



Radio galaxy shape measurement with Hamiltonian Monte Carlo in the visibility domain

M. Rivi,^{1,2★} M. Lochner,^{1,3,4★} S. T. Balan,¹ I. Harrison^{5★} and F. B. Abdalla^{1,6}

¹Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK

²INAF - Istituto di Radioastronomia, via Gobetti 101, I-40129 Bologna, Italy

³African Institute for Mathematical Sciences, 6 Melrose Road, Muizenberg 7945, South Africa

⁴South African Radio Astronomy Observatory, The Park, Park Road, Pinelands, Cape Town 7405, South Africa

⁵Jodrell Bank Centre for Astrophysics, School of Physics & Astronomy, The University of Manchester, Manchester M13 9PL, UK

⁶Department of Physics and Electronics, Rhodes University, PO Box 94, Grahamstown 6140, South Africa

Accepted 2018 October 2. Received 2018 September 17; in original form 2018 May 29

ABSTRACT

Radio weak lensing, while a highly promising complementary probe to optical weak lensing, will require incredible precision in the measurement of galaxy shape parameters. In this paper, we extend the Bayesian Inference for Radio Observations model fitting approach to measure galaxy shapes directly from visibility data of radio continuum surveys, instead of from image data. We apply a Hamiltonian Monte Carlo (HMC) technique for sampling the posterior, which is more efficient than the standard Monte Carlo Markov Chain method when dealing with a large dimensional parameter space. Adopting the exponential profile for galaxy model fitting allows us to analytically calculate the likelihood gradient required by HMC, allowing a faster and more accurate sampling. The method is tested on SKA1-MID simulated observations at 1.4 GHz of a field containing up to 1000 star-forming galaxies. It is also applied to a simulated observation of the weak lensing precursor survey SuperCLASS. In both cases we obtain reliable measurements of the galaxies' ellipticity and size for all sources with signal-to-noise ratio ≥ 10 , and we also find relationships between the convergence properties of the HMC technique and some source parameters. Direct shape measurement in the visibility domain achieves high accuracy at the expected source number densities of the current and next SKA precursor continuum surveys. The proposed method can be easily extended for the fitting of other galaxy and scientific parameters, as well as simultaneously marginalizing over systematic and instrumental effects.

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

1 INTRODUCTION

The next generation of radio telescopes, such as the Square Kilometre Array (SKA),¹ will reach high enough sensitivity to provide a density of detected galaxies sufficient for weak lensing measurements in the radio band, with the advantage of reaching higher redshifts compared with optical surveys (Brown et al. 2015; Harrison et al. 2016). Standard techniques for the measurement of cosmic shear are based on the observation of the shapes of faint star-forming (SF) galaxies, as it quantifies their coherent distortion by a large-scale foreground matter distribution. Such methods have

been developed for optical surveys so far and require source shapes be measured accurately in order for errors to be dominated by statistics, rather than systematics (Mandelbaum et al. 2015). Therefore, they assume stringent and specific requirements on image fidelity. Radio instruments do not detect images of the observed sky; they provide its Fourier Transform at a finite number of points instead (*visibilities*). The image reconstruction from these data using iterative deconvolution methods such as CLEAN (Högbom 1974; Schwarz 1978) is a highly non-linear process that does not necessarily converge in a well-defined manner when dealing with extended sources resolved by high-resolution telescopes. Moreover, the Fourier Transform of visibilities makes the noise in radio images highly correlated. SKA simulations have shown that even on high signal-to-noise ratio (SNR) objects, this process produces images with structures in the residuals that dominate the cosmological signal, confirming that this bias may be induced by the procedure

* E-mail: m.rivi@ucl.ac.uk (MR); dr.michelle.lochner@gmail.com (ML); ian.harrison-2@manchester.ac.uk (IH)

¹<https://www.skatelescope.org/>

adopted for turning the visibility data into images (Patel et al. 2015). Cross-correlation of real data images shows that there is no evidence of high levels of correlation between the optical and radio intrinsic shape of the matched objects (Patel et al. 2010; Tunbridge, Harrison & Brown 2016), possibly due to systematics of the imaging pipeline. However, a significant percentage of contaminating active galactic nuclei (AGNs) sources (as they carry a much higher shape noise) and the astrophysical scatter between optical and radio position angle due to the different emission mechanisms may also be the reasons for such results. A more natural approach to adopt with radio data is to work directly in the visibility domain, where the noise is purely Gaussian and the data not yet affected by the systematics introduced by the imaging process, with the advantage of an exact modelling of the sampling function. However, this approach faces difficult statistical and computational challenges because sources are no longer localized in the Fourier domain and a telescope like SKA generates a large number of visibilities. Nevertheless, the advantages of being able to reduce, as well as holistically incorporate systematic effects coupled with Gaussian noise properties make modelling in the visibility domain a compelling option.

Bayesian Inference for Radio Observations (BIRO) (Lochner et al. 2015) combines simulation tools, such as *MeqTrees* (Norrdam & Smirnov 2010), capable of predicting visibilities from a given sky model and observational set-up, with Bayesian inference samplers, like Markov-Chain Monte Carlo (MCMC) (Metropolis et al. 1953; Hastings 1970) or Nested Sampling (Skilling 2004). This allows full forward modelling of scientific properties, such as the flux and positions of sources, as well as instrumental effects like pointing errors in telescopes that can have strong effects on the scientific parameters. In addition, by making full use of the visibility information (instead of gridding as imagers do), BIRO can dramatically increase the resolving power of a telescope. BIRO could thus potentially be a useful tool for radio weak lensing, if the galaxies in the survey can be correctly modelled. Other available tools for source fitting in the visibility domain have been developed for general purposes (Martí-Vidal et al. 2014). They provide a variety of galaxy shape models, but they are not sufficiently accurate or capable on wide fields of large numbers of sources for weak lensing measurements.

Currently, two analytical galaxy models have been considered for shear measurement in the visibility domain. The first one uses *shapelets* (Refregier 2003; Refregier & Bacon 2003), where galaxy images are decomposed through an orthonormal basis of functions corresponding to perturbations around a circular Gaussian. Shapelets are invariant under Fourier Transform (up to a rescaling) allowing the adoption in the visibility domain of the same approaches used with images. Finding the best-fitting shape via minimizing the chi-squared is nominally linear in the coefficients and therefore can be performed simultaneously for all sources by simple matrix multiplications. However, this is only the case once a size scale β and number of shapelet coefficients to include have been chosen, which can in itself be a highly non-linear and time-consuming problem. Chang & Refregier (2002) successfully applied this technique to data from the Faint Images of the Radio Sky at Twenty cm (FIRST) survey (Becker, White & Helfand 1995), conducted with the NRAO Very Large Array (VLA). They were also able to detect cosmic shear with a significance of 3.6σ after an accurate treatment of systematic effects (Chang, Refregier & Helfand 2004). However, shapelets introduce a shear bias as they cannot accurately model steep brightness profiles and highly elliptical galaxy shapes (Melchior et al. 2010). For this reason, *Sérsic*

models are commonly used in optical weak lensing surveys analysis (Mandelbaum et al. 2015), where the disc component is described by the exponential brightness profile (*Sérsic* index $n = 1$) and the bulge is approximated by the deVacouleur’s profile (*Sérsic* index $n = 4$). A reasonable assumption is to use a single optical disc-like component in the radio band because the radio-emitting plasma should follow the distribution of stars in galaxy discs, as it is due to the synchrotron radiation of the interstellar medium. The possible bias arising for this model has been discussed for the optical domain in Voigt & Bridle (2010), Miller et al. (2013), and Kacprzak et al. (2014). A similar model bias should be expected in the visibility domain.

In Rivi et al. (2016b) the exponential profile has been Fourier transformed analytically and used to extend the optical semi-Bayesian *lensfit* method (Miller et al. 2013) to radio data in the visibility domain. This method, called *RadioLensfit*, fits the visibilities of a single source at a time and applies a Bayesian marginalization of the likelihood over uninteresting parameters. Visibilities of each galaxy are isolated by a source extraction algorithm described in Rivi & Miller (2018). This approach is very fast computationally but may be limited by the source number density in the field of view, because of nearby galaxies residuals in the extraction procedure (‘neighbour bias’).

We investigate in this paper a different approach for galaxy shape measurement in Fourier space to perform a joint model fitting of all the sources in the field of view. We adopt the same analytical galaxy model used in *RadioLensfit* and introduce it into the BIRO formalism. Unlike in Lochner et al. (2015), we now have a large number of parameters to cope with so we use a Hamiltonian Monte Carlo (HMC) sampler (Neal 2011). We first test the method on two SKA1-MID simulated data, investigating also the potentiality of fitting a large number of sources. Then we apply the method on the simulation of a real observation: SuperCLASS (Battye et al., in preparation), whose UV coverage is quite complex as it is composed of the baselines of two different SKA precursor radio telescopes (e-MERLIN and JVLA). Finally, we discuss the convergence of individual parameters depending on source properties.

Our paper is organized as follows. In Section 2, we outline the method and its implementation. In Sections 3 and 4, we describe data simulation and results for SKA1 and SuperCLASS, respectively. In Section 5, we discuss HMC convergence. We also discuss how to handle AGN contamination in real observations in Section 6. Finally, conclusions are presented in Sections 7.

2 THE METHOD

In this work, we propose to measure radio galaxy shapes directly from the raw visibilities (i.e. in the Fourier space) by sampling the joint posterior distribution of the shape parameters of all the detected galaxies in the field of view. Since the number of galaxies in a single pointing may be very large, conventional MCMC methods using random walk samplers (e.g. Metropolis–Hastings) are not suitable because they require a prohibitive number of samples to explore high-dimensional parameter spaces. We use the HMC approach that reduces random walk behaviour by applying the Hamiltonian dynamics of particles in potential wells (Section 2.1). It requires the computation of the likelihood gradient that we can compute analytically for our purpose, avoiding numerical differentiation (Section 2.2). Mean and standard deviation of the sampled chain provide the measure and related error of the model parameters.

2.1 HMC algorithm

HMC (Neal 2011; Betancourt 2017) is an MCMC method that adopts physical system dynamics to explore a target distribution efficiently, resulting in faster convergence and maintaining a reasonable efficiency even for high-dimensional problems (Hanson 2001; Taylor, Ashdown & Hobson 2008).

