracy in galaxy detection and characterization. Observations of large fields of view, or with non-coplanar baselines, will be also affected by direction-dependent effects (DDE) that should be taken into account in the data analysis (Wrobel & Walker 1999). An alternative approach to usage of image reconstruction techniques is to work directly in the visibility domain, where the data originate and are not yet affected by the systematics introduced by the imaging process. DDE modelling, e.g. Smirnov (2011b,c), may also be easily introduced in model fitting techniques for galaxy parameter estimation. However, adopting a source model fitting approach to visibilities is very challenging because sources are no longer localized in the Fourier domain and their flux is mixed together in a complicated way. Furthermore, the computational challenges for a telescope such as SKA are great, given the large size of data sets (order of PetaBytes) that must be processed and the expected high source number densities.

Available generic tools for model fitting in the Fourier domain are based on simple source brightness profiles or models obtained from the combination of basic shapes that are not sufficiently realistic (Martí-Vidal et al. 2014). Other methods are specific for galaxy shape measurement for radio weak lensing and rely on information, such as the number of sources and galaxy positions, determined via some pre-processing of the data set or data analysis in the image domain. The first one decomposes galaxy shapes through an orthonormal basis of functions, *shapelets* (Refregier & Alexandre

★ E-mail: amalyali@mpe.mpg.de
[1]https://skatelescope.org/

2003; Refregier & Bacon 2003), and coefficients are jointly fitted solving normal equations (Chang, Refregier & Helfand 2004). The second one, *RadioLensfit* (Rivi et al. 2016; Rivi & Miller 2018), is an adaptation to the radio data of the optical method *lensfit* (Miller et al. 2013), where a source extraction technique has been implemented in order to fit visibilities of a single galaxy at a time. The single source ellipticity fitting is performed adopting a realistic star-forming galaxy brightness profile and marginalizing the likelihood over flux, position, and size. Another approach (Rivi et al. 2019) applies Hamiltonian Monte Carlo (HMC) technique for a joint fitting of all sources in the field of view (FOV) adopting the same galaxy model used for *RadioLensfit*. Although HMC accelerates posterior sampling convergence with respect to standard Monte Carlo Markov Chain (MCMC) with a reasonable efficiency even for high-dimensional problems, it is still a computationally demanding approach to apply to the very large number of sources that will be observed in future radio surveys. Moreover, if the number of sources in the data set is unknown, then the posterior sampling algorithm must be able to handle changing dimensions of the parameter space, such as via reversible-jump MCMC (Green 1995) for the case of a multiple source model. A single object model approach can avoid this changing dimensionality issue due to forming a multimodal posterior in the single object parameter space, with the bonus that it is typically computationally cheaper as the number of parameters to fit is limited to the ones of a single source.

A first attempt to detect discrete objects in astronomical data sets with a single object model is presented in Hobson & McLachlan (2003), where two iterative methods (using MCMC and a simulated annealing simplex technique) are applied to microwave maps dominated by emission from primordial cosmic microwave background anisotropies in order to detect the thermal and kinetic Sunyaev–Zel'dovich effects from clusters of galaxies. Feroz, Marshall & Hobson (2008) also demonstrate the strength of the single model approach using multimodal nested sampling to detect galaxy clusters and estimate their parameters from images of *N*-body simulations.

In this work, we explore a similar Bayesian technique for detecting and characterizing star-forming galaxies in the Fourier domain. We fit a single galaxy model to simulated interferometer data and use the multimodal nested sampling algorithm MULTINEST (Feroz & Hobson 2008; Feroz, Hobson & Bridges 2009; Feroz et al. 2013) for detecting multiple galaxy sources in the data and measuring each galaxy's properties. This approach is self-consistent, as it does not require knowledge in advance of the number of galaxies in the FOV. We present the relevant theoretical background and our method in Section 2. We then explore the detection ability of this approach on simulated radio observations in Section 3, and conclude with a discussion and summary of the effectiveness of this new method in Section 4.

2 METHODOLOGY

We follow the model fitting approach presented in Feroz et al. (2008) adopting a single galaxy model. This model allows for working with a small number of parameters, but returns a highly multimodal posterior that is then sampled using MULTINEST. Each mode should correspond to a different galaxy in the FOV; however, noise and source contaminations usually produce a number of false positives that must be recognized. We begin by briefly introducing the relevant underlying Bayesian framework in Section 2.1 and then present details of the adopted model and parameters' prior distributions in Sections 2.2 and 2.3.

2.1 Bayesian statistics

Bayesian methods for the estimation of a set of parameters Θ in a model/hypothesis H for given data $\mathbf{D}$ estimate the posterior probability distribution $\Pr(\Theta \mid \mathbf{D}, H)$ of the values of Θ. Bayes' theorem reads:

$$\Pr(\Theta \mid \mathbf{D}, H) = \frac{\Pr(\mathbf{D} \mid \Theta, H)\Pr(\Theta \mid H)}{\Pr(\mathbf{D} \mid H)}, \tag{1}$$

where the likelihood $\mathcal{L}(\Theta) \equiv \Pr(\mathbf{D} \mid \Theta, H)$ encodes the constraints imposed by observations, $\pi(\Theta) \equiv \Pr(\Theta, H)$ is the probability distribution of the parameters based on prior knowledge of the parameter values (prior), and $\mathcal{Z} \equiv \Pr(\mathbf{D} \mid H)$ is a normalization factor, known as the *Bayesian evidence*, corresponding to the average of the likelihood weighted by the prior over its *N*-dimensional parameter space:

$$\mathcal{Z} = \int \mathcal{L}(\Theta)\pi(\Theta)\mathrm{d}^N\Theta. \tag{2}$$

Parameter estimates along with their associated uncertainties are derived by sampling the un-normalized posterior, as the evidence is independent of the parameters Θ. $\mathcal{Z}$ becomes relevant for model selection, where a choice between two models H_0 and H_1 has to be decided based on the comparison of their respective posterior probabilities given the observed data $\mathbf{D}$. Applying Bayes' theorem to invert the order of conditioning in the evidence, we obtain the model posterior probability:

$$\Pr(H_k \mid \mathbf{D}) = \frac{\Pr(\mathbf{D} \mid H_k)\Pr(H_k)}{\Pr(\mathbf{D})}, \quad k = 0, 1. \tag{3}$$

Then, the ratio between the models' posteriors is:

$$R = \frac{\Pr(H_1 \mid \mathbf{D})}{\Pr(H_0 \mid \mathbf{D})} = \frac{\mathcal{Z}_1}{\mathcal{Z}_0}\frac{\Pr(H_1)}{\Pr(H_0)}. \tag{4}$$

The ratio of the models' evidences is called the *Bayes factor* and reflects the relative strength of support for each model given the data. More details about Bayesian inference in the context of astrophysics are provided in Trotta (2008).

