






# Non-Gaussianity constraints using future radio continuum surveys and the multitracer technique

Zahra Gomes , <sup>1</sup>★ Stefano Camera , <sup>2,3,4,5</sup> Matt J. Jarvis , <sup>1,5</sup> Catherine Hale <sup>1,6</sup>  
and José Fonseca <sup>5,7,8</sup>

<sup>1</sup>*Oxford Astrophysics, Department of Physics, University of Oxford, Keble Road, Oxford OX1 3RH, UK*

<sup>2</sup>*Dipartimento di Fisica, Università degli Studi di Torino, via P. Giuria 1, I-10125 Torino, Italy*

<sup>3</sup>*INFN – Istituto Nazionale di Fisica Nucleare, Sezione di Torino, via P. Giuria 1, I-10125 Torino, Italy*

<sup>4</sup>*INAF – Istituto Nazionale di Astrofisica, Osservatorio Astrofisico di Torino, strada Osservatorio 20, I-10025 Pino Torinese, Italy*

<sup>5</sup>*Department of Physics, University of the Western Cape, Bellville 7535, South Africa*

<sup>6</sup>*CSIRO Astronomy and Space Science, 26 Dick Perry Avenue, Kensington, Perth 6151, WA, Australia*

<sup>7</sup>*Dipartimento di Fisica e Astronomia ‘G. Galilei’, Università degli Studi di Padova, via Marzolo 8, I-35131 Padova, Italy*

<sup>8</sup>*INFN – Istituto Nazionale di Fisica Nucleare, Sezione di Padova, via Marzolo 8, I-35131 Padova, Italy*

Accepted 2019 December 18. Received 2019 December 17; in original form 2019 August 20

## ABSTRACT

Tighter constraints on measurements of primordial non-Gaussianity (PNG) will allow the differentiation of inflationary scenarios. The cosmic microwave background bispectrum – the standard method of measuring the local non-Gaussianity – is limited by cosmic variance. Therefore, it is sensible to investigate measurements of non-Gaussianity using the large-scale structure. This can be done by investigating the effects of non-Gaussianity on the power spectrum on large scales. In this study, we forecast the constraints on the local PNG parameter  $f_{\text{NL}}$  that can be obtained with future radio surveys. We utilize the multitracer method that reduces the effect of cosmic variance and takes advantage of the multiple radio galaxy populations that are differently biased tracers of the same underlying dark matter distribution. Improvements on previous work include the use of observational bias and halo mass estimates, updated simulations, and realistic photometric redshift expectations, thus producing more realistic forecasts. Combinations of Square Kilometre Array simulations and radio observations were used as well as different redshift ranges and redshift bin sizes. It was found that in the most realistic case the  $1\sigma$  error on  $f_{\text{NL}}$  falls within the range 4.07–6.58, rivalling the tightest constraints currently available.

**Key words:** galaxies: active – cosmological parameters – inflation – large-scale structure of Universe – radio continuum: galaxies.

## 1 INTRODUCTION

Primordial non-Gaussianity (PNG) describes deviations from a Gaussian random field in the initial density field present after inflation. These small fluctuations are believed to seed the formation of all the large-scale structure (LSS) present in the Universe today as well as the cosmic microwave background (CMB) temperature anisotropies that we observe and in the inflationary scenario they were formed from quantum fluctuations during or just after inflation. Currently, there exists a large number of models for inflation which all fit with current observations, but because different models provide different predictions for the extent of PNG present, this measurement is a valuable tool for discriminating between these

models (see Bartolo et al. 2004 for a review). Most inflationary models predict that the non-Gaussianity depends only on the local value of the potential, such PNG is said to be of the ‘local type’ and it is parametrized by the  $f_{\text{NL}}$  parameter that quantifies the deviation from a Gaussian random field  $\phi$

$$\Phi = \phi + f_{\text{NL}} (\phi^2 - \langle \phi \rangle^2). \quad (1)$$

Simple slow-roll, single-field inflationary models predict almost Gaussian density fields with  $f_{\text{NL}} \ll 1$  ( $\mathcal{O}(10^{-2})$ ) (Maldacena 2003; Creminelli & Zaldarriaga 2004) while multifield models allow larger deviations from Gaussianity with  $f_{\text{NL}} \gtrsim 1$  (Lyth, Ungarelli & Wands 2003; Zaldarriaga 2004). The current tightest constraint on the local PNG parameter is  $f_{\text{NL}} = 0.8 \pm 5$ , which was obtained using Planck 2015 CMB data (Planck Collaboration XVII 2016). Planck 2018 data (Planck Collaboration IX 2019) produced the slightly less constraining result  $f_{\text{NL}} = 0.9 \pm 5.1$ . This recent result was less