The basic idea is to sample a distribution $P(\mathbf{x})$ of n parameters \mathbf{x} according to the Hamiltonian of a $2n$ -dimensional dynamical system, where the parameters correspond to the positions of n particles and $P(\mathbf{x})$ is the canonical distribution of the potential energy function $U(\mathbf{x})$. Momentum variables \mathbf{p} are introduced to be coupled with positions \mathbf{x} to allow Hamiltonian dynamics to operate according to a kinetic energy function $K(\mathbf{p})$ whose canonical distribution is the zero-mean Gaussian with a diagonal covariance matrix $\mathbf{M} = \text{diag}(m_1, \dots, m_n)$, where m_i represents the mass of particle i . This means:

$$U(\mathbf{x}) = -\log[P(\mathbf{x})], \quad K(\mathbf{p}) = \frac{1}{2} \mathbf{p}^T \mathbf{M}^{-1} \mathbf{p}. \quad (1)$$

Samples from the parameters distribution are obtained by marginalizing the canonical distribution of the corresponding Hamiltonian function $H(\mathbf{x}, \mathbf{p}) = U(\mathbf{x}) + K(\mathbf{p})$:

$$\exp[-H(\mathbf{x}, \mathbf{p})] = P(\mathbf{x}) \exp\left[-\frac{1}{2} \mathbf{p}^T \mathbf{M}^{-1} \mathbf{p}\right] \quad (2)$$

over \mathbf{p} .

In practice, at each iteration of HMC a new sample \mathbf{x}' is generated from the previous one \mathbf{x}_0 as follows. First a proposal \mathbf{p}_0 for the momentum variables is generated randomly according to the canonical distribution of the kinetic energy, i.e. an n -dimensional uncorrelated Gaussian. Then, a deterministic proposal for positions \mathbf{x} is computed by allowing the system to evolve for a fixed time τ from the starting point $(\mathbf{x}_0, \mathbf{p}_0)$ according to Hamilton's equations:

$$\frac{dx_i}{dt} = \frac{\partial H}{\partial p_i} = [\mathbf{M}^{-1} \mathbf{p}]_i \quad (3)$$

$$\frac{dp_i}{dt} = -\frac{\partial H}{\partial x_i} = -\frac{\partial U(\mathbf{x})}{\partial x_i}. \quad (4)$$

These equations are usually solved numerically by applying the leapfrog method, where Hamilton's equations are discretized with a time step of small size ε as follows:

$$p_i\left(t + \frac{\varepsilon}{2}\right) = p_i(t) - \frac{\varepsilon}{2} \frac{\partial U}{\partial x_i} \Big|_{\mathbf{x}(t)} \quad (5)$$

$$x_i(t + \varepsilon) = x_i(t) + \frac{\varepsilon}{m_i} p_i\left(t + \frac{\varepsilon}{2}\right) \quad (6)$$

$$p_i(t + \varepsilon) = p_i\left(t + \frac{\varepsilon}{2}\right) - \frac{\varepsilon}{2} \frac{\partial U}{\partial x_i} \Big|_{\mathbf{x}(t+\varepsilon)}. \quad (7)$$

The new proposal for \mathbf{x} is obtained after k steps as $\mathbf{x}' = \mathbf{x}(\tau)$, with $\tau = k\varepsilon$. It is accepted with Metropolis criterion, i.e. with probability:

$$\min\{1, \exp(-\delta H)\}, \quad (8)$$

where

$$\delta H = H(\mathbf{x}', \mathbf{p}') - H(\mathbf{x}_0, \mathbf{p}_0).$$

This way the rate of acceptance is close to unity when numerical solutions of Hamilton's equations are very close to the exact ones because along such a trajectory energy is conserved. Neal (2011) shows that $\varepsilon < 2$ must be required in order to have stable trajectories

computed with the leapfrog method. Our C++ implementation of the HMC algorithm will soon be available within a BIRO repository.

2.2 Likelihood and gradient computation

In order to sample the posterior of the parameters, we need to define the likelihood function obtained by comparing data visibilities \tilde{V}_i with the model visibilities V_i dependent on parameters \mathbf{x} :

$$\mathcal{L}(\mathbf{x}) = \left(\prod_i^N \frac{1}{\sqrt{2\pi\sigma_i^2}} \right) \exp\left[-\sum_i \frac{|\tilde{V}_i - V_i(\mathbf{x})|^2}{2\sigma_i^2}\right], \quad (9)$$

where σ_i is the standard deviation of the Gaussian noise on the i -th data visibility.

Visibilities are evaluated at the interferometer baseline vectors, whose coordinates (u, v, w) are measured in wavelengths at the centre frequency of the signal band with respect to a coordinate system where w points towards the phase tracking centre (Thompson, Moran & Swenson 1986). For a sky containing N galaxies with exponential profile, we adopt the *RadioLensfit* analytical visibility model (Rivi & Miller 2018):

$$V(u, v, w) = \left(\frac{\lambda_{\text{ref}}}{\lambda}\right)^\beta \sum_{s=1}^N \frac{S_{s, \lambda_{\text{ref}}} e^{2\pi i (ul_s + vm_s + w\sqrt{1-l_s^2 - m_s^2})}}{(1 + 4\pi^2 \alpha_s^2 |\mathbf{A}_s^{-T} \mathbf{k}|^2)^{3/2}}, \quad (10)$$

where $\beta = -0.7$ is the spectral index for the synchrotron radiation emitted by the galaxy disc, $\mathbf{k} = (u, v)$, and for $s = 1, \dots, N$ we have the following source s parameters: position coordinates (l_s, m_s) , flux $S_{s, \lambda_{\text{ref}}}$ at reference wavelength λ_{ref} , scale length α_s , and ellipticity components $e_{1, s}, e_{2, s}$. The ellipticity parameters are contained in the matrix \mathbf{A}_s that transforms the circular exponential profile² in elliptical

$$\mathbf{A}_s = \begin{pmatrix} 1 - e_{1, s} & -e_{2, s} \\ -e_{2, s} & 1 + e_{1, s} \end{pmatrix}. \quad (11)$$

We assume the following ellipticity definition:

$$\mathbf{e} = e_1 + ie_2 = \frac{a - b}{a + b} e^{2i\theta}, \quad (12)$$

where a and b are the galaxy major and minor axes, respectively, and θ is the galaxy orientation. Note that in case future radio observations show that a two-component galaxy model is also required in the radio domain, a similar analytical model for the bulge in the Fourier domain approximating the deVacouleur's profile is provided in Spergel (2010).

A preliminary data analysis for measuring flux and position of each source may allow the reduction of the number of parameters to $3N$. This can be achieved either in the image domain³ or, if necessary, in the visibility domain by a multimodal nested sampling approach adopting a single source model (Malyali et al., in preparation). On the other hand, including these parameters in the full HMC analysis will require a tight prior on the positions in order to have a small additional computational complexity. In this work we assume them well known.

Since the visibility model is defined analytically we can compute the exact likelihood gradient with respect to the parameters of our interest: scale length and ellipticity components of each source in the primary beam. The computation of the galaxy exponential

² $J(r) = I_0 \exp(-r/\alpha)$

³Based on SKA1 level 0 science requirements, SKA1-MID shall provide astrometric accuracy of at least 1 per cent.

model and the corresponding likelihood gradient (see Appendix A) has been added to the open-source software MONTBLANC⁴ (Perkins et al. 2015) that is a GPU-accelerated implementation of the Radio Interferometric Measurement Equation (RIME) (Hamaker, Bregman & Sault 1996; Smirnov 2011a,b) in support of BIRO. The RIME is a powerful framework to easily describe what happens to a signal as it travels from source to the interferometer in terms of Jones matrices, where multiple effects along the signal propagation path correspond to multiplication of matrices describing each effect. For example, equation (10) can be rearranged as follows:

$$V_{tpq\lambda} = \sum_{s=1}^N V_{tpq\lambda s} = \sum_{s=1}^N K_{tps\lambda} B_{s\lambda} K_{tqs\lambda}^H, \quad (13)$$

where uvw points are identified by the baseline formed by antennas p and q at time step t and channel with centre wavelength λ . $K_{tps\lambda}$ and $B_{s\lambda}$ are, respectively, the phase and brightness matrices for source s (H denotes matrix hermitian transpose). We use a brightness matrix defined by the exponential model of an extended source⁵:

$$B_{s\lambda} = \left(\frac{\lambda_{\text{ref}}}{\lambda}\right)^\beta \frac{S_{s,\lambda_{\text{ref}}}}{(1 + 4\pi^2 \alpha_s^2 |\mathbf{A}_s^{-T} \mathbf{k}|^2)^{3/2}}. \quad (14)$$

Given the general formalism of the RIME, any instrumental effects that can be modelled, such as primary beam shape, pointing errors and instrumental noise, can be incorporated and marginalized over with our formalism.

2.3 Priors

As source scale length and ellipticity priors we use the distributions presented in Rivi & Miller (2018), whose parameters are estimated from VLA observations:

- (i) $p(\alpha)$ is a lognormal function with mean $\mu = \ln(\alpha_{\text{med}})$ and variance $\sigma \sim 0.3$, where α_{med} is given by equation (16);
- (ii) for the modulus e of the intrinsic galaxy ellipticity

$$p(e) = \frac{Ne \left(1 - \exp\left[-\frac{e - e_{\text{max}}}{c}\right]\right)}{(1 + e)(e^2 + e_0^2)^{1/2}}, \quad (15)$$

with $c = 0.2298$, $e_0 = 0.0732$, and $e_{\text{max}} = 0.804$.

These distributions are used for both the galaxy catalogue simulation and the prior in the HMC sampling.

2.4 HMC tuning

Selecting suitable values of the step size ε and the number of steps k for the leapfrog method is crucial as discussed in Neal (2011). After preliminary runs, we choose $\varepsilon = 0.05$ and $k = 10$ as a trade-off between the accuracy of the Hamilton's equations solutions (i.e. high acceptance rate) and the computational time due to a too long trajectory.