Nested sampling (Skilling 2004; Skilling et al. 2006) is a method mainly used for efficient and accurate evidence evaluation relative to traditional MCMC methods, yet also provides draws from the posterior as a by-product. The boost in efficiency arises from converting a multidimensional integral into effectively a one-dimensional integral, achieved through the relation between enclosed prior mass, $X(\lambda)$, and likelihood:

$$X(\lambda) = \int_{\mathcal{L}(\Theta)>\lambda} \pi(\Theta)\mathrm{d}^N\Theta. \tag{5}$$

The nested sampling algorithm begins with drawing a set of independent random points from the prior (*live points*). In each iteration, the point with lowest likelihood, $\mathcal{L}_{\rm old}$, is removed and replaced with a new point randomly drawn from the prior, subject to the constraint that its likelihood $\mathcal{L}$ is greater than $\mathcal{L}_{\rm old}$. This process continues until convergence is reached; typically when the fraction of live enclosed posterior mass is a small fraction of the overall mass (Keeton 2011). The evidence is then obtained through quadrature:

$$\mathcal{Z} \approx \sum_{i=1}^{N_{\rm p}} \mathcal{L}_i w_i = \frac{1}{2}\sum_{i=1}^{N_{\rm p}} \mathcal{L}_i(X_{i-1} - X_i), \tag{6}$$

where $N_{\rm p}$ is the number of discarded points, $\mathcal{L}_i$ is the likelihood of the *i*-th discarded point, $w_i = (X_{i-1} - X_i)/2$ is the weight assigned

to each point. Draws from the posterior, p_i, are obtained as:

$$p_i = \frac{\mathcal{L}_i w_i}{\mathcal{Z}} \tag{7}$$

and these samples can then be used to estimate the mean and standard deviation of the model parameters.

MULTINEST[2] is an implementation of nested sampling based on ellipsoidal rejection sampling and tailored for sampling multimodal posteriors. It identifies peaks in the posterior distribution and returns a set of posterior samples for each mode. MULTINEST also returns the *local evidence* $\mathcal{Z}_{\rm loc}$ of each mode based on the set of posterior samples associated with it (as defined in section 5.7 of Feroz et al. 2009). Since we use only a six parameter model, MULTINEST has the best performance amongst other available nested sampling implementations, such as POLYCHORD (Handley, Hobson & Lasenby 2015a,b), which scale well to higher dimensional settings.

2.2 Galaxy likelihood

We adopt a single star-forming galaxy visibility model defined analytically as the Fourier transform of a Sérsic model of index $n = 1$ (exponential), to estimate a galaxy's position (l, m), flux S, scale length α and ellipticity components (e_1, e_2) (Rivi et al. 2016). This galaxy model is defined as only having a disc component because radio emission is dominated by synchrotron radiation from relativistic electrons in the interstellar medium of the galaxy disc.

Following Rivi et al. (2019), model visibilities are simulated by using the GPU-accelerated implementation of the radio interferometer measurement equation (Hamaker, Bregman & Sault 1996; Smirnov 2011a,b), provided by the open-source software MONTBLANC[3] (Perkins et al. 2015). This tool, developed in support of Bayesian inference of radio observations (Lochner et al. 2015), also returns the likelihood computation given the observational data and the noise variance.

2.3 Priors

We adopt a uniform prior on source position over the interferometer's FOV. For the other galaxy parameters we use distributions presented in Rivi & Miller (2018) and estimated from measurements of faint sources observed with the Very Large Array (VLA) radio telescope. In particular the flux and scale length priors are obtained by the analysis of the deep radio VLA-SWIRE field catalogue:

$$\pi(S) \propto S^{-1.34} \tag{8}$$

and $\pi(\alpha)$ is a lognormal distribution (flux independent) with mean $\mu \sim 0.27$ arcsec and variance $\sigma \sim 0.31$ arcsec. The prior on the ellipticity modulus is obtained by fitting the function proposed in Miller et al. (2013) to the VLA-COSMOS field data, using a maximum ellipticity value of 0.804 as in the optical case.

3 TESTING ON SKA1-MID SIMULATED RADIO OBSERVATIONS

3.1 Data simulation

SKA-MID will be a dish array located in South Africa made, in Phase 1, of 64 MeerKAT dishes in a moderately compact core with a diameter of about 1 km and 133 SKA1 dishes distributed in the core and in three logarithmically spaced spiral arms emanating from the centre and extending out to a maximum radius of 80 km, with a maximum baseline of 150 km. We use the antennae configuration provided in Heystek (2015) to simulate 8 h observations of a 1 deg^2 FOV at declination of -30° (i.e. at the zenith), in a single smeared-out frequency channel between 1280 and 1520 MHz and with visibilities sampled once every $\tau_{\rm acc} = 60$ s.

We generate realistic star-forming galaxy populations as in Rivi & Miller (2018), where sources are randomly distributed within the FOV and scale lengths are flux dependent according to the following linear relation between the log median values: $\ln[\alpha_{\rm med}/{\rm arcsec}] = -0.93 + 0.33\ln[S/\mu{\rm Jy}]$. Observed visibilities are computed adopting the exponential profile, as for the model visibilities, and we add uncorrelated Gaussian noise to these with variance given in Wrobel & Walker (1999) assuming all the antennas are SKA1-MID dish antennae.

3.2 Source detection results

We use the PYTHON implementation of MULTINEST, PyMultiNest[4] (Buchner et al. 2014) for sampling the posterior, running it in multimodal mode and setting the sampling efficiency to 0.8. This corresponds to the enlarging of the ellipsoidal bounds by a factor of 1.25 between iterations. The number of live points, $N_{\rm live}$, used determines the resolution of the algorithm at finding posterior modes. If $N_{\rm live}$ is too small, then MULTINEST can miss drawing samples from a subset of the modes and no information can be obtained about that region of the posterior (see Fig. 4 where lower $N_{\rm live}$ values result in fewer detected sources). This could lead to inferences between different seeded runs of MULTINEST being inconsistent, due to two different runs potentially having vastly different sampled posteriors, thus it is crucial that $N_{\rm live}$ is set to a sufficiently high value. On the other side a very large number of live points increases the computational cost and may produce a significant number of fake modes when the likelihood function is not sufficiently smooth. This could happen in the case of too much noise in the data, or due to interference with the signal of neighbourhood sources. Since the number of posterior modes is dependent on the number of simulated galaxies in the FOV, $N_{\rm gal}$, we vary $N_{\rm live}$ between simulations according to $N_{\rm gal}$.[5] We observe that the total number of modes detected by MULTINEST is typically proportional to the number of galaxies and the number of live points used. For more accurate object detection, we set clustering on the live points to be performed only on the source position parameters l, m within the sampling routine. In these parameter spaces, posterior samples from each mode generally have the least overlap with other modes (effectively non-degenerate in cases of non-overlapping sources), compared with the highly degenerate spaces of scale length, flux, and ellipticity (α, S, e_1, and e_2).

As in Feroz et al. (2008), we find that the high sensitivity of MULTINEST to structure within the posterior results in it returning a larger number of posterior modes than the number of sources. There is typically a population of true modes that correspond to galaxies, and 'fake' modes that are spurious. The fake modes can

[2] https://ccpforge.cse.rl.ac.uk/gf/project/multinest/
[3] https://github.com/ska-sa/montblanc/
[4] https://johannesbuchner.github.io/PyMultiNest
[5] Based on instrument detection sensitivity and simulations of extragalactic radio source populations, estimates of the expected source number densities may be computed for real data (e.g. see Bacon et al. 2018 for the expected number of galaxies for the two planned continuum surveys with SKA1).