\* E-mail: zahra.gomes@physics.ox.ac.uk

constraining due to the use of more realistic polarization simulations with higher noise levels for estimating the errors. In addition, for LSS measurements, general relativistic projection effects and non-linear evolution are expected to produce scale-dependent effects on LSS measurements very similar to those of PNG with  $f_{\text{NL}}$  of order unity (Yoo 2010; Bruni et al. 2012; Jeong, Schmidt & Hirata 2012; Camera, Maartens & Santos 2015). Therefore, in order to start distinguishing between single-field and multifield inflation,  $\sigma(f_{\text{NL}}) \sim 1$  is required, and therefore this is normally the target of such analyses (de Putter, Gleyzes & Doré 2017).

Local PNG has – to date – been measured most precisely using the bispectrum of the CMB temperature anisotropy maps (Komatsu et al. 2003; Planck Collaboration XVII, IX 2016, 2019). The main disadvantage of this method is that such measurements are limited by cosmic variance on large scales and by Silk damping on small scales (Alvarez et al. 2014; Ferraro & Smith 2015). LSS measurements are a viable alternative to the CMB bispectrum as the non-Gaussianity also leaves an imprint on the LSS. In particular,  $f_{\text{NL}} > 0$  leads to an increase in the 3D power spectrum on large scales, which corresponds to an increase in the total halo bias on large scales (Carbone, Verde & Matarrese 2008; Dalal et al. 2008; Matarrese & Verde 2008). Large galaxy surveys (both optical and radio) such as the Sloan Digital Sky Survey (SDSS) and the National Radio Astronomy Observatory (NRAO) Very Large Array (VLA) Sky Survey (NVSS) have been used to measure the clustering of dark matter and constrain the effects of PNG (Slosar et al. 2008; De Bernardis et al. 2010; Xia et al. 2010; Ross et al. 2013). One approach to this LSS measurement is to use quasar populations as these are highly biased dark matter tracers and cover large volumes of space. Leistedt, Peiris & Roth (2014) and Castorina et al. (2019) study the clustering of the SDSS quasar sample and provide the tightest constraints with LSS methods to date. Combinations of galaxy surveys and CMB maps have also been considered (Giannantonio & Percival 2014).

As this PNG effect is most prominent on large scales, very large surveys are used, but these are still limited by cosmic variance on the larger scales and forecasts of the tightest constraints possible with single tracer surveys have found that  $\sigma(f_{\text{NL}}) \sim 1$  is not possible. The multitracer method, first proposed by Seljak (2009) significantly reduces the effect of cosmic variance by cross-correlating multiple dark matter tracers, with different biases located in the same sky volume. These populations have the same underlying density distributions and thus the ratio of their biases will cancel the effects of cosmic variance resulting in a measurement that is only limited by shot noise. This multitracer technique, applied to the large upcoming galaxy surveys with multiple differently biased galaxy populations is expected to rival the current CMB constraints (see Ferramacho et al. 2014; Yamauchi, Takahashi & Oguri 2014; Alonso & Ferreira 2015).

An alternative avenue for probing the effects of PNG on the LSS is intensity mapping and future surveys of this type seem promising due to the combination of their large sky coverage and the potential to push to high redshifts (Joudaki et al. 2011; Camera et al. 2013; Li & Ma 2017; Fonseca, Maartens & Santos 2018). The multitracer method has also been suggested for use on HI intensity mapping surveys. For example, Fonseca et al. (2018) forecast that performing such an analysis on the HI survey from the Square Kilometre Array (SKA; Square Kilometre Array Cosmology Science Working Group 2018) and the H $\alpha$  survey from SPHEREx (Doré et al. 2014) could lead to constraints on  $f_{\text{NL}}$  on the order of  $\sigma(f_{\text{NL}}) \sim 1$ . Applying the multitracer method to combinations of intensity mapping and photometric galaxy surveys is also possible and was first suggested

by Alonso & Ferreira (2015) and Fonseca et al. (2015). More recently, the combination of intensity mapping from SKA and galaxy surveys from the Large Synoptic Survey Telescope (LSST; LSST Science Collaboration 2009) was suggested and investigated by Witzemann et al. (2018) and intensity mapping from MeerKAT and galaxy surveys from the Dark Energy Survey (DES; The Dark Energy Survey Collaboration 2005) was suggested by Fonseca, Maartens & Santos (2017) and forecasts have found that these analyses should provide constraints more than two times better than Planck (Fonseca et al. 2017). Recently, a forecast analysis has been performed on the use of the multitracer method for combining HI intensity mapping with both photometric galaxy surveys and CMB lensing and the constraints of  $\sigma(f_{\text{NL}}) \sim 1$  were obtained (Ballardini, Matthewson & Maartens 2019).