Since most of galaxy ellipticities should be close to zero, according to the ellipticity modulus distribution, we choose as starting points for the ellipticity components random numbers uniformly distributed between -0.1 and 0.1 . For scale length parameters we choose as starting points the median values depending on the source

flux density according to the relation estimated in Rivi et al. (2016a) from the VLA 20 cm continuum radio source catalogue in the SWIRE⁶ field (Owen & Morrison 2008):

$$\ln[\alpha_{\text{med}}/\text{arcsec}] = -0.93 + 0.33 \ln[S/\mu\text{Jy}]. \quad (16)$$

For the convergence of HMC chains, it is very important to choose good ‘step sizes’ of the momentum samples. They are defined by the inverse mass in the kinetic energy function and are typically given by the variance of the parameters. In this application parameter variance is dependent on the source SNR, and therefore integrated source flux. We estimate a relation between the standard deviation for source scale length and ellipticity components as a function of its flux density by fitting the error bars measured from a simulated observation of 100 sources (as described in Section 3). For both ellipticity components and scale length we obtain a power-law relation between source flux (in μJy) and measured standard deviation. For example, for SKA1-MID observations we obtain the following relations (see Fig. 1):

$$\sigma_{e_1}, \sigma_{e_2} = -0.0022 + 1.77S^{-0.92}, \quad (17)$$

$$\sigma_\alpha[\text{arcsec}] = -0.0914 + 0.61S^{-0.31}. \quad (18)$$

3 SKA1-MID SIMULATIONS

We simulate SKA1-MID 8 h observations pointing close to the zenith ($\delta \sim -30^\circ$). Visibilities are sampled every 60 s at the frequency 1.4 GHz (the corresponding UV coverage is plotted in Fig. 2) and we choose a single large channel with bandwidth 240 MHz, as no smearing effects are added. Visibilities are simulated using equation (10) and adding an uncorrelated Gaussian noise whose variance is dependent on the antenna system equivalent flux density of SKA1 dishes, the frequency channel bandwidth, and the time sampling according to the formula given in Wrobel & Walker (1999).

3.1 Source catalogues

The source catalogues are simulated as in Rivi & Miller (2018), where positions are generated randomly according to a uniform distribution over a circular field of view, the flux distribution is $p(S) \propto S^{-1.34}$, scale length and ellipticity follow the prior distributions listed in Section 2.3.

The minimum integrated source flux detectable at a $\text{SNR}_{\text{vis}}^7 = 10$ is $10 \mu\text{Jy}$, and the source density expected in the current planned SKA1 radio weak lensing survey ($2.7 \text{ gal arcmin}^{-2}$, Brown et al. 2015) is obtained at the maximum flux of $200 \mu\text{Jy}$. We choose a field of view of 400 arcmin^2 to test the method at such density for 1000 galaxies (‘SKA1 1000’). To compare measurement accuracy and the speed of convergence as a function of the number of sources to fit, we also test the case of 100 sources with the same flux range and field of view (‘SKA1 100’).

3.2 Results

In Fig. 3, we show the one-dimensional marginalized posterior for a few example parameters of ‘SKA1 100’ showing they are

⁴https://github.com/ska-sa/montblanc/tree/chi_sqrd_gradient

⁵Other galaxy models available in Montblanc are Gaussian and point sources. For Gaussians a simpler computation of the likelihood gradient can be added.

⁶<http://heasarc.gsfc.nasa.gov/W3Browse/radio-catalog/vlasdf20cm.html>

⁷We compute the signal-to-noise ratio in the visibility domain as $\text{SNR}_{\text{vis}} = \sqrt{\sum_{i=1}^{\text{nv}_{\text{vis}}} |V_i|^2 / \sigma_i^2}$, where V_i are the simulated visibilities without noise.

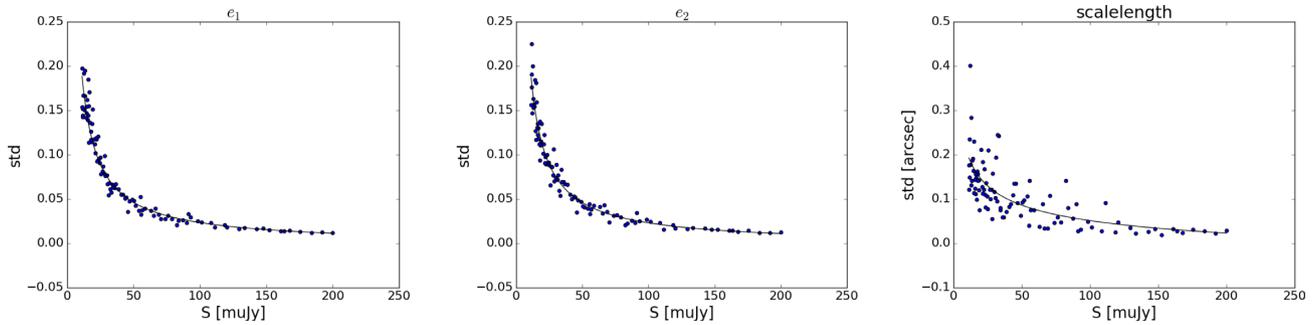


Figure 1. Power-law function fitting the standard deviation of the parameters versus integrated source flux from the SKA1-MID measurements of 100 sources with $\text{SNR}_{\text{vis}} \geq 10$.

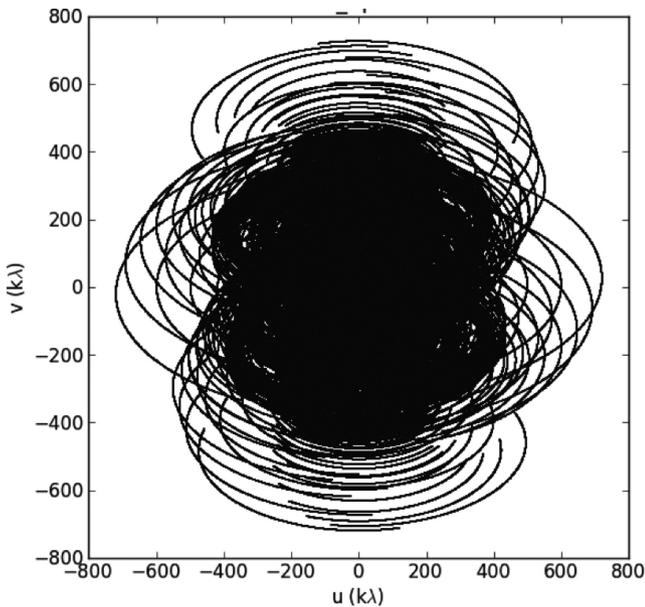


Figure 2. UV coverage of SKA1 observations.

well constrained and accurately recover input parameters. We show the difference between the measured and input parameters (which should be close to zero) for both SKA1 cases in Fig. 4. As expected, brighter objects have much lower uncertainty in the parameters. We find that the shape parameters are largely uncorrelated although they could be highly correlated with instrumental effects that we do not model here (Lochner et al. 2015).

Table 1 contains the values of the best-fitting lines coefficients for the scale length and the ellipticity components obtained by fitting the simulated observations of 100 and 1000 sources in the field of view, respectively. They are almost the same for both test cases, where only the uncertainty is obviously larger for a smaller number of sources. The goodness of fit for all shape parameters, estimated as a straight line fit of measured versus true parameters, shows that the measurements are very precise.

Comparing the best-fitting slope of ellipticities for 1000 sources, i.e. at the source density of the proposed SKA1 weak lensing survey, with the one obtained at the same density with *RadioLensfit* we see a significant improvement. In fact, in Rivi & Miller (2018) the best-fitting slopes of the ellipticity components, measured from a similar source population and SKA1 UV coverage, were found to be $a_1 = 0.9365 \pm 0.0017$ and $a_2 = 0.9262 \pm 0.0017$, respectively. This improvement is expected because joint fitting avoids the

source extraction bias of the *RadioLensfit* method, but with a large computational cost (see Table 3). We also observe that ellipticity error bars of our approach are larger, but correctly measured, mainly for two reasons: (i) *RadioLensfit* may underestimate the parameters uncertainty by sampling the likelihood only in a neighbourhood of the maximum point (see section 3.3 of Rivi et al. 2016b), while we marginalize the posterior instead; (ii) in this work, we have a larger number of free parameters as we are measuring both ellipticity and scale length of all sources simultaneously. Thus, the larger error bars are actually an indication of the technique correctly incorporating additional sources of uncertainty.

4 APPLICATION TO SUPERCLASS

We also investigate the ability of the method to infer shear using current-generation data, from the SuperCLASS survey. This allows us to both assess the possibilities in near-term data, and also understand the effect of realistic UV coverages, which are not typically observed as a single long track, rather multiple interleaved pointings. We emphasize that the results we generate here should not be used as a measure of the relative shape measurement ability of SuperCLASS and an SKA survey; shape measurement ability will be a strong function of UV coverage and we use a very realistic coverage for SuperCLASS and a highly simplified (and much smaller in volume) coverage for SKA.

4.1 The survey

The SuperCLASS (Super CLuster Assisted Shear Survey)⁸ (Battye et al, in preparation) is a legacy survey on the UK’s e-MERLIN radio telescope, with the express goal of making a convincing detection of a weak lensing signal in radio data. For the field, containing four massive galaxy clusters at $z \sim 0.2$, observations are taken at L band (around 1.4 GHz) using both e-MERLIN and the Jansky Very Large Array (JVLA). This gives coverage of a wide range of spatial scales, with sensitivity to angular scales between 1 and 10 arcsec principally coming from the JVLA, and smaller scales from the longer baselines available to e-MERLIN. The field has also been observed at multiple lower radio frequencies (to assist in Rotation Measure synthesis and source classification), sub-mm, and optical and near-IR wavelengths (to obtain photometric redshifts and optical shear measurements). The full data release of the survey is expected to contain thousands of galaxies at a density of ~ 1 gal arcmin⁻² over 1 deg². Shape