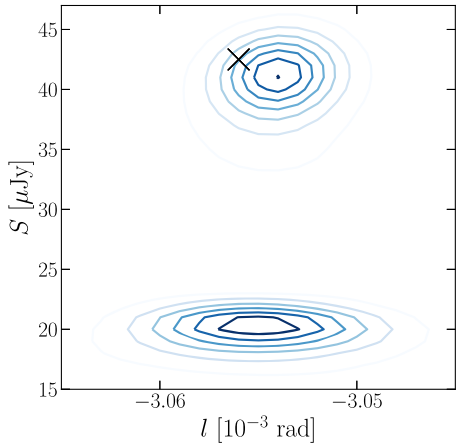

Figure 1. Position l versus flux density S of posterior samples plotted for two modes returned by MULTINEST. The black cross marks the position and flux of the true galaxy. Only the top mode, which has the highest $\mathcal{Z}_{\rm loc}$ within the cluster, samples from the correct region. The second one is an example of an F1 mode. Coordinate m for each mode is consistent with each other, but it is omitted for illustrative purposes. The darker the contour, the higher the density of posterior samples in that parameter space region.

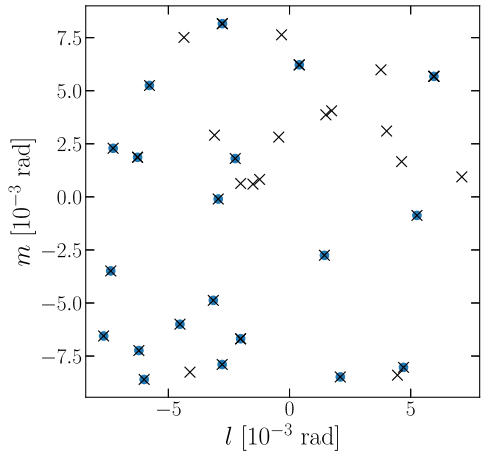

Figure 2. An example of the problem caused by 'fake' modes. Blue circles represent the positions of the 20 simulated galaxies, with flux ranging between 10 and 200 μJy, whilst black crosses are the l, m estimates for all modes returned by MULTINEST (in some cases they are single modes whereas in others they are clusters of modes). The lower bound of the flux prior is 9 μJy. F2 modes generally have fluxes between 9 and 10 μJy.

be further classified as two different types, which we call here as F1 and F2 modes. F1 modes generally have position estimates that are consistent with a galaxy source, resulting in small clusters of modes forming very close to the locations of true modes.[6] However, their α, S, e_1, e_2 estimates are inconsistent with the corresponding true galaxy parameters – for example, their fluxes are greatly underestimated (see Fig. 1) and they have a lower local evidence $\mathcal{Z}_{\rm loc}$ than the true mode in the cluster.

F2 modes are those which appear at seemingly random positions within the FOV, which vary (in number and also parameter estimates) between different seeded sampling runs. In Fig. 2, we show an example of modes (plotted as black crosses) detected by MULTINEST from a simulated observation of 20 galaxies with flux ranging between 10 and 200 μJy, corresponding to a signal-to-noise ratio SNR ≥ 10. F2 modes are the crosses which have no true galaxy counterpart (blue circles), whilst F1 modes generally overlap each other and are located on the blue circles (thus not distinguishable by eye). F2 modes typically have much lower local evidence and their fluxes lie close to the lower bound of the flux prior due to posterior sampling for these modes mainly driven by the shape of the prior. In addition, their estimated scale lengths are much larger than expected on average from real sources at such low flux (see Fig. 3), resulting in the peak brightness and SNR of F2 modes being distinctly low relative to the population of true modes.

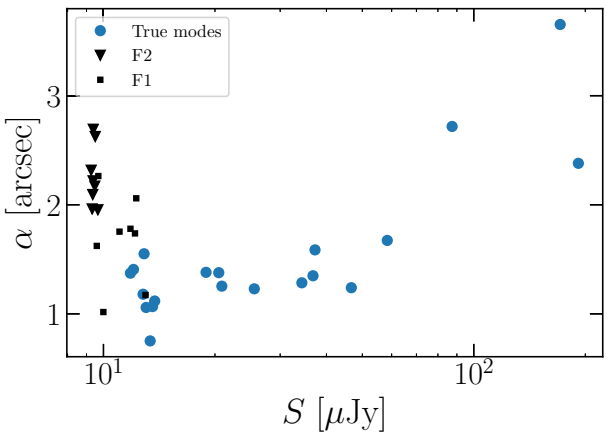

Figure 3. Estimated scale lengths and fluxes for all modes returned by MULTINEST on a simulated population of 20 galaxies, with flux prior ranging between 9 and 200 μJy. Blue circles represent the 'true' estimated modes, whilst F1 and F2 modes are shown in black squares and triangles, respectively. We see that F2 modes have significantly larger than expected scale lengths given their low estimated fluxes (relative to the population of true modes).

3.3 Modal selection

If we naïvely assumed all returned modes were galaxies, then the presence of the fake modes would lead to a high number of false positives that would make accurate galaxy detection unfeasible. In the following, we present a method to remove F1 modes via clustering (see Section 3.3.1) and then we discuss in Section 3.3.2 two approaches to pick out from the remaining F2 modes the true detections, assuming no prior knowledge about the number of galaxies within the FOV.

3.3.1 *Clustering*

Since F1 modes are grouped in clusters around the true galaxy positions, we use an unsupervized, non-parametric clustering algorithm called *mean shift*, implemented in the PYTHON package SCIKIT-LEARN[7] and run with bandwidth 10^{-4} rad (approximately the uncertainty in positional parameter estimates), to identify the

[6]Despite performing clustering only in positional space, it seems that the default tolerance of MULTINEST for splitting live points into separate clusters is too stringent for our application, resulting in multiple modes effectively at the same position. Therefore, we need to perform some additional clustering (see Section 3.3.1).

[7]http://scikit-learn.org/stable/modules/generated/sklearn.cluster.MeanShift.html

centre positions of these mode clusters and the corresponding modes contained in each. For each detected modal cluster, we identify the mode with the highest local evidence and discard the remaining modes. Our final parameter estimate for the galaxy at the cluster centre is then constructed from the posterior samples of the remaining mode. We only select this one, and do not combine the posterior samples from all modes within a cluster to construct parameter estimates for the galaxy, because only the highest $\mathcal{Z}_{\rm loc}$ mode typically samples from a region of the posterior consistent with the true parameter values (see Fig. 1). This results in only true or spurious source detections.

3.3.2 Removing spurious detections

As mentioned in Section 3.2, F2 modes have low SNR and flux estimates heavily dependent on the range of the flux prior provided to MULTINEST. Due to this, a natural choice for removing them is to choose a suitable lower bound for the flux prior and discarding all modes below an SNR threshold.