Further research on improving the methods currently being used for measuring the  $f_{\text{NL}}$  signal is also being done by improving the measurement of the galaxy bispectrum and accounting for systematics (Tellarini et al. 2016; Welling, van der Woude & Pajer 2016; Mueller, Percival & Ruggeri 2017; Karagiannis et al. 2018; Uhlemann et al. 2018). Another example is a recently proposed method that involves the creation of a sample with zero bias, allowing the effect of PNG on the clustering to be measured more easily (Castorina et al. 2018). Similarly, studies on the optimal galaxy survey for providing  $f_{\text{NL}}$  constraints of  $\sigma(f_{\text{NL}}) = 1$  are also being done (de Putter & Doré 2017).

Radio continuum surveys are appealing for use with multitracer techniques for a number of reasons: radio surveys cover very large areas of sky, which is particularly advantageous since the increased bias signal is detected on large scales, and is also useful for reducing shot noise. Just as importantly, radio surveys contain a range of distinct radio galaxy populations that have significant spread in their biases (Hale et al. 2018). Other advantages of radio continuum surveys are that they allow the observation of galaxies to very high redshifts and without dust obscuration (Jarvis et al. 2015), which has been shown to provide advantages in determining cosmological parameters (Camera et al. 2012). Contrary to intensity mapping, Galactic and extragalactic foregrounds are not significant issues for the continuum surveys that use resolved galaxy populations. On the other hand, one disadvantage of using radio surveys is that it is very difficult to get precise redshifts for the individual sources: either multiwavelength data or HI 21-cm line measurements are necessary (Jarvis et al. 2015).

The SKA (Carilli & Rawlings 2004; Dewdney et al. 2009) will be the world's largest radio telescope with thousands of dishes and antennas distributed over South Africa and Australia. It will consist of a mid-frequency instrument (SKA-MID) and a low-frequency instrument (SKA-LOW) and will span a frequency range of 50 MHz to 15 GHz with unprecedented angular resolution and sensitivity. It will be constructed in two main phases, with the first composed of 10 per cent of the intended eventual collecting area. The SKA's radio continuum survey will cover  $2\pi$  steradians and redshifts  $<5$ , while its spectroscopic HI survey will detect galaxies up to redshift  $\sim 2$ . HI intensity mapping surveys that cover redshifts up to 5 may also be possible with the SKA (Quinn et al. 2015; Santos et al. 2015).

In this paper, we update the work by Ferramacho et al. (2014) who perform a tomographic multitracer analysis driven by multiple radio galaxy populations with simulated masses and then use this to forecast the PNG constraints possible with the SKA Phase 1. We will perform a similar analysis, using both the galaxy masses estimated by the  $S^3$  simulation (Wilman et al. 2008) and those measured by Hale et al. (2018) who use VLA COSMOS field

observations to estimate the average bias values and corresponding halo masses of the radio galaxy populations. In addition, redshift distributions predicted by both  $S^3$  and the more recent Tiered Radio Extragalactic Continuum Simulation (T-RECS) by Bonaldi et al. (2018) are used and compared. In Section 2, we present an overview of the theoretical framework used for this analysis, followed by a discussion of our choice of radio galaxy populations in Section 3. Our methods are described in Section 4 and our results and analysis are presented in Section 5. Finally, we conclude in Section 6. We use a fiducial cosmology with  $H_0 = 67.74$ ,  $\Omega_\Lambda = 0.6911$ ,  $\Omega_{\text{CDM}} = 0.26$ ,  $\Omega_b = 0.05$ ,  $A_s = 2.142 \times 10^{-9}$ , and  $f_{\text{NL}} = 0$  taken from the Planck 2015 results (Planck Collaboration XVII 2016).