⁸<http://www.e-merlin.ac.uk/legacy/projects/superclass.html>

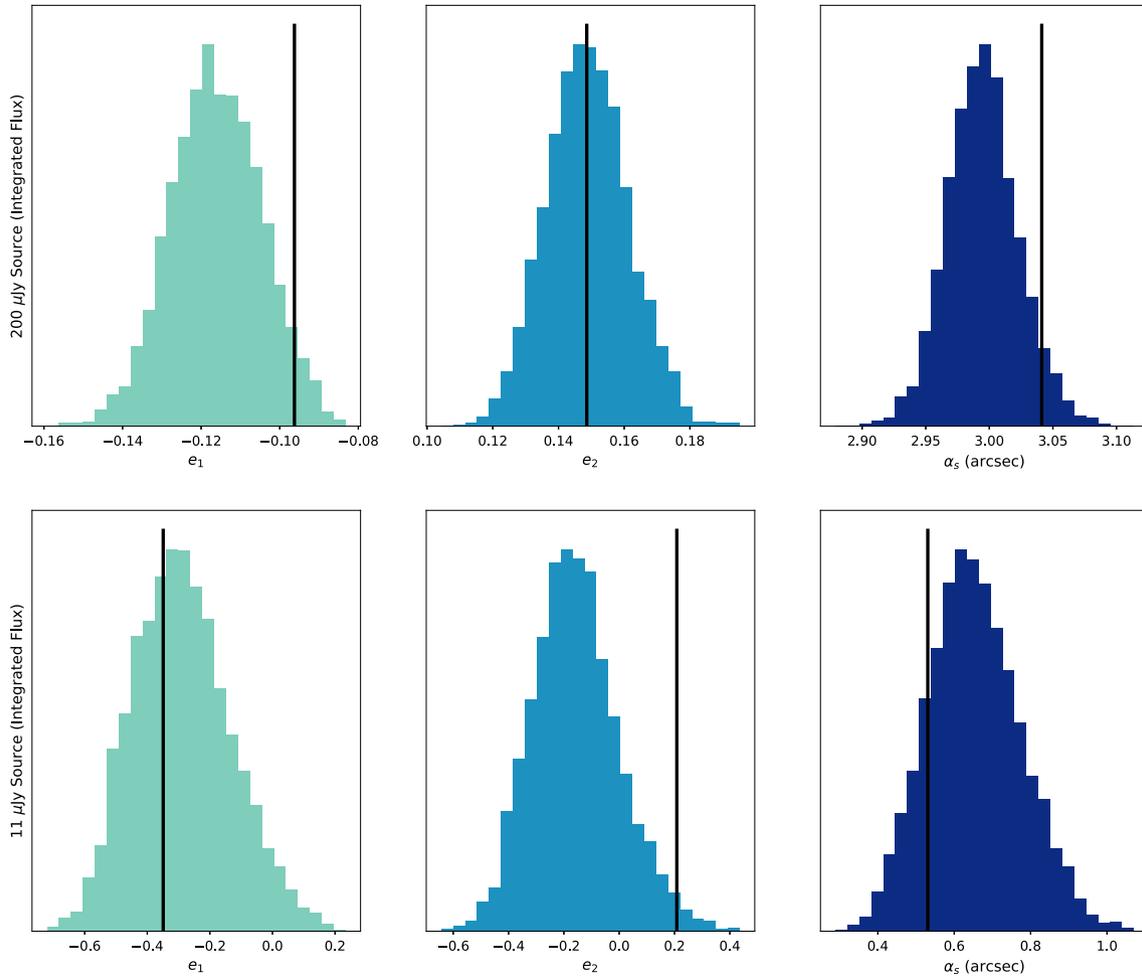


Figure 3. Example 1D marginalized posterior plots for the SKA1 100 source case, for the brightest (top panel) and the faintest (bottom panel) sources in the catalogue. The input value for each parameter is shown by a solid line, illustrating the ability of our approach to accurately recover both the parameter value and its uncertainty.

measurement for the survey is being performed using both calibrated image-plane and hybrid image/visibility-plane techniques.

4.2 Data simulation

We create simulated UV coverages that closely match the true ones in the SuperCLASS survey. This provides us with an idealized but realistic version of the experiment; though we do not include effects from calibration errors and missing data from telescope outages or Radio Frequency Interference (RFI, an appreciable problem for e-MERLIN, Peck & Fenech 2013), we do recreate the real shape of the UV coverage (corresponding to the shape of the PSF). The UV coverages for both telescopes are simulated as measurement sets using the CASA simulator tool (McMullin et al. 2007).

4.2.1 e-MERLIN

For e-MERLIN, we simulate eight IFs of 512 channels of width 125 kHz, covering a bandwidth of 512 MHz upwards from a starting frequency of 1.25 GHz. The UV coverage is then generated as it is for a real observation: during a single 8 h observing run (one epoch), seven different pointing centres are observed, along with amplitude

and phase calibration sources. Observations cycle around the seven pointing centres, spending alternately 12.5 min observing the source and 2.5 min observing calibrators. An individual pointing centre is observed over four epochs, giving a total amount of time on source of just over 8 h per pointing centre. This gives an image plane noise level of $\sim 14 \mu\text{Jy beam}^{-1}$ before different pointings are mosaiced together, a process which reduces the noise by a factor of 2. The UV coverage resulting from this procedure is shown in the left-hand panel of Fig. 5.

4.2.2 JVLA

For the JVLA telescope observations of the SuperCLASS field, we simulate the full 16 IFs of 256 channels each, with channel widths of 250 kHz, covering a bandwidth of 1024 MHz upwards from a starting frequency of 1.01 GHz. For the JVLA, each individual pointing is observed over six epochs, with a total time of 11 min 6 s on source. The greater filling of the Fourier plane on the sample scales by the JVLA when compared to e-MERLIN means that a similar image-plane depth of $\sim 14 \mu\text{Jy beam}^{-1}$ is reached in this time. The UV coverage resulting from this procedure is shown in the right-hand panel of Fig. 5.

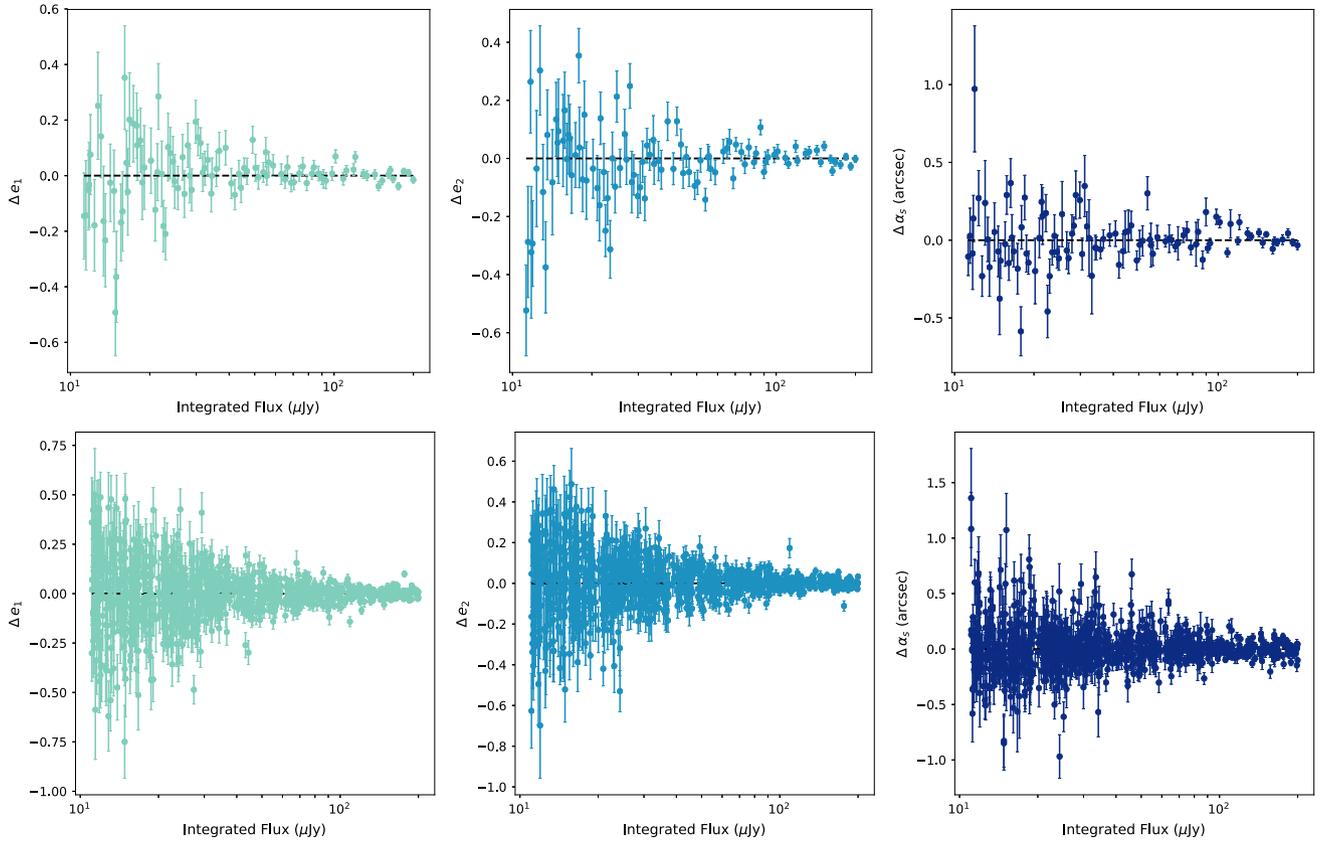


Figure 4. The measured minus true value of each parameter as a function of integrated flux for the ellipticity parameters (e_1 and e_2) and the scale length parameter (α_s), for SKA1 100 sources (top panel) and SKA1 1000 sources (bottom panel). The error bar is estimated from the 1D marginalized posterior for that parameter. It is clear that brighter sources have lower uncertainty on the parameters.

Table 1. Multiplicative (a) and additive (c) coefficients of the best-fitting lines for galaxy shape measurements with SKA1 at $\text{SNR}_{\text{vis}} \geq 10$.

Observation	e_1		e_2		scale length	
	a	c	a	c	a	c
SKA1 100	0.9559 ± 0.0116	-0.0018 ± 0.0032	0.9780 ± 0.0118	-0.0014 ± 0.0031	1.0015 ± 0.0091	-0.0019 ± 0.0161
SKA1 1000	0.9704 ± 0.0043	0.0001 ± 0.0010	0.9718 ± 0.0040	-0.0002 ± 0.0010	1.0048 ± 0.0030	-0.0090 ± 0.0051

4.2.3 Visibilities

Because of restrictions on available memory we sample visibilities every 10 s instead of the real sampling time of 1 s. No time or bandwidth smearing effects are accounted for in the forward modelling. As in Section 3, we simulate visibilities observed in a single pointing according to equation (10) and add uncorrelated Gaussian noise with variance dependent on the telescope configuration. In order to have the source SNR distribution expected for the real data, we artificially decrease this noise by a factor of 2 (as the real data will benefit from the mosaicing effect of overlapping pointings), considering only the pointing inner region of 100 arcmin² as field of view. The source number density of 1 gal arcmin⁻² at 10σ in the image domain⁹ is obtained for a minimum input integrated source flux $S_{\text{min}} \sim 500 \mu\text{Jy}$ and, according to our flux prior, a maximum flux

density of 400 mJy. Note that the corresponding minimum source SNR in the visibility domain is about 18.

4.3 Shape measurements

We are able to recover with a good accuracy the shape parameters of all sources. Fig. 6 shows the difference between the measured and input galaxy shape parameters, and Table 2 contains the best-fitting lines coefficients of the measured parameter values. The goodness of fit is better than for SKA1 simulations mainly because in this case the SNR is higher ($\text{SNR}_{\text{vis}} \geq 18$), but also because the UV coverage is much larger (see Table 3).