Since this approach could be too empirical, we compare it with a more rigorous method such as the *Bayesian model selection* described in Section 2.1 and used in Feroz et al. (2008) for galaxy cluster detection. It consists in choosing between two contesting hypotheses based on the value of a parameter of the source model for which the prior distribution is known. Therefore, we perform this comparison testing the source flux and applying a threshold to the flux instead of the SNR. If $S_{\rm cut}$ is too low, then not all fake detections are removed. On the other side, if it is too high, then we may cut out all false positives but also begin to remove true ones, limiting our ability to detect low-flux galaxies (see Fig. 4 based on multiple tests on simulated source populations between 10 and 50 galaxies in the FOV).

Similarly to Feroz et al. (2008), we define the following hypotheses:

H_0: a galaxy with flux $S_{\rm min} < S \leq S_{\rm lim}$ is contained in the FOV Ω,

H_1: a galaxy with $S_{\rm lim} < S < S_{\rm max}$ is contained in Ω,

where $S_{\rm lim}$ is a lower bound of interest for the source flux. The model selection ratio given in equation (4) is calculated as follows. For each hypothesis H_k, the associated evidence is:

$$\mathcal{Z}_k = \int \mathcal{L}(\Theta)\pi_k(\Theta)\mathrm{d}^N\Theta, \tag{9}$$

where

$$\pi_k(\Theta) = \pi(l, m)\pi_k(S)\pi(\alpha)\pi(e), \tag{10}$$

for $k = 0$, 1 are the priors defining the hypotheses. The source parameter priors are defined as in Section 2.3, and they differ between the two hypotheses only for the flux: $\pi_0(S)$ and $\pi_1(S)$ are the power-law distribution defined in equation (8) normalized over the ranges $S_{\rm min} < S \leq S_{\rm lim}$ and $S_{\rm lim} < S < S_{\rm max}$, respectively, and zero elsewhere. The corresponding local evidences are returned by MULTINEST for each mode and we then cross-match the returned posterior modes between runs for each hypothesis by their position estimates using the software STILTS.[8] We find that the H_0 run typically returns much larger number of modes than for H_1. As F1 modes are removed by clustering first, this matching only involves F2 and true modes. The two hypotheses return different sets of F2 modes, but generally have at least one mode present at the expected position of a true mode, enabling us to associate the modes

[8]http://www.star.bris.ac.uk/ mbt/stilts/

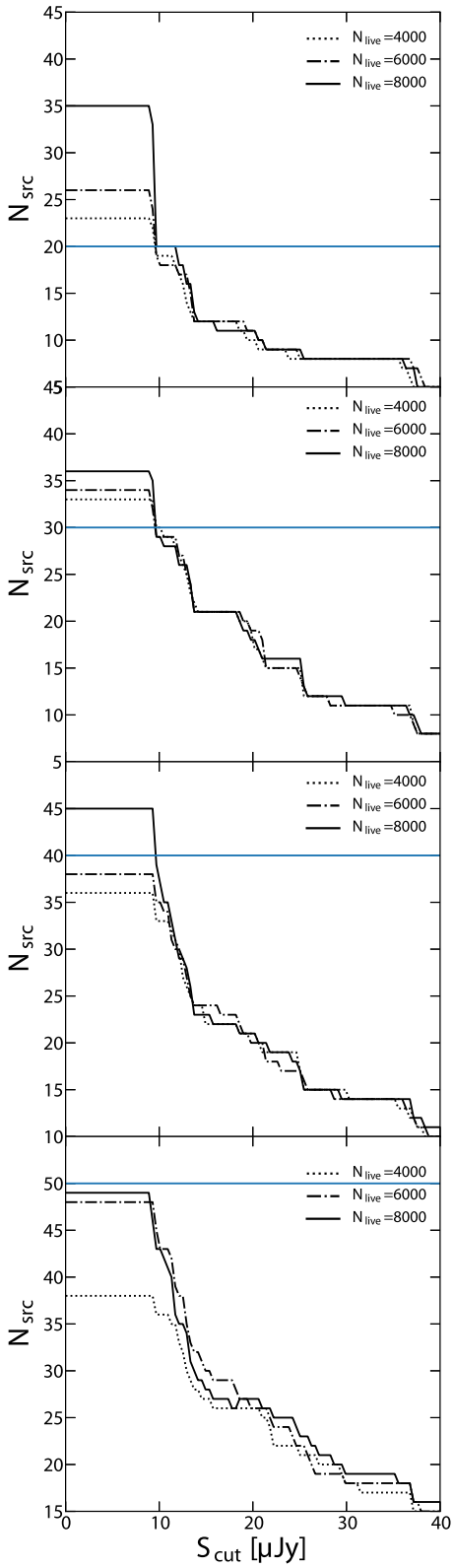

Figure 4. The number of clustered modes, each corresponding to a single source detection, remaining after discarding modes below different flux cuts. From top to bottom, results refer to $N_{\rm gal}$ = 20, 30, 40, 50, with blue line representing $N_{\rm gal}$ in each plot. For each case we show plots corresponding to three different numbers of live points: $N_{\rm live}$ = 4000, 6000, 8000. The simulated sources have flux $S \geq 10$ μJy and flux prior down to 9 μJy.

between the runs. The prior ratio $\Pr(H_1)/\Pr(H_0)$ is obtained from the cumulative flux prior distribution:

$$\frac{\Pr(H_1)}{\Pr(H_0)} = \frac{\int_{S_{\rm lim}}^{S_{\rm max}} \pi(S) dS}{\int_{S_{\rm min}}^{S_{\rm lim}} \pi(S) dS} = \frac{S_{\rm max}^{-0.34} - S_{\rm lim}^{-0.34}}{S_{\rm lim}^{-0.34} - S_{\rm min}^{-0.34}}. \quad (11)$$

We are thus able to compute the models' posteriors ratio R, and the probability of the mode being due to a detection of a galaxy with flux $S > S_{\rm lim}$ is:

$$\Pr(H_1 \mid \mathbf{D}) = R \cdot \Pr(H_0 \mid \mathbf{D}) = \frac{R}{1+R}. \quad (12)$$

We compare performance of the flux cut and Bayesian approaches for detecting 50 galaxies within the FOV in two different cases. The tunable parameters of MULTINEST are kept fixed between runs.

(a) *Population with flux above 10 μJy, corresponding to the typical weak lensing surveys threshold of SNR ≥10.*
We generate visibility data of a galaxy population with fluxes distributed according to equation (8) defined between 10 and 200 μJy. For the Bayesian modal selection, we ran MULTINEST twice on the simulated visibility data: once with flux prior defined between 3–10 μJy and the second between 10 and 200 μJy, corresponding to H_0 and H_1 as above. Each run returns the local evidence for each posterior mode identified for the given hypothesis. After removing F1 modes through the clustering algorithm and matching mode detections between the two runs via position estimates, we use the local evidence values to compute the probability of each detection as belonging to a galaxy with flux greater than 10 μJy as described above. The final sample of galaxies is chosen by selecting modes with probability greater than 0.5. For 50 simulated galaxies, 47 modes are selected as true galaxies with no false positives. An analysis using a single MULTINEST run with flux prior ranging between 9 and 200 μJy and applying a flux cut of 10 μJy, corresponding to the minimum source flux of our simulated catalogue, led to all spurious being removed without the discarding of true positives and returned the same number of detected galaxies as the Bayesian selection.