## 2 PRIMORDIAL NON-GAUSSIANITY IN LARGE-SCALE STRUCTURE

A non-zero local PNG signal will increase the power of the halo power spectrum on large scales, corresponding to a scale-dependent correction to the halo bias given by the expression below (Carbone et al. 2008; Dalal et al. 2008; Matarrese & Verde 2008; Slosar et al. 2008)

$$b_h(M, z) = b_L(M, z) + f_{\text{NL}} \delta_c [b_L(M, z) - 1] \frac{3\Omega_m H_0^2}{c^2 k^2 T(k) D(z)}, \quad (2)$$

where  $b_h$  is the total halo bias,  $b_L$  is the Gaussian linear halo bias,  $f_{\text{NL}}$  is the amplitude of the non-Gaussianity (defined in equation 1),  $\delta_c$  is the critical overdensity for spherical collapse at redshift  $z = 0$ ,  $T(k)$  is the linear transfer function, and  $D(z)$  is the growth function. The PNG parameter  $f_{\text{NL}}$  is therefore normally determined from large-scale structure by measuring the galaxy power spectrum or bispectrum (De Bernardis et al. 2010; Ross et al. 2013). On the other hand, in situations where spectroscopic redshift data are not available (which will be the case with some current and future large-scale surveys with volumes and depths that make it unfeasible to measure spectroscopic redshifts), and only the less accurate photometric redshifts or estimated redshift distributions are present, neither the 3D matter power spectrum nor the bispectrum can be reliably computed. Instead, a tomographic analysis with angular power spectra can be used. The photo- $z$  estimates are used to place the galaxies into redshift bins and then the angular power spectrum, which is the projection of the line-of-sight haloes (within a particular redshift bin), into the 2D plane can be computed. Although the angular power spectra do not provide as much information as the full 3D analysis, it is a sufficient estimate. Some analyses that have used this method are: Slosar et al. (2008) and Xia et al. (2010).

For the case of a multitracer analysis using these angular power spectra, the statistical information is given by the auto- and cross-correlation power spectra that are split into multipoles (see Huterer, Knox & Nichol 2001)

$$C_l^{i,j} = \frac{2}{\pi} \int_{k_{\text{min}}}^{k_{\text{max}}} k^2 P_\delta(k) W_l^i(k) W_l^j(k) dk. \quad (3)$$

$P_\delta(k)$  is the dark matter power spectrum at redshift 0.  $W_l^i$  are window functions that incorporate the redshift distribution ( $dn/dz$ ) and bias  $b$  of the tracer  $i$  as well as the angular geometry of multipole  $l$  via a spherical Bessel function  $j_l(kr)$  where  $r$  is the comoving radial distance to redshift  $z$ . The form of these window functions is given below

$$W_l^i = \int \frac{dn^i}{dz} D(z) b_h^i(z) j_l(kr) dz. \quad (4)$$

In this analysis, we will compute these auto- and cross-power spectra for a variety of tracers, defined by radio galaxy type and redshift bin, using estimated bias functions and redshift distributions given the constraints of the SKA telescope. Following this, we will perform a Fisher analysis of this multitracer method to determine the constraints on the  $f_{\text{NL}}$  parameter that will be possible with the SKA. The relatively large width of the redshift bins used in this analysis allows us to neglect the effects of redshift space distortions.

## 3 CHOICE OF RADIO GALAXY POPULATIONS

There have been a number of attempts to analyse radio galaxies by population and determine their average dark matter halo masses and biases (Wilman et al. 2008; Lindsay et al. 2014; Magliocchetti et al. 2017; Bonaldi et al. 2018; Hale et al. 2018). Wilman et al. (2008) performed a semi-empirical simulation of the extragalactic radio continuum sky covering an area of  $20 \times 20 \text{ deg}^2$  with a flux density limit of 10 nJy over a range of frequencies. This simulation was created as part of the SKA Design Study Simulated Skies ( $S^3$ ). Five distinct galaxy populations were identified: star-forming galaxies (SFGs), star bursts (SBs), Fanaroff–Riley Class I (FR Is), Fanaroff–Riley Class II (FR IIs), and radio quiet quasars (RQQs) and average halo masses and biases were computed for each population. The very low flux limits of  $S^3$  make it easily scalable to the proposed 5  $\mu\text{Jy}$  flux density limit at 1.4 GHz of SKA Phase 1 and the five distinct populations of radio galaxies make this set-up appealing for forecasting the constraints on  $f_{\text{NL}}$  from the SKA using the multitracer method. This was done by Ferramacho et al. (2014) who used the redshift distributions and the mean masses of these five populations to compute auto- and cross-correlation angular power spectra and perform a Fisher analysis.