5 CONVERGENCE ANALYSIS

5.1 Tests for convergence

Any Bayesian inference sampler, such as HMC, is only guaranteed to converge to the true posterior with an infinite number of steps. Given that we obviously can only run a chain for a finite

⁹Usually the source signal-to-noise in the image domain is given by the ratio between the source peak intensity and the image noise standard deviation. Since we are dealing with extended sources, this SNR tends to be lower than what measured in the visibility domain.

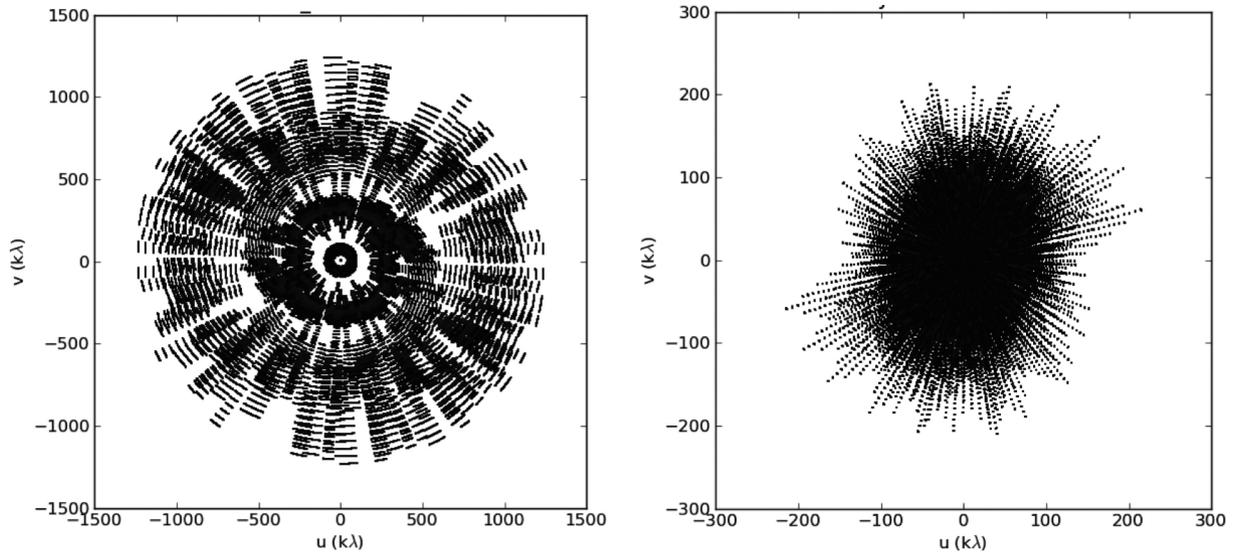


Figure 5. SuperCLASS UV coverage. Left: e-MERLIN. Right: JVLA (note the difference in scales).

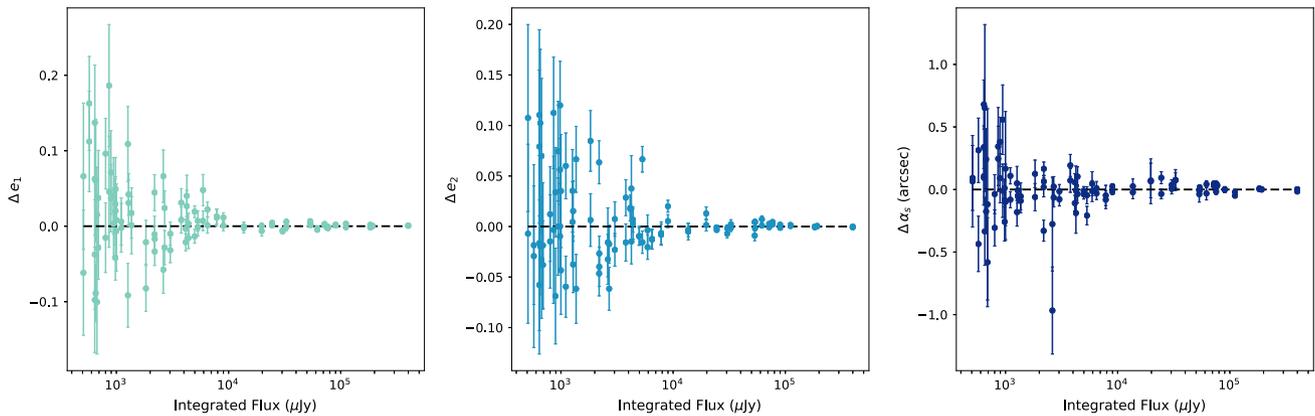


Figure 6. The measured minus true value of each parameter as a function of integrated flux for the ellipticity parameters (e_1 and e_2) and the scale length parameter (α_s) for SuperCLASS with 100 sources.

Table 2. Multiplicative (a) and additive (c) coefficients of the best-fitting lines for galaxy shape measurements with SuperCLASS at $\text{SNR}_{\text{vis}} \geq 18$ (the lowest SNR in the SuperCLASS sample).

Observation	e_1		e_2			scale length	
	a	c	a	c	a	c	
SuperCLASS-100	1.0001 ± 0.0010	0.0006 ± 0.0002	1.0011 ± 0.0020	0.0004 ± 0.0002	1.0002 ± 0.0005	-0.0047 ± 0.0079	

time, convergence must be tested for. We test for convergence of the HMC chains using two different metrics: the commonly-used Gelman–Rubin (GR) statistic (Gelman & Rubin 1992) and the autocorrelation function (Sokal 1997).

5.1.1 The Gelman–Rubin statistic

The GR is a measure of how similar a set of HMC chains are: it compares the variance within a chain to the variance between chains and should be close to one for a converged set of chains. Ideally, one would run several independent chains, starting from

different starting points. However, as we are computation limited in this work, we settle for running one chain that we split up into three separate chains and compare.

For the SKA1 100 source chain and the SuperCLASS 100 source chain, we find $\text{GR} \approx 1.1$ for all parameters. However, for the SKA1 1000 source chain, the parameters converge more slowly and for a few parameters, $\text{GR} \approx 2$. While we still use this chain in our analysis, because we observe no strong bias with respect to the true parameters and find the effective sample size to be sufficient (see below), if used on real data a longer chain should ideally be run and convergence ensured.

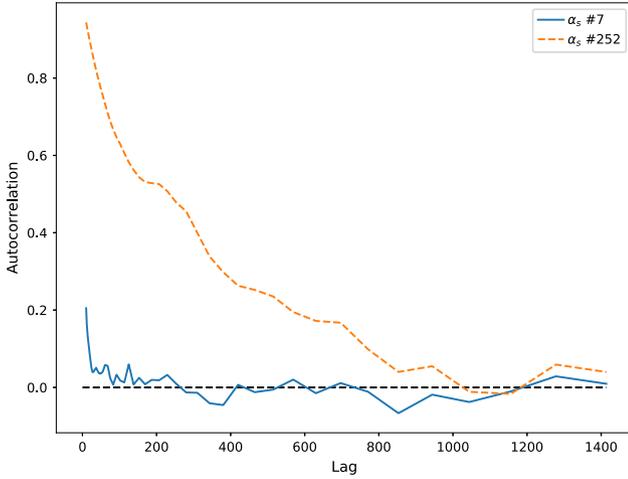


Figure 7. Autocorrelation curves for two example parameters of the SKA1 100 sources chain. When the autocorrelation is close to zero, steps separated by the corresponding lag are uncorrelated, independent samples. A short autocorrelation length corresponds to a well-mixed chain. Plotted here are the autocorrelation functions for the scale length of a bright, round source (solid, blue line) and faint, highly elliptical source (dashed, orange line), for which the steps of the chain are still highly correlated indicating a difficulty to converge.

5.1.2 Auto-correlation function

An additional test for convergence is the lag- k autocorrelation function that computes the autocorrelation between posterior samples separated by k steps. The autocorrelation function for parameter θ , with mean $\bar{\theta}$, sampled by a length- N chain is given by:

$$\rho_k = \frac{\sum_i^{N-k} (\theta_i - \bar{\theta})(\theta_{i+k} - \bar{\theta})}{\sum_i^{N-k} (\theta_i - \bar{\theta})^2}. \quad (19)$$

$\rho_k = 0$ indicates samples separated by k are independent. Fig. 7 shows examples lag- k autocorrelation functions for some parameters of the ‘SKA1 100’ chain. Samples in a chain from HMC are, in general, not statistically independent. The autocorrelation function is a measure of over what ‘time-scale’ steps are correlated. The faster the autocorrelation function drops to zero, the more independent samples can be drawn and the shorter the chain can be to achieve precision on parameter estimates. For a chain of a given length, this has direct impact on the error with which a parameter is measured.

If the uncertainty on a given parameter as measured from the posterior is σ , then the *sampling uncertainty* σ_{samp} is given by (Sokal 1997):

$$\sigma_{\text{samp}}^2 = \frac{\sigma^2 t_{\text{int}}}{N}, \quad (20)$$

where N is the length of the chain. t_{int} is the *integrated autocorrelation time* and is defined as:

$$t_{\text{int}} = 1 + 2 \sum_{k=0}^N \hat{\rho}_k, \quad (21)$$

where $\hat{\rho}_k$ is the normalized autocorrelation function ρ_k/ρ_0 . σ_{samp}^2 is the variance on the estimate of a given parameter which is due to finite sampling and is also sometimes referred to as the Monte Carlo standard error.

The integrated autocorrelation time is thus a key metric for analysing the convergence of a chain in an intuitive way: t_{int} dic-

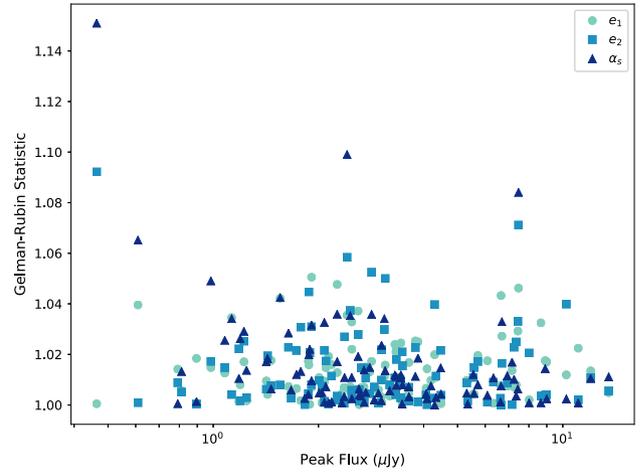


Figure 8. The Gelman–Rubin statistic as a function of peak flux of the source for all three parameters for the SKA1 100 sources chain. This indicates the chain is sufficiently converged and also that there may be a loose relationship between how easily a chain converges (i.e. a lower GR value) and peak flux. This relationship is further investigated using the integrated autocorrelation time in Fig. 9.

tates the number of steps required before a chain will reach a given level of precision in the estimate of a parameter, since it defines the number of independent samples that are there in a chain. For instance, an *effective sample size* (Nt_{int}) of 100 would contribute an additional ~ 10 per cent to the uncertainty on a given parameter. As the effective sample size increases, the contribution to the error budget from sampling becomes negligible when compared with the intrinsic posterior uncertainty.