(b) *Population with low-flux range: 5–10 μJy, corresponding to the typical galaxy catalogue surveys threshold of SNR ≥5.*
Visibility data of 50 galaxies is now simulated with fluxes distributed between 5 and 10 μJy. For implementing the Bayesian modal selection, we considered as two competing hypotheses a flux prior defined between 3 and 5 μJy (H_0) and between 5 and 10 μJy (H_1). In this case, we find that the Bayesian selection returned 61 detections of which 41 true positives and 20 spurious whereas a run with flux prior down to 4 μJy and a flux cut of 5 μJy resulted in detecting 57 sources with 45 real and 12 spurious. Further investigation in the low-SNR regime is explored in Section 3.5.

We conclude that the threshold approach is the most efficient because it produces a lower fraction of false positives and is computationally cheaper as it only needs to run MULTINEST once on the data set. We note that there is an option within MULTINEST to only accept modes with evidence above a minimum value provided by the user, but found that the flux cut approach described above was more robust to changes in the number of galaxies in the FOV and led to a lower number of false positives.

Subsequent detections presented in this paper are obtained by applying an SNR cut as it takes into account not only the source flux but also the size of the source, improving the mode selection. For example, Fig. 5 shows that cutting at SNR = 8 recovers a slightly purer selection of the lowest SNR modes relative to applying a flux cut at 10 μJy. We stress that our computation of the SNR is performed in the visibility domain as in Rivi et al. (2016), which usually returns a larger value than the standard peak brightness over the rms noise method used in the image domain (e.g. see section 4.2.3 of Rivi et al. 2019).

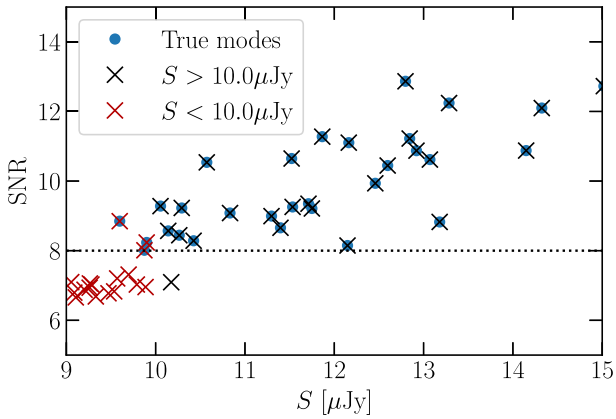

Figure 5. SNR versus estimated flux for all modes returned by MULTINEST on a simulated population of 100 galaxies, with flux ranging between 10 and 200 μJy. Blue circles show the 'true' modes of the run (based on a cross-match with the input galaxy population); 98/100 are selected using an SNR threshold of 8 with no false positives. On the other hand, using $S_{\rm cut} = 10.0$ μJy returns 96/100 galaxies with 1 false positive. Note that we have zoomed in to the low SNR and S region of the plot to highlight the improvement of the SNR cut over a flux cut.

3.4 Galaxy parameter estimates

We report our parameter estimates for each galaxy with the mean and standard deviation of the posterior samples returned for each posterior mode identified as 'true' by our clustering and SNR cut selection algorithm. If needed, our approach also returns the full posterior distribution for each inferred source. Within a run, MULTINEST typically convergences on the source positions earliest, before later converging on the shape parameters. If $N_{\rm live}$ is not sufficiently high, then generally no convergence between runs is attained for S, α, e_1, and e_2.

We show results obtained from simulated visibilities of populations in the radio weak lensing regime, i.e. with SNR $\gtrsim$ 10. In Fig. 6, we plot our parameter estimates for the fitting of 20 galaxies within the FOV, using $N_{\rm live} = 8000$ live points. All sources are detected and with no false positives. We find that source positions and fluxes are fit accurately (although fluxes tend to be underestimated), while scale lengths and ellipticity estimates deviate more from their true value (as expected, since these parameters are notoriously harder to fit). More specifically, scale lengths tend to be overestimated for small input α values, whereas ellipticities are biased towards 0. We find similar behaviour in parameter estimates for the fitting of 50 galaxies (Fig. 7), but now with a source recovery rate of 94 per cent when using $N_{\rm live} = 14\,000$ (increased due to requiring a higher resolution at finding modes) and 100 galaxies (Fig. 8), where we recover 98 per cent of sources using $N_{\rm live} = 35\,000$. For the detected sources in the 20, 50, and 100 source simulations, the mean of the 1σ uncertainty in position and flux parameter estimates is 0.4 arcsec and 1.2 μJy, respectively. Underestimation of source fluxes, which seems greater at larger values of the original flux, is probably due to the finite sampling of the UV coverage causing a loss of flux detection at short baselines, while overestimation of the size for small sources may be due to the resolution of the telescope.

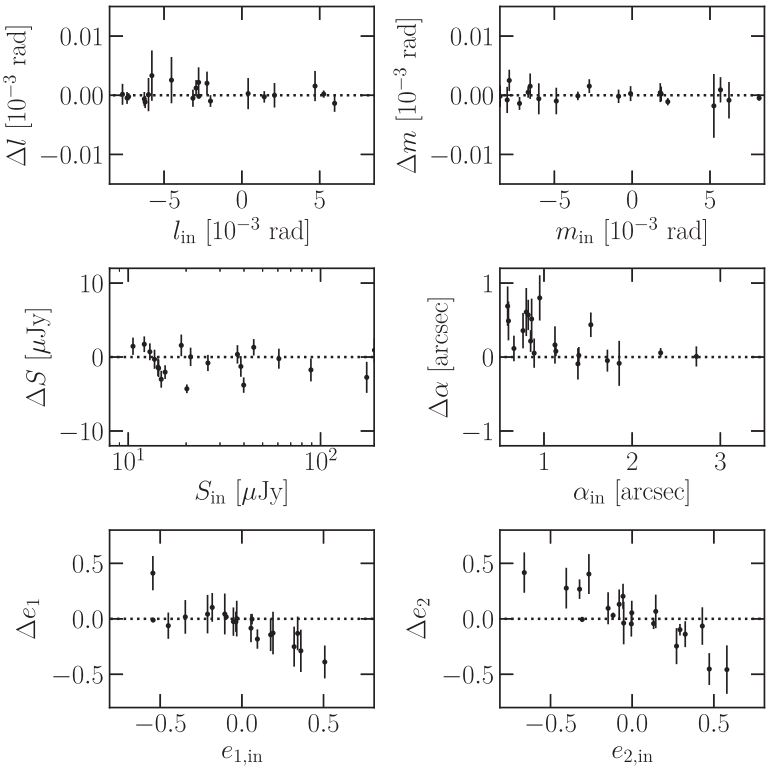

Figure 6. 20/20 sources with input fluxes ranging between 10 and 200 μJy (SNR $\gtrsim$ 10) distributed according to equation (8), detected using $N_{\rm live} = 8000$ points. For each detected galaxy, we plot the residuals of its estimated parameters (measured minus true value). As expected, position and flux measurements are much more accurate than scale length and ellipticity components. Combined with estimates presented in Figs 7 and 8, we find that position estimates appear free from systematic biases, whereas fluxes are generally underestimated. Smaller α values are overestimated, whilst the ellipticities are biased towards zero.