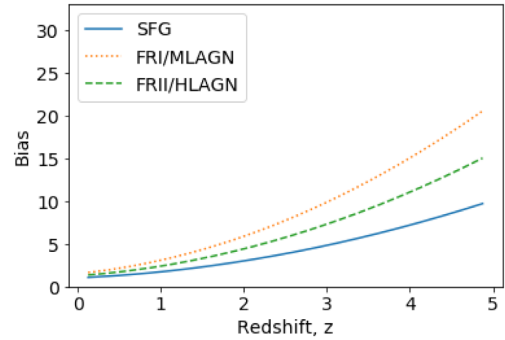
The T-RECS (Bonaldi et al. 2018) is an updated simulation created for the SKA with a sky area of  $25 \text{ deg}^2$  and a 150 MHz to 20 GHz frequency range. Bonaldi et al. (2018) do not provide independent estimates for halo masses of galaxy populations but do compare clustering measurements for the T-RECS sources to Hale et al. (2018), and find that the observed biases for the SFGs are higher than predicted while for the active galactic nuclei (AGNs) there is reasonable agreement.

In contrast to the simulated data of  $S^3$  and T-RECS, Hale et al. (2018) explored the clustering of different radio-selected galaxy populations taken from 3 GHz VLA imaging over the  $\sim 2 \text{ deg}^2$  COSMOS field (see Smolčić et al. 2017b) using a  $5.5\sigma$  cut which corresponds to a mean flux density limit of 22  $\mu\text{Jy}$  at 1.4 GHz. Using the angular two-point correlation functions of the radio galaxies, they estimated average bias values for the median redshifts of the samples and corresponding halo masses for these galaxy populations. See fig. 10 of Hale et al. (2018) for a comparison of the measured biases and the  $S^3$  simulated bias trends as a function of redshift. The populations considered were: high and low radio luminosity SFGs, high and low radio luminosity AGNs, and high and moderate luminosity AGNs (HLAGNs and MLAGNs). The radio galaxies were classified into these categories by Smolčić et al. (2017a) who gathered a wealth of multiwavelength [optical, near-UV, near-infrared (IR), mid-infrared, and X-ray] data to perform the classifications.

The two most common ways of classifying radio AGNs are by morphology and by accretion mode. The morphological method or Fanaroff–Riley classification (Fanaroff & Riley 1974) splits AGNs based on the ratio of the distance between their regions of greatest surface brightness and the total length of the galaxy. Galaxies with their brightest regions near their centres are called FR Is while

those with bright regions at the edges of extended lobes are labelled FRIIs. This is a relatively straightforward classification that can be applied when only radio images, with sufficient resolution, are present (not all AGNs will fall neatly into one of these categories, making them difficult to classify). The second method, the accretion mode classification distinguishes AGNs based on their radiative efficiency. Radiatively efficient AGNs are those with a radiative accretion mode, meaning that they have an accretion disc and follow the orientation-based AGN unification model (Urry & Padovani 1995). They also have luminosities of  $L \gtrsim 0.01L_{\text{Edd}}$ , where  $L_{\text{Edd}}$  is the Eddington luminosity and are therefore called HLAGNs. When these AGNs are radio loud and extended they are also called high-excitation radio galaxies (HERGs), but QSOs are also included in this HLAGN category. On the other hand, the radiatively inefficient AGNs accrete in the jet mode, they have no accretion disc and instead their accretion flow is advection dominated. These have luminosities  $L \lesssim 0.01L_{\text{Edd}}$  and are called MLAGNs. The radio-loud variants are called low-excitation radio galaxies (LERGs) and radio quiet AGN LINERS are the other members of this group. See section 2.1 of Heckman & Best (2014) for a full description of the classification. Unlike the FR I/FR II classification, this method requires a wealth of multiwavelength data that will not always be present.