In an ideal world, an HMC chain would always be run long enough for the sampling uncertainty to be too small to be concerned with. However, in reality computational resources are limited and the sampling uncertainty must be taken into account. The required level of precision depends on the final analysis to be done with the sample. MCMC analyses have been done in the context of shear estimation in the case of the Dark Energy Survey (Jarvis et al. 2016) where chains of a few thousand steps were found to be sufficient. Estimates of the required sampling uncertainty needed would critically depend on the design and cosmological requirements of that survey, which is beyond the scope of this paper. So, rather than enforcing a particular precision, in this work we use t_{int} and σ_{samp} to investigate which parameters are difficult to constrain and gain some physical insight in the problem.

5.2 Convergence of individual parameters

Given that some parameters converge more quickly than others, it is natural to ask what properties about a source make it easy or difficult to constrain its parameters. We investigate this here using the integrated autocorrelation time (see equation 21) and the sampling uncertainty (equation 20).

We first test the hypothesis that brighter sources should be easier to constrain. Fig. 8 shows the relationship between the GR statistic and the peak flux for the SKA1 100 sources case. There appears to be some relationship that brighter sources tend to have ‘more converged’ parameters but it’s difficult to be certain. We investigate this further with the integrated autocorrelation time. As seen in Fig. 9, the integrated autocorrelation time only correlates strongly with flux in the SuperCLASS case where the sources are at a

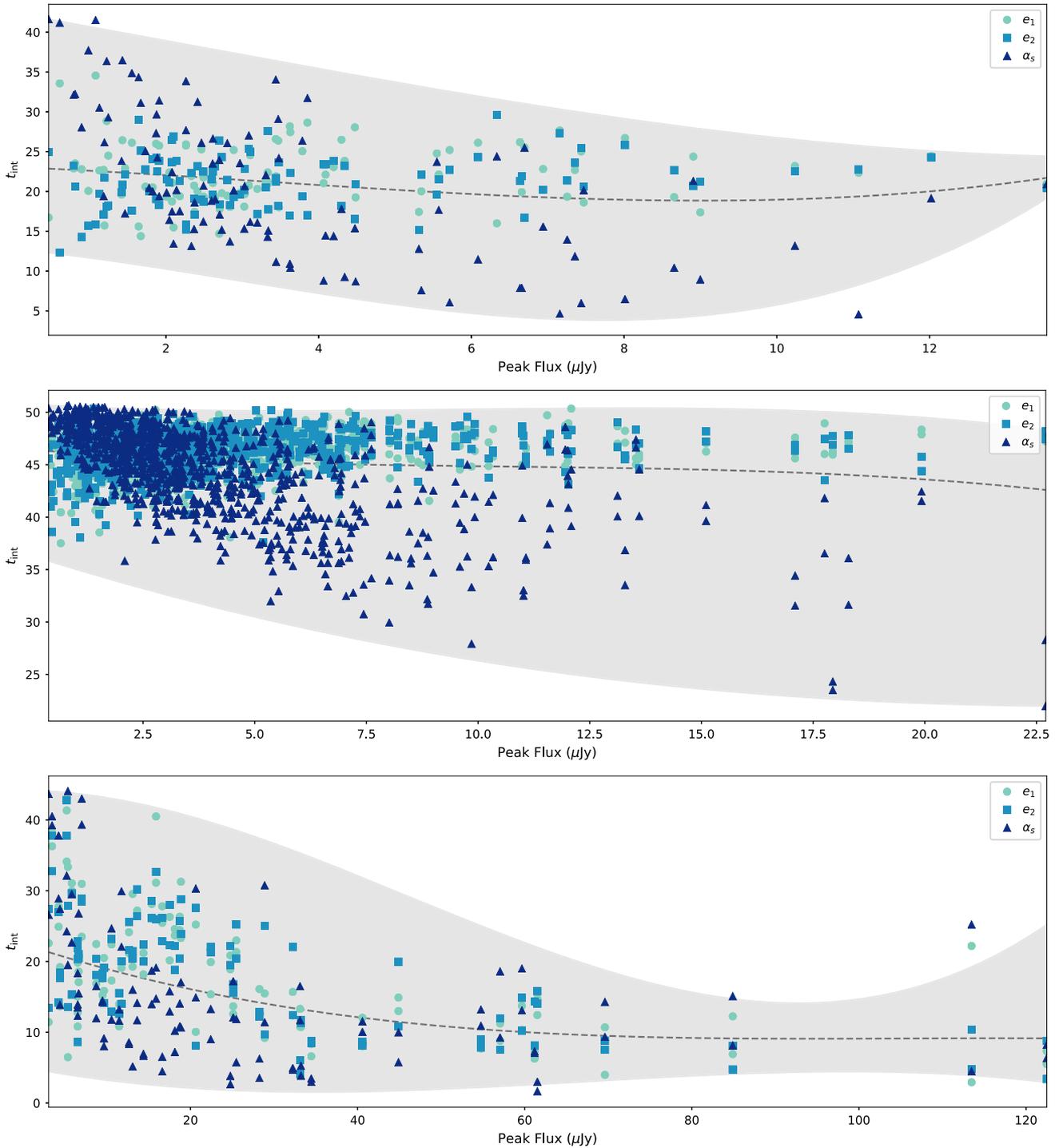


Figure 9. The integrated autocorrelation time as a function of peak flux of the galaxy for all three parameters, for SKA1 100 sources (top panel), SKA1 1000 sources (middle panel), and SuperCLASS 100 sources (bottom panel). The smaller this value, the more statistically independent samples can be obtained from a chain. In each plot, lines are added to guide the eye that indicate the mean (dashed line) and minimum/ maximum (shaded envelope) in broad bins, interpolated with a cubic spline. For each case, it is clear that the scale length of brighter sources is easier to constrain, while the effect on ellipticity is more subtle. The SuperCLASS results indicate that very high SNR sources are significantly easier to constrain.

significantly higher signal-to-noise range. For the SKA1 case, there is a clear indication that the scale length is easier to constrain for brighter sources, but interestingly there does not seem to be a strong relationship between t_{int} and peak flux for the ellipticity parameters. This is not to say the uncertainty on the measured ellipticity parameters does not decrease with flux (see Figs 4 and 6) but the ellipticity

parameters of sources with a higher peak flux are not more easy to measure than faint ones. We do, however, see a strong correlation with source size in Fig. 10.

We also consider the sampling uncertainty, which is the uncertainty on the estimate of the mean of a parameter, due to having a finite number of samples in the posterior. Fig. 10 shows the plots for