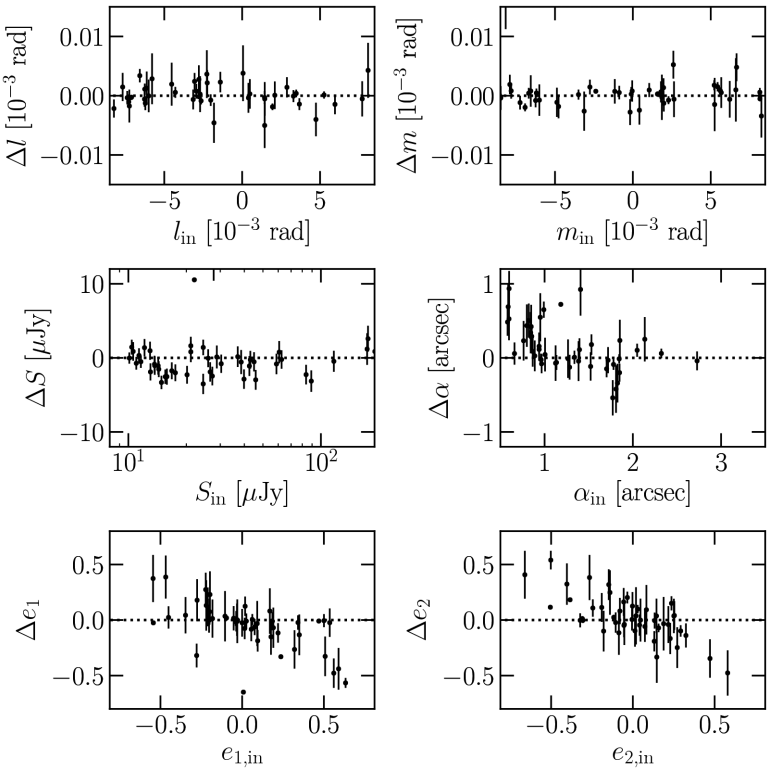

Figure 7. Residual plots for 47/50 sources with input fluxes ranging between 10 and 200 μJy detected using $N_{\rm live} = 14000$ points. 6 per cent of the estimates for S, α, e_1, and e_2 are inconsistent with the true value to within 5σ and these also typically have very small uncertainties in their estimates.

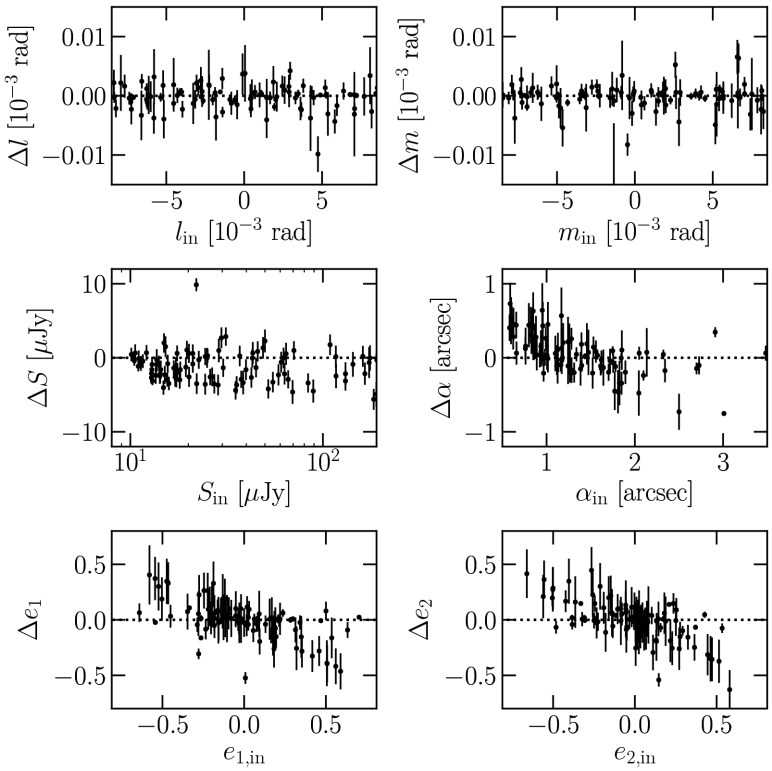

Figure 8. Residual plots for 98/100 sources with input fluxes ranging between 10 and 200 μJy, detected using $N_{\rm live} = 35\,000$ points. About 3 per cent of the estimates for S, α, e_1, and e_2 are inconsistent with the true value to within 5σ.

Relative to existing Bayesian methods in the visibility domain our approach suffers from lower accuracy in shape estimates even in this low source number density regime. This is mainly due to the usage of a single source model which does not account for interference between emission from other galaxies in the FOV. This is shown in Rivi & Miller (2018) where, although the source extraction method tries to remove as much as possible the contamination from the other sources, a neighbouring bias in the shear measurement is still estimated comparing to the case where there is a single source in the FOV. In Rivi et al. (2019) the multisource model provides better shape measurement accuracy as expected. Also, we are measuring all the six galaxy parameters simultaneously, whereas such methods assume true values for source position and flux are known in the model fitting. As this method may be very useful for providing accurate source flux and positions, it could be combined with one of the above methods for a second fine-grained fit of the shape parameters. Shape estimates returned by MULTINEST may be used to accelerate convergence of the HMC-based approach, or as the initial sky model for *RadioLensfit*.

We now explore the ability of the method to detect sources with low SNR for constructing reliable galaxy catalogues for future SKA surveys.

3.5 Signal-to-noise detection threshold

To characterize modal classification ability and to find the optimal SNR threshold for detection, we construct a *receiver operating characteristic* (ROC) curve (see Fig. 9), which is a plot of the true positive rate, TPR, defined as:

$$\mathrm{TPR} = \frac{N_{\rm TP}}{N_{\rm TP} + N_{\rm FN}}, \tag{13}$$

against the false positive rate:

$$\mathrm{FPR} = \frac{N_{\rm FP}}{N_{\rm TN} + N_{\rm FP}}, \tag{14}$$

for different SNR threshold values, where $N_{\rm TP}$, $N_{\rm FP}$, $N_{\rm TN}$, $N_{\rm FN}$ are the number of true positives, false positives, true negatives, and false negatives, respectively, returned by our mode selection algorithm. The TPs, FPs, TNs, and FNs are defined in the context of classifying a mode as real or spurious. We define the real modes as those which have their position estimates consistent with the position of a galaxy in the input source catalogue, and spurious modes as those which do not have a match. Consistency is checked only for the positional parameters as these were found to be the most reliable and the easiest parameters to fit, while the fitting is generally less consistent for fluxes and shape parameters. After selecting the mode with the highest $\mathcal{Z}_{\rm loc}$ in the cluster and performing the SNR cut, we have two sets of modes: those with SNR above and below the $\mathrm{SNR_{cut}}$, which we define as the positives and negatives, respectively. The true and false positives are modes above the $\mathrm{SNR_{cut}}$ that correspond to real and spurious sources, respectively, whilst the true and false negatives are modes below $\mathrm{SNR_{cut}}$ that correspond to real and spurious sources, respectively.