In order to conduct this multitracer analysis, distinct populations with different biases must be identified. The biases of SFGs have been found to be much lower than those of AGNs (Hale et al. 2018) and this is expected because AGNs are hosted by more massive galaxies than the blue SFG population. We choose not to distinguish between low- and high-luminosity SFGs or SFG and starburst galaxies (SBs) as these distinctions are difficult to make with only radio continuum data. The physically motivated choice of classification for the AGNs is the accretion mode classification because MLAGNs and HLAGNs are expected to reside in different environments and evolve differently over time due to their different ionization properties. Hale et al. (2018) found that MLAGNs are significantly more biased than HLAGN and therefore inhabit more massive haloes, this might be because the hot gas in the most massive haloes is not easily accreted on to AGNs, leading to more massive haloes hosting less efficient AGNs. This result supports the findings of Hardcastle (2004), Tasse et al. (2008), Janssen et al. (2012), Ramos Almeida et al. (2013), and Gendre et al. (2013). Unfortunately, the wealth of multiwavelength data required for classifications of MLAGN/HLAGN to be made will not be present for large volumes of sky in the early stages of the SKA. On the other hand, it will be possible to identify the FR I and FR II galaxies from radio images, as long as the angular resolution is better than 1 arcsec (Wilman et al. 2008; Ferramacho et al. 2014). Therefore, since radio-loud MLAGNs are normally associated with FR I galaxies while HLAGNs are associated with FR II, for the purposes of this forecast analysis, we assume that these populations are the same although this is not strictly true. This assumption allows us to use the biases obtained from  $S^3$  and Hale et al. (2018; see Section 4.1) and the number distributions found for FR I/FR II from the  $S^3$  and T-RECS simulations (see Section 4.2). Therefore, in this analysis, with the aim of predicting constraints that are realistic, we use only the populations: SFGs, FR I/MLAGNs and FR II/HLAGNs. It is important to keep in mind the caveat that FR Is and FR IIs are similar to, but not the same as HLAGNs and MLAGNs, respectively. For instance, the HLAGNs/MLAGNs distinction of the VLA-COSMOS 3 GHz survey (Smolčić et al. 2017a) includes the radio quiet quasars, which are included as a distinct population in Wilman et al. (2008).



**Figure 1.** Bias redshift evolution for the different galaxy populations assuming a Gaussian distribution of masses for each population, using the central masses determined by Hale et al. (2018).

## 4 FISHER ANALYSIS FOR SKA FORECASTS

### 4.1 Halo bias

Bias values for the haloes of each galaxy population and for each redshift bin were required to perform the multitracer analysis (see equations 3 and 4). In order to find these bias distributions, a mean halo mass,  $M_{\text{cent}}$  was adopted for each population, based on the observed/simulated estimations and it was assumed that the galaxy haloes spanned a Gaussian distribution of masses,  $f$ , with a mean  $M_{\text{cent}}$  and a standard deviation of  $0.2M_{\text{cent}}$ . The linear halo bias distribution over a Gaussian distribution of masses,  $b^i(z)$  was then calculated using the below equation

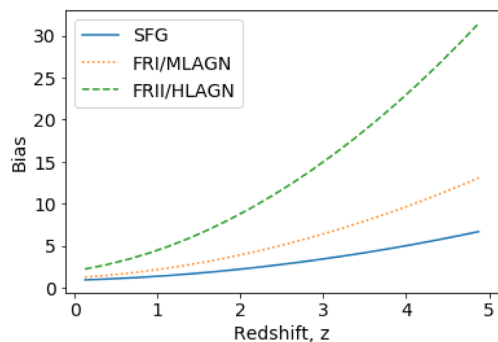
$$b^i(z) = \int f^i(M, M_{\text{cent}}^i) b_L^i(M, z) dM, \quad (5)$$

where  $i$  represents the galaxy population and  $b_L^i(M, z)$  is the linear bias as a function of halo mass and redshift. Ferramacho et al. (2014) found that this simplistic approach was sufficient to describe the mass distributions and that changing this form did not notably change the constraints obtained. In order to calculate the Gaussian linear bias function above, equation (6) was used along with the estimated average halo masses (obtained from  $S^3$  and Hale et al. 2018) and the mass variance  $\sigma$  found using the halo mass function (the number density of haloes as a function of mass) calculator from Murray, Power & Robotham (2013). The linear bias is given by

$$b_L(M, z) = 1 + \frac{qv - 1}{\delta_c(0)} + \frac{1}{\delta_c(0)} \frac{2p}{1 + (qv)^p}, \quad (6)$$

where  $v = \delta_c^2(0)/\sigma^2(M, z)$ ,  $\delta_c(0)$  is the critical overdensity for spherical collapse at redshift  $z = 0$ , and  $\sigma(M, z)$  is the mass variance, the root mean square fluctuation in spheres which, on average, contain mass  $M$  at the initial time. The accepted values of  $p$  and  $q$  are 0.3 and 0.75, respectively (Sheth & Tormen 1999). The angular power spectra code utilized in this analysis (Fonseca et al. 2015) required a single bias for each redshift bin and therefore we extracted the average bias in each bin from the models found and used these values.