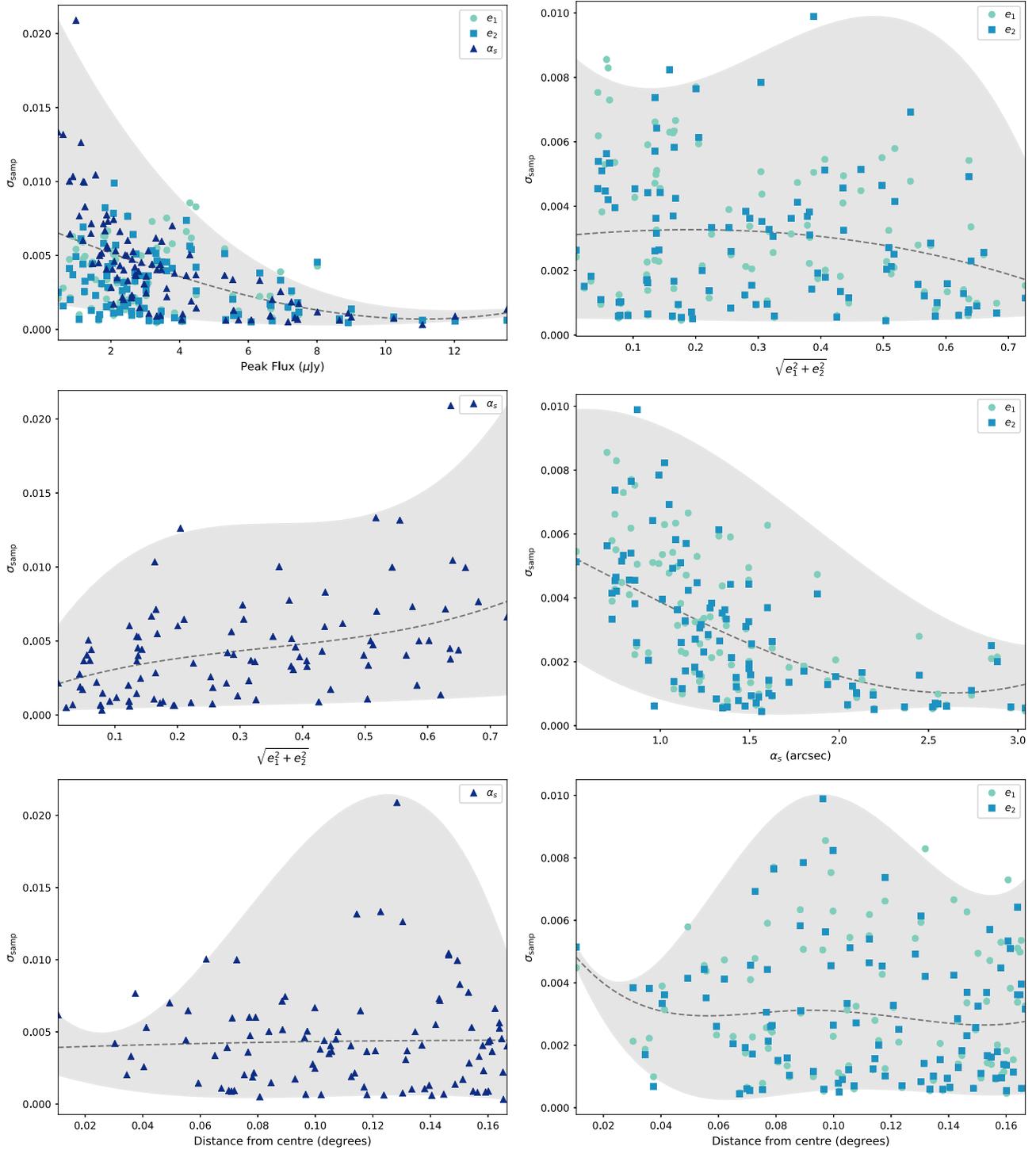


Figure 10. The sampling uncertainty (that is the contribution to the error on a parameter measurement due to having a chain of finite independent samples) on each of the parameters as a function of different quantities. *Top left:* All parameters as a function of flux. *Top right:* The ellipticity parameters as a function of the amplitude of \mathbf{e} . *Middle left:* The scale length as a function of the distance of the source from phase centre. *Middle right:* The ellipticity parameters as a function of scale length. *Bottom left:* The scale length as a function of the distance of the source from phase centre. *Bottom right:* The ellipticity parameters as a function of the distance of the source from phase centre. In each plot, lines are added to guide the eye that indicate the mean (dashed line) and minimum/ maximum (shaded envelope) in broad bins, interpolated with a cubic spline. All plots use the SKA1 100 sources chain, although the results are similar for the others. The conclusion from this analysis is that the easiest sources to constrain the parameters tend to be sources that are bright and large. The scale length is more precisely estimated for round sources, although the corresponding ellipticity uncertainty may be a bit larger. The maximum sampling uncertainty increases with distance from the phase centre, although there is significant scatter.

Table 3. A summary of the description of each experiment and the corresponding computational cost to reach convergence (i.e. $GR \approx 1$ for all parameters) in terms of likelihood samples and computational time using 1 CPU core and 2 NVIDIA Tesla K40. ‘Burn-in’ refers to the number of samples removed in the initial part of the chain as it finds its way to the best-fitting parameter values. ‘n. samples total’ refers to the total length of the chain before burn-in is removed.

Observation	n. visibilities	field of view [arcmin ²]	n. sources	n.samples burn-in	n. samples total	CPU time per sample [min]
SKA1	9266 880	400	100	250	15 100	0.36
SKA1	9266 880	400	1000	1000	20 900	4.64
SuperCLASS	345 907 200	100	100	250	7800	8.47

‘SKA1 100’. As expected, the uncertainty decreases with increased flux, even though there is only a subtle relationship between the effective number of samples and flux as shown in Fig. 9. We also find that the sampling error on the ellipticity parameters only has a weak dependence on the actual value of the ellipticity. What is interesting, however, is the relationship between the ellipticity and the scale length as shown in the middle two panels of Fig. 10. The sampling uncertainty of the scale length increases as the ellipticity does, while the sampling uncertainty for the ellipticity is lower for larger scale length. For highly elliptical sources, both e_1 and e_2 need to be determined precisely to get an accurate measure of α_s , whereas for more circular sources, this is less important. The middle right-hand panel suggests that for larger sources, the ellipticity is simply easier to measure, as expected since these sources are both easier to resolve and, for a similar peak flux, have a higher integrated flux. The bottom panel indicates an increase in the maximum value of the sampling uncertainty for sources far from the phase centre, which is likely related to a decrease in sensitivity in the beam at the edges of the field of view.

Finally, within these interesting trends there is a great deal of scatter, implying that the properties of a source alone do not determine how quickly its parameters can be constrained. Because each source interacts with every other source due to the complicated beam pattern of the telescope, it is likely that the exact position of a source relative to the others will also influence constraining power.

5.3 Computational cost

Table 3 shows, for all the tests presented in the paper, the number of samples taken before a chain is converged that naturally increases as more sources are added to the problem. It also shows the average CPU time for computing a chain sample on a single core of Intel Xeon E5-2650 exploiting 2 NVIDIA Tesla K40 GPUs. Obviously, the computational time increases both with the number of sources and visibilities. Note that most of the computational time may be wasted in data transfer between CPU and GPU. Since GPU memory size is much smaller compared to the CPU one, large data sets as the ones produced by radio observations must be split in chunks to be sent through PCIe connection to the available GPUs. MONTBLANC implements this trying to overlap GPU computation with data transfer, however the current serial version of the code is not scalable with the number of GPUs used because of the many data chunks and transfer bandwidth bottleneck. The new generation of NVIDIA Tesla Pascal GPUs should overcome this issue by exploiting the new NVLink¹⁰ interconnection that maximizes the throughput of multi-GPUs and GPU/CPU system configurations through a larger bandwidth, more links, and a better scalability.

Moreover, a parallel version of the code is in preparation, implementing the Message Passing Interface paradigm (MPI¹¹) that will also allow the distribution of data and computation among more CPUs of hybrid multinode architectures. This implementation will be required when dealing with full SKA size data sets although their size may be reduced by working with gridded visibilities, as long as they had been gridded with an appropriate set of gridding kernels (Harrison & Brown 2015).

6 DISCUSSION ON AGN CONTAMINATION

The continuum faint radio sky observed at 1–2 GHz for weak lensing surveys is dominated by SF galaxy populations, however, a non-negligible fraction of sources is expected to be associated with AGNs (Jarvis et al. 2015; Smolcic et al. 2017; Bonaldi et al. 2019; Owen 2018). Therefore, an investigation about AGN classification and visibility shape modelling should be performed in order to handle their contamination. In deep radio fields two AGN populations are detected: radio-loud (RL), where synchrotron emission is dominated by large-scale relativistic jets and the lobes that the jets inflate, and radio-quiet (RQ), not showing jet-related emission and featuring much weaker radio emission. RL AGNs are well known to dominate the bright portion of the radio counts above 0.5 mJy at 1.4 GHz, but moving towards fluxes below few hundreds μ Jy this population should be progressively outnumbered by RQ AGN (Mancuso et al. 2017).

In Chang et al. (2004), a first attempt to model the shape of AGN components in the visibility domain is applied to the VLA FIRST survey where the dominating AGN population is RL due to the low sensitivity of the radio survey. They used shapelets to fit simultaneously all sources in a given pointing, including AGN lobes, thus not contaminating fainter sources in the primary beam and enabling cosmic shear measurement. However, while AGN will indeed be lensed along with SF galaxies, their complicated morphologies mean noise due to intrinsic shape dispersion will be large, and fitting simple parametrized profiles will lead to large model biases. Hence, we expect to remove RL AGN to get more accurate results. Several classification methods, mainly based on the comparison with other wave-bands catalogues of the same observed area, are already available and/or under further investigation (e.g. Padovani 2016; Barger et al. 2017). Such classification schemes will never be completely perfect, and leakage of RL AGN into the weak lensing sample will potentially create residual model biases and increase in shape noise. Initial studies indicate that successful classification rates of ~ 90 per cent should be sufficient (McCallum 2016).

¹⁰<https://www.nvidia.com/en-us/data-center/nvlink/>

¹¹<http://www.mpi-forum.org/>

Deep SKA precursor surveys such as e-MERGE¹² highlight that in the weak lensing regime a sizeable fraction (20–30 per cent) of SF galaxies may host RQ AGN (Bonzini et al. 2013; Delvecchio et al. 2017; Owen 2018), and thus particular interest should be given to these systems. Observed data, e.g. Guidetti et al. (2017), suggest that radio emission from RQ AGN is compact. Therefore, they may be modelled by a combination of exponential discs and Gaussians.

7 CONCLUSIONS

We have presented a new Bayesian method for measuring SF galaxy shape parameters from visibility data of radio observations. The method extends the BIRO technique by implementing a joint model fitting of the ellipticity and scale length of all exponential sources in the field of view. It applies an HMC technique, and can be easily extended for the measurement of other galaxy parameters such as position, flux, and in-band spectral index. Since we follow the RIME approach (see Section 2.2) for computing the model visibilities, it can also be extended for a simultaneous inference of scientific and instrumental parameters (Lochner et al. 2015).

We tested the method on simulated observations of up to 1000 galaxies adopting the SKA-MID Phase 1 UV coverage at 1.4 GHz. We were able to recover with a good accuracy the original galaxy shape parameters (see Table 1) at the source density of the proposed SKA1 radio weak lensing survey (2.7 gal arcim⁻²) with SNR_{vis} ≥ 10. As expected, the joint fitting approach improves measurements of galaxy ellipticities obtained with the *RadioLensfit* method at the same source density because it removes the neighbour bias introduced by the source extraction procedure. On the other hand, HMC has a long computation time, as shown in Table 3, which can be reduced with the new GPUs architectures and a higher level of code parallelization exploiting a hybrid multinode HPC system.

We also applied this method to the simulation of a fraction (100 arcmin²) of SuperCLASS, a precursor radio weak lensing survey performed combining observations of e-MERLIN and JVLA radio telescopes (although again we caution not to directly compare results from the simplified SKA simulation and realistic SuperCLASS simulation). Since the assumed minimum 10σ detection in the image domain corresponds to an SNR_{vis} ≥ 18, we obtained a faster convergence of the chains and better fitting than with SKA1 experiments. In fact, the convergence analysis showed that the parameters of bright sources are easier to constrain. It also showed that ellipticity sampling uncertainty is strongly correlated with the source size and it is lower for large sources.

These results show that working with visibilities may provide a more accurate source characterization, and more reliable and complete source catalogues, potentially offering a novel measurement approach for future SKA surveys. In this case, more computing resources may be required in order to deal with the large volume of data that will be produced by the future surveys and the complexity of the data analysis algorithms. Moreover, further investigation on AGN shape modelling should be performed as real observations will also contain a non-negligible fraction of AGN population.

ACKNOWLEDGEMENTS

We thank Simon Perkins for support with MONTBLANC, Isabella Prandoni for useful discussions about AGN and e-MERLIN and

JVLA extended source detection. We also thank the SuperCLASS collaboration for access to and assistance with their simulations.

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at are those of the authors and are not necessarily to be attributed to the NRF. 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 University Research Fellowship award.