We simulate the visibilities of 50 galaxies within the FOV with fluxes limited to 3–10 μJy (i.e. SNR ranging between 3 and 13). We set the domain of our flux prior to be defined between 3–200 μJy and its normalization factor is adjusted accordingly. The resulting ROC curve lies far from the line one would expect to obtain from random guessing modes as true or spurious (dashed line in Fig. 9) and the area under the curve (AUC) is 0.91, so our modal classifier still performs well on low-SNR sources. The $\mathrm{SNR_{cut}}$ at which we have a good trade-off between the TPR and FPR is at $\mathrm{SNR_{cut}} \sim 3.3$ in the

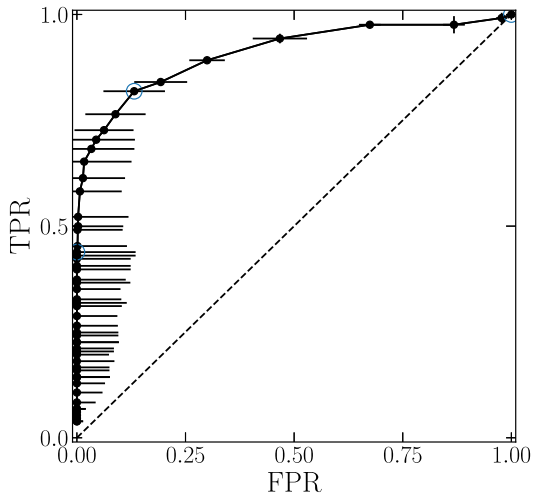

Figure 9. ROC plot for simulated sources within the 3–10 μJy range. Black markers are the mean value for a given SNR cut of FPR and TPR across three different 50 source population simulations, with error bars showing the range of FPR, TPR values at each SNR cut. Circled markers from left to right represent SNR cuts in the visibility domain of ∼4.5, 3.3, and 2.6. Dashed line is $y = x$, which would be the expected curve for randomly guessing the mode type. The SNR threshold at which we have a good trade-off between TPR and FPR is at 3.3.

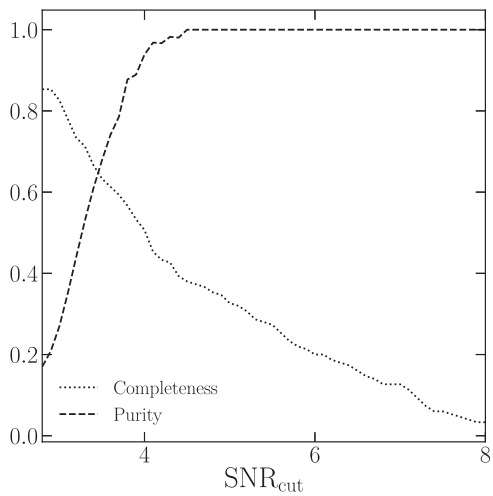

Figure 10. Completeness and purity plots corresponding to the ROC curve shown in Fig. 9. Purity increases with the SNR threshold but lowering the completeness of our final selection of galaxies. We find purity = 1 for SNR cuts above 4.5.

visibility domain; this is estimated as the value which maximizes TPR − FPR. To accompany the ROC curve, we also show in Fig. 10 the *purity* ($N_{\rm TP}/(N_{\rm TP} + N_{\rm FP})$), and *completeness* ($N_{\rm TP}/N_{\rm src}$) plot of our sample of galaxies after modal classification as a function of $\rm SNR_{cut}$. We find that as before, if the threshold is too low, then our approach suffers from low purity but high completeness in the final modes returned. For example, when no selection is applied (at $\rm SNR_{cut} \sim 2.6$) then 16 per cent of the detected modes are true and only 85 per cent of the original population is detected, as we find the very low-SNR sources are not detectable. As the threshold is raised, we increase purity but lower completeness. If high purity is preferred over completeness of the recovered galaxy population, then no false positives are generally found when applying $\rm SNR_{cut} \sim 4.5$.

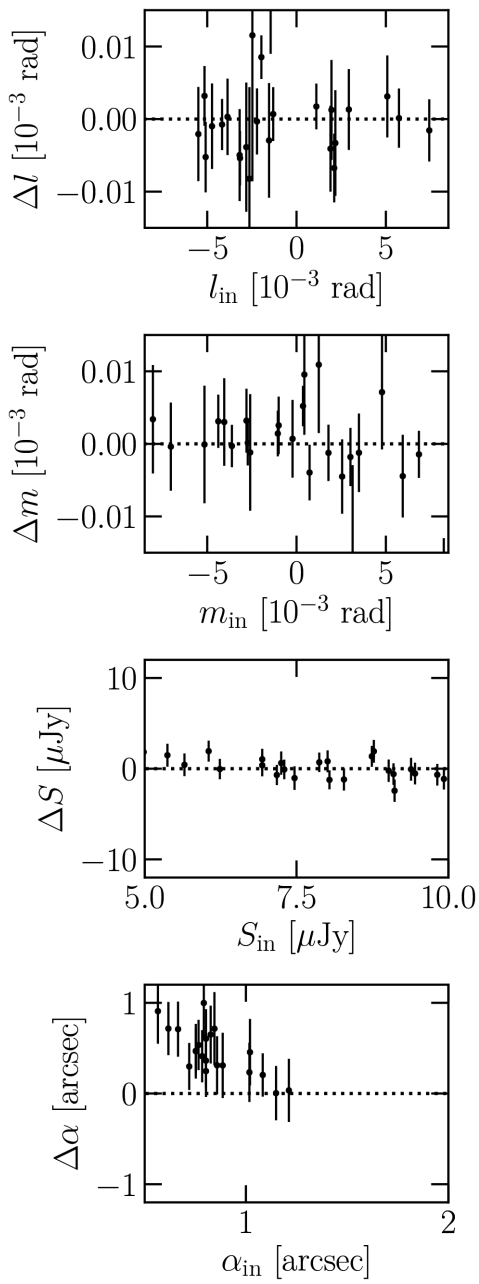

Figure 11. Parameter estimates for fitting of 50 sources in the FOV with SNR ranging approximately between 3 and 13. We use a flux prior down to 3 μJy and select modes with SNR above 4.5 (the minimum $\rm SNR_{cut}$ that produces no false positives), which returns 25/50 galaxies for this particular run, i.e. 58 per cent of the population with SNR ≥ 5.

We note that, as our method typically overestimates sizes in this regime, the SNR of true modes are usually lower than the SNR of the corresponding simulated sources. Therefore in order to recover most of the population above a given SNR, it may be required to apply a cut to the modes less than this SNR. For example, at $\rm SNR_{cut} \sim 5$ we detect on average only ∼50 per cent of the simulated population with SNR ≥ 5, whereas lowering the SNR threshold would result in a greater recovery percentage of this population.

We show the galaxy parameter estimates of the detected sources of one simulation in Fig. 11. As expected for the low-SNR regime, the ability to constrain the galaxy fluxes and size parameters is worse than with higher SNR sources, with the magnitude of galaxy

sizes typically overestimated, but position measurements are still accurate. Across the recovered galaxies, the mean 1σ uncertainty in position and flux parameters is 1.4 arcsec and 1.2 μJy, respectively. Our method does not suffer any significant deterioration in the accuracy of recovered positional and flux estimates, although there can be a larger number of false positives in the low-flux ranges.