Two sets of mean halo masses were used to produce bias functions. One set of mass values were from the radio observations of Hale et al. (2018) and the other was from  $S^3$ . The bias functions obtained for both the Hale et al. (2018) and the  $S^3$  masses can be seen in Figs 1 and 2, respectively. The bias values found by Hale et al. (2018) for the MLAGNs (FR I) were higher than those from  $S^3$ , values for the SFGs were slightly lower and values for the HLAGNs (FR II) were significantly lower than those predicted by  $S^3$ .



**Figure 2.** Same as Fig. 1 using the  $S^3$  simulation by Wilman et al. (2008).

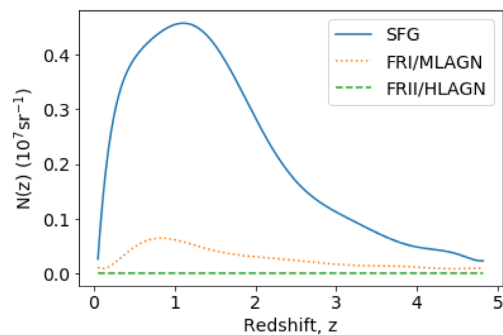
Hale et al. (2018) identified one population of SFGs, which consisted of both normal SFGs and SBs, while the  $S^3$  simulation separated these populations. In this analysis, we do not wish to distinguish between SFGs and SBs as we do not expect these to be distinguished with SKA data and therefore we choose to combine these galaxy populations of the  $S^3$  simulation and used these combined biases (and redshift distributions, which will be discussed in the following section). We find the effective bias for the combined SFG and SB using a weighted mean based on the numbers of each population in each redshift bin

$$b^{\text{eff}}(z) = \frac{\sum_i n^i(z)b^i(z)}{\sum_i n^i(z)}. \quad (7)$$

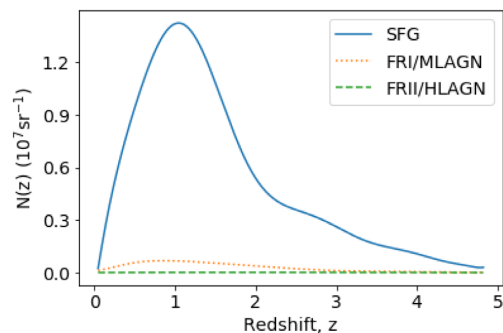
Following this, we find the single halo mass that best fits this bias distribution and use this halo mass to conduct our analysis. In the case in which we use the  $S^3$  biases and the T-RECS number distributions, we do not have differentiated SB and SFG distributions from T-RECS and therefore we assume that the ratio of SFG to SB remains the same as in  $S^3$ , and we use the same bias function. The best-fitting halo mass was found to be  $1.1 \times 10^{12} M_\odot h^{-1}$  and the corresponding bias function is plotted in Fig. 2. We notice that the combined bias of the SFG and SB of the  $S^3$  simulation was similar in magnitude to the corresponding bias obtained by Hale et al. (2018).

When exploring the constraints obtained with different populations, we also consider the case in which we are unable to differentiate the MLAGN/FR I from the HLAGN/FR II and therefore we combine the two populations, and calculate the combined bias with the method shown above.

The bias functions presented above are extrapolated to  $z = 5$  while the bias measurements obtained from observations such as Hale et al. (2018) do not extend upwards of  $z \sim 2$ , Wilman et al. (2008) suggest a cut on the bias function at  $z = 1.5$  for AGNs and  $z = 3$  for SFGs with a flat bias after this in order to prevent the bias from becoming unrealistically large, but there is little observational evidence for such an abrupt cut. Indeed, recent work on measuring the bias of Lyman-break galaxies at high redshifts suggest that the bias could be as high as  $b \sim 8-11$  at  $z \sim 6$  (Hatfield et al. 2018). This is therefore in the range of the assumed bias for our FR I/MLAGN and SFG samples, which are the dominant radio populations at these redshifts for the survey depth considered. Furthermore, one might expect that the host galaxies of the FR I/MLAGN population are more massive than the Lyman-break galaxies at similar redshifts, thus a higher bias may be expected. However, due to the uncertainty present at redshifts above  $z \sim 2$  analyses performed with redshifts below this are more trustworthy. In this analysis, we provide results for the case in which redshift is restricted to  $z < 2$  and compare it to the case in which  $z < 5$ .



**Figure 3.** Redshift distributions using a redshift bin size of  $\Delta z = 0.25$  for the SFG, FR I, and FR II populations using the  $S^3$  simulation.