REFERENCES

- Barger A. J., Cowie L. L., Owen F. N., Hsu L.-Y., Wang W.-H., 2017, *ApJ*, 835, 95
- Becker R., White R., Helfand D. J., 1995, *ApJ*, 450, 559
- Betancourt M., 2017, preprint ([arXiv:1701.02434](https://arxiv.org/abs/1701.02434))
- Bonaldi A., Bonato M., Galluzzi V., Harrison I., Massardi M., Kay S., Zotti G. D., Brown M., 2019, *MNRAS*, 482, 2
- Bonzini M., Padovani P., Mainieri V., Kellermann K. I., Miller N., Rosati P., Tozzi P., Vattakunnel S., 2013, *MNRAS*, 436, 3759
- Brown M. et al., 2015, in *Advancing Astrophysics with the Square Kilometre Array*, Vol. 2, Proc. Sci., SISSA, Trieste, p. 1365
- Chang T., Refregier A., 2002, *ApJ*, 570, 447
- Chang T., Refregier A., Helfand D., 2004, *ApJ*, 617, 794
- Delvecchio I. et al., 2017, *A&A*, 602, A3
- Gelman A., Rubin D., 1992, *Stat. Sci.*, 7, 457
- Guidetti D. et al., 2017, *MNRAS*, 471, 210
- Hamaker J. P., Bregman J. D., Sault R. J., 1996, *A&AS*, 117, 137
- Hanson K., 2001, in Milan S., Kenneth M. H., eds, *Proc. SPIE Conf. Ser. Vol. 4322, Medical Imaging 2001: Image Processing*. SPIE, Bellingham, p. 456
- Harrison I., Brown M. J., 2015, preprint ([arXiv:1507.06639](https://arxiv.org/abs/1507.06639))
- Harrison I., Camera S., Zuntz J., Brown M., 2016, *MNRAS*, 463, 3674
- Hastings W. K., 1970, *Biometrika*, 57, 97
- Högbom J., 1974, *A&A*, 15, 417
- Jarvis M. et al., 2015, in *Advancing Astrophysics with the Square Kilometre Array*, Vol. 2, Proc. Sci., SISSA, Trieste, PoS(AASKA14)018
- Jarvis M. et al., 2016, *MNRAS*, 460, 2245
- Kacprzak T., Bridle S., Rowe B., Voigt L., Zuntz J., Hirsch M., MacCrann N., 2014, *MNRAS*, 441, 2528
- Lochner M., Natarajan I., Zwart J. T. L., Smirnov O., Bassett B. A., Oozeer N., Kunz M., 2015, *MNRAS*, 450, 1308
- Mancuso C. et al., 2017, *ApJ*, 842, 95
- Mandelbaum R. et al., 2015, *MNRAS*, 450, 2963
- Martí-Vidal I., Vlemmings W. H. T., Muller S., Casey S., 2014, *A&A*, 563, A136
- McCallum N., 2016, Master's thesis, Univ. Manchester
- McMullin J. P., Waters B., Schiebel D., Young W., Golap K., 2007, in Shaw R. A., Hill F., Bell D. J., eds, *ASP Conf. Ser. Vol. 376, Astronomical Data Analysis Software and Systems XVI*. Astron. Soc. Pac., San Francisco, p. 127
- Melchior P., Böhnert A., Lombardi M., Bartelmann M., 2010, *A&A*, 510, A75
- Metropolis N., Rosenbluth A. W., Rosenbluth M. N., Teller A. H., Teller E., 1953, *J. Chem. Phys.*, 21, 1087
- Miller L. et al., 2013, *MNRAS*, 429, 2858
- Neal R. M., 2011, in Brooks S., Gelman A., Jones G., Meng X.-L., eds, *MCMC using Hamiltonian dynamics*. Chapman & Hall/CRC Press, Florida
- Noordam J. E., Smirnov O. M., 2010, *A&A*, 524, A61
- Owen F., 2018, *ApJS*, 235, 34
- Owen F., Morrison G., 2008, *AJ*, 136, 1889
- Padovani P., 2016, *A&AR*, 24, 13
- Patel P., Bacon D. J., Beswick R. J., Muxlow T. W. B., Hoyle B., 2010, *MNRAS*, 401, 2572

¹²<http://www.e-merlin.ac.uk/legacy/projects/emerge.html>

- Patel P. et al., 2015, in *Advancing Astrophysics with the Square Kilometre Array*, Vol. 2, Proc. Sci., SISSA, Trieste, p. 1427
- Peck L. W., Fenech D. M., 2013, *Astron. Comput.*, 2, 54
- Perkins S., Marais P., Zwart J., 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., 2016a, in *The many facets of extragalactic radio surveys: towards new scientific challenges*. Proc. Sci., SISSA, Trieste, PoS(EXTRA-RADSUR2015)033
- Rivi M., Miller L., Makhathini S., Abdalla F. B., 2016b, *MNRAS*, 463, 1881
- Schwarz U., 1978, *A&A*, 65, 345
- Skilling J., 2004, in *Rainer F., Roland P., Udo von T.*, eds, *AIP Conf. Proc.* Vol. 735, *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*. Am. Inst. Phys., New York, p. 395
- Smirnov O. M., 2011a, *A&A*, 527, A106
- Smirnov O. M., 2011b, *A&A*, 527, A107
- Smolcic V. et al., 2017, *A&A*, 602, A1
- Sokal A., 1997, in *DeWitt-Morette C., Cartier P., Folacci A.*, eds, *Monte Carlo Methods in Statistical Mechanics: Foundations and New Algorithms*. Springer US, Boston, p. 131
- Spergel D., 2010, *ApJS*, 191, 58
- Taylor J. F., Ashdown M. A. J., Hobson M. P., 2008, *MNRAS*, 389, 1284
- Thompson A., Moran J., Swenson G., 1986, *Interferometry and Synthesis in Radio Astronomy*. John Wiley & Sons, New York, p. 554
- Tunbridge B., Harrison H., Brown M., 2016, *MNRAS*, 463, 3339
- Voigt L. M., Bridle S. L., 2010, *MNRAS*, 404, 458
- Wrobel J., Walker R., 1999, in *Taylor G., Carilli C., Perley R.*, eds, *ASP Conf. Ser.* Vol. 180, *Synthesis Imaging in Radio Astronomy II*. Astron. Soc. Pac., San Francisco, p. 171

APPENDIX A: LIKELIHOOD GRADIENT

The MONTBLANC code returns the chi-squared value computed comparing data visibilities \tilde{V}_i with the sky model visibilities $V_i(\mathbf{x})$:

$$\chi^2(\mathbf{x}) = \sum_i \frac{|\tilde{V}_i - V_i(\mathbf{x})|^2}{\sigma_i^2}. \quad (\text{A1})$$

From it the likelihood can be easily obtained as $\mathcal{L}(\mathbf{x}) \propto \exp[-\chi^2(\mathbf{x})/2]$. Similarly we can compute the likelihood gradient with respect to a set of parameters p_1, \dots, p_M by adding to MONTBLANC the computation of the corresponding chi-squared partial

derivatives:

$$\begin{aligned} \frac{\partial \chi^2(\mathbf{x})}{\partial p_j} &= -2 \sum_i \frac{\Re \tilde{V}_i - \Re V_i(\mathbf{x})}{\sigma_i^2} \cdot \frac{\partial \Re V_i(\mathbf{x})}{\partial p_j} + \\ &\quad -2 \sum_i \frac{\Im \tilde{V}_i - \Im V_i(\mathbf{x})}{\sigma_i^2} \cdot \frac{\partial \Im V_i(\mathbf{x})}{\partial p_j}, \end{aligned} \quad (\text{A2})$$

where \Re and \Im denote the real and imaginary part of the complex visibilities. Adopting visibility formulation as in equation (13), galaxy shape parameters $p_{s,k}$ are contained only in the brightness matrix of source s , therefore

$$\frac{\partial V_{ipq\lambda}}{\partial p_{s,k}} = \frac{\partial V_{ipq\lambda s}}{\partial p_{s,k}} = K_{tps\lambda} \frac{\partial B_{s\lambda}}{\partial p_{s,k}} K_{tqs\lambda}^H. \quad (\text{A3})$$

From equation (14), we have the following partial derivatives of the model visibilities with respect to Sérsic shape parameters $\alpha_s, \mathbf{e}_s = (e_{1,s}, e_{2,s})$ of each galaxy $s = 1, \dots, N$ in the field of view:

$$\frac{\partial V_{ipq\lambda}}{\partial \alpha_s} = -\frac{12\pi^2 \alpha_s |\mathbf{A}_s^{-T} \mathbf{k}|^2}{1 + 4\pi^2 \alpha_s^2 |\mathbf{A}_s^{-T} \mathbf{k}|^2} \cdot V_{ipq\lambda s}, \quad (\text{A4})$$

$$\frac{\partial V_{ipq\lambda}}{\partial e_{1,s}} = -\frac{12\pi^2 \alpha_s^2 f_1(\mathbf{e}_s)}{1 + 4\pi^2 \alpha_s^2 |\mathbf{A}_s^{-T} \mathbf{k}|^2} \cdot V_{ipq\lambda s}, \quad (\text{A5})$$

$$\frac{\partial V_{ipq\lambda}}{\partial e_{2,s}} = -\frac{12\pi^2 \alpha_s^2 f_2(\mathbf{e}_s)}{1 + 4\pi^2 \alpha_s^2 |\mathbf{A}_s^{-T} \mathbf{k}|^2} \cdot V_{ipq\lambda s}, \quad (\text{A6})$$

with

$$\begin{aligned} f_1(\mathbf{e}_s) &= \frac{[(1 + e_{1,s})u + e_{2,s}v]u}{(1 - e_{1,s}^2 - e_{2,s}^2)^2} + \frac{2e_{1,s} [(1 + e_{1,s})u + e_{2,s}v]^2}{(1 - e_{1,s}^2 - e_{2,s}^2)^3} \\ &\quad - \frac{[e_{2,s}u + (1 - e_{1,s})v]v}{(1 - e_{1,s}^2 - e_{2,s}^2)^2} + \frac{2e_{1,s} [e_{2,s}u + (1 - e_{1,s})v]^2}{(1 - e_{1,s}^2 - e_{2,s}^2)^3} \end{aligned}$$

and

$$\begin{aligned} f_2(\mathbf{e}_s) &= \frac{[(1 + e_{1,s})u + e_{2,s}v]v}{(1 - e_{1,s}^2 - e_{2,s}^2)^2} + \frac{2e_{2,s} [(1 + e_{1,s})u + e_{2,s}v]^2}{(1 - e_{1,s}^2 - e_{2,s}^2)^3} \\ &\quad + \frac{[e_{2,s}u + (1 - e_{1,s})v]u}{(1 - e_{1,s}^2 - e_{2,s}^2)^2} + \frac{2e_{2,s} [e_{2,s}u + (1 - e_{1,s})v]^2}{(1 - e_{1,s}^2 - e_{2,s}^2)^3}. \end{aligned}$$

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.