4 CONCLUSIONS

We have explored a novel Bayesian method for detecting galaxies from radio interferometric data using a single source model and sampling the resulting multimodal posterior with MULTINEST. Without the need for any transformation to the image domain, our approach extracts out galaxies within visibility data sets, and estimates their positions and other properties from the appropriate posterior modes. We have tested this approach on simulated SKA1-MID observations of up to 100 star-forming galaxies in the FOV and searched for an acceptable SNR threshold where detections are reliable. To remove fake modes we propose to apply the mean shift clustering algorithm to group modes with similar estimated galaxy positions, setting the bandwidth for clustering to roughly the mean of the 1σ uncertainty in positional parameter estimates (i.e. ∼0.4 arcsec). Once clustered, we select the mode with the highest evidence within each cluster as a true detection. The remaining false positives usually have low fluxes and large sizes, so may be removed with a suitable SNR cut. A more rigorous Bayesian hypothesis selection seems to be less efficient, besides the fact that it is much more computationally expensive.

We find that an estimated SNR threshold of ∼4.5 is reasonable for mode selection as spurious modes should be a negligible fraction of the detected sources. From our tests, we also expect not to find spurious modes above SNR ~ 10. We note that spurious modes which remain after performing an SNR cut could be identified in real data sets via cross-matching results with other surveys of the same field at different wave bands. Shape parameters fitting at SNR $\gtrsim 10$ result to be less accurate than other methods proposed for radio weak lensing in the Fourier domain and using the same galaxy model (Rivi & Miller 2018; Rivi et al. 2019). This is expected because these methods take into account the signal interference between nearby galaxies by either using a multisource model or removing an approximation of such source contamination. This is implemented assuming true flux and positions are well-known, whereas these parameters are free in our method. A further investigation for reducing this issue as well as estimating the reliability of this approach at large source densities, where galaxies may not be spatially well-separated, should be performed in future work. Moreover, the impact of AGN structure must be studied as a non-negligible fraction of AGN population should be contained even in the faint radio sky. An initial discussion about this is presented in Rivi et al. (2019).

Since our approach makes no assumptions about the number of galaxies within the FOV, it could become a useful tool for the development of accurate, reliable galaxy catalogues for the next generation of radio interferometers. In order to extend this method to larger source number densities similar to those expected for the SKA, and without using too large a number of MULTINEST live points which would eventually lead to unfeasibly large computing times, a possible solution could be to split the pointing FOV into a number of tiles and then run the code on each one with positional prior limited to the space enclosed by each tile. Once these improvements are made, this approach could also be used in conjunction with *RadioLensfit* or HMC methods for a solely visibility domain-based analysis of galaxy populations for radio weak lensing.

As recent developments in image reconstruction techniques have been able to provide uncertainty quantification (see for e.g. Cai et al. 2018a,b), it would also be interesting to perform a detailed comparison of performance of our approach with these methods for detecting and characterizing the observed galaxy population.

ACKNOWLEDGEMENTS

We thank Edward Edmondson for support with the UCL computer cluster, and Mike Hobson for usage advice on MULTINEST. We also thank the anonymous reviewer for useful comments.

MR acknowledges the support of the Science and Technology Facilities Council via an SKA grant. FBA acknowledges the support of the Royal Society via a Royal Society URF award. This work is partially supported by EPSRS by grant EP/M011089/1, by STFC grant ST/M00113X/1, and by the Leverhulme Trust.

REFERENCES

Bacon D. et al., 2018, preprint (arXiv:1811.02743)
Bhatnagar S., Cornwell T. J., 2004, A&A, 426, 747
Brown M. L. et al., 2015, Proc. Sci., 2, 1365
Buchner J. et al., 2014, A&A, 564, A125
Cai X., Pereyra M., McEwen J. D., 2018a, MNRAS, 480, 4154
Cai X., Pereyra M., McEwen J. D., 2018b, MNRAS, 480, 4170
Chang T.-C., Refregier A., Helfand D. J., 2004, ApJ, 617, 794
Cornwell T. J., 2008, IEEE J. Sel. Top. Signal Process., 2, 793
Dewdney P. E., Hall P. J., Schilizzi R. T., Lazio T. J. L., 2009, Proc. IEEE, 97, 1482
Feroz F., Hobson M. P., 2008, MNRAS, 384, 449
Feroz F., Marshall P. J., Hobson M. P., 2008, preprint (arXiv:0810.0781)
Feroz F., Hobson M. P., Bridges M., 2009, MNRAS, 398, 1601
Feroz F., Hobson M., Cameron E., Pettitt A., 2013, preprint (arXiv:1306.2144)
Green P. J., 1995, Biometrika, 82, 711
Hamaker J. P., Bregman J. D., Sault R. J., 1996, A&AS, 117, 137
Handley W., Hobson M., Lasenby A., 2015a, MNRAS: Letters, 450, L61
Handley W., Hobson M., Lasenby A., 2015b, MNRAS, 453, 4384
Heystek L., 2015, SKA Memo SKA-TEL-INSA-0000537
Hobson M. P., McLachlan C., 2003, MNRAS, 338, 765
Högbom J. A., 1974, A&AS, 15, 417
Keeton C. R., 2011, MNRAS, 414, 1418
Lochner M., Natarajan I., Zwart J. T. L., Smirnov O., Bassett B. A., Oozeer N., Kunz M., 2015, MNRAS, 450, 1308
Martí-Vidal I., Vlemmings W. H. T., Muller S., Casey S., 2014, A&A, 563, A136
Miller L. et al., 2013, MNRAS, 429, 2858
Perkins S. J., Marais P. C., Zwart J. T. L., Natarajan I., Tasse C., Smirnov O., 2015, Astron. Comput., 12, 73
Refregier A., 2003, MNRAS, 338, 35
Refregier A., Bacon D., 2003, MNRAS, 338, 48
Rivi M., Miller L., 2018, MNRAS, 476, 2053
Rivi M., Miller L., Makhathini S., Abdalla F. B., 2016, MNRAS, 463, 1881
Rivi M., Lochner M., Balan S. T., Harrison I., Abdalla F. B., 2019, MNRAS, 482, 1096
Skilling J., 2004, in Fischer R., Preuss R., Toussaint U. V., eds, AIP Conf. Proc. Vol. 735, Bayesian Inference and Maximum Entropy Methods in Science and Engineering. AIP, Am. Inst. Phys., New York, p. 395
Skilling J. et al., 2006, Bayesian Anal., 1, 833
Smirnov O. M., 2011a, A&A, 527, A106
Smirnov O. M., 2011b, A&A, 527, A107
Smirnov O. M., 2011c, A&A, 527, A108
Trotta R., 2008, Contemp. Phys., 49, 71
Wrobel J., Walker R., 1999, in Taylor G. B., Carilli C. L., Perley R. A., eds, ASP Conf. Ser. Vol. 180, Synthesis Imaging in Radio Astronomy II. Astron. Soc. Pac., San Fancisco, p. 171

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