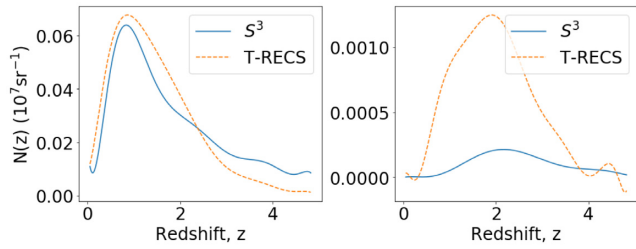


**Figure 4.** Same as Fig. 3 using the T-RECS simulation.

## 4.2 Redshift distributions

Another requirement for this analysis was redshift distributions of the galaxy populations over the redshift range of interest. Due to a lack of observed radio data at the flux limits expected for SKA (the VLA COSMOS data from Smolčić et al. 2017b is limited to  $22 \mu\text{Jy}$  at 1.4 GHz), the redshift distributions obtained from SKA simulations, namely, the  $S^3$  and T-RECS simulations were used. The MLAGN/FR I, HLAGN/FR II and SFG galaxy populations were extracted, restricted to a flux limit of  $5 \mu\text{Jy}$  (a conservative flux limit for SKA Phase 1), a redshift  $z < 5$  and then rescaled to 1 steradian of sky. The SFG population for the  $S^3$  simulation was obtained by combining the SFG and SB galaxy populations (T-RECS has one population including both of these galaxy types). The numbers of galaxies with different redshifts were then plotted in histograms using bins of size  $\Delta z = 0.25$  and functional forms were fitted. The resulting redshift distributions can be seen in Figs 3 and 4 and comparisons of the FR I and FR II distributions are shown in Fig. 5. Clearly, the T-RECS simulation predicts significantly higher numbers of SFGs than  $S^3$  and these numbers seem to be more reasonable, as observations (Smolčić et al. 2017b) have suggested that  $S^3$  underestimates SFGs (see also Bonaldi et al. 2016). A larger number of FR II/HLAGN was predicted by T-RECS while similar numbers of FR I/MLAGN were predicted by both simulations.

Unfortunately, since redshifts cannot be obtained from radio continuum surveys as the synchrotron radiation is featureless, it might be difficult to obtain detailed redshift distributions for large numbers of radio galaxies over large areas of sky. The emergence of large optical/near-IR surveys that will have accurate photo- $z$ 's and overlapping sky coverage will mitigate this problem. On the



**Figure 5.** Comparison of the redshift distributions for the FR I/MLAGN (left) and FR II/HLAGN (right) populations using the  $S^3$  and T-RECS simulations. A redshift bin size of  $\Delta z = 0.25$  is used.

other hand, some smaller, deep fields such as those used by the MIGHTEE survey (Jarvis et al. 2016) will soon have all the radio continuum, redshift, and multiwavelength data necessary for such an analysis, albeit on much smaller scales. These surveys can be used to inform the key observations necessary for the large-scale  $f_{\text{NL}}$  studies.

The accuracy of the photometric redshifts of the survey used for cross-matching must also be sufficiently high to produce these redshift distributions, although as only distributions are necessary, precise point estimates are less important. A number of methods have been proposed for estimating redshift distributions and significant improvements on accuracy and error estimation have been made over the last few years (Almosallam, Jarvis & Roberts 2016; Gomes et al. 2018; Soo et al. 2018; Sánchez & Bernstein 2019). In addition, apart from standard template fitting, machine learning, or hybrid methods, there are now techniques for determining the redshift distributions directly (without the point estimates). Two such methods are clustering redshifts (Newman 2008; Matthews & Newman 2010; Ménard et al. 2013; Schmidt et al. 2013) and using magnitude space reweighting technique (Lima et al. 2008), both of which require a spectroscopic reference set. Most of these methods still require cross-matching the radio galaxies to an optical/near-IR survey, but the clustering redshift method does not require this as it involves the spatial cross-correlation between the galaxies with unknown redshifts and a distribution with known redshifts. This method has been shown to produce reasonable redshift distributions for radio galaxy surveys (Ménard et al. 2013). The accuracy of the  $N(z)$  distribution will become increasingly important as smaller binning schemes are chosen for analyses such as these. The redshift bin sizes of  $\Delta z = 0.25$  and  $\Delta z = 1$  were used in this analysis.

An alternative option is to use the spectroscopic HI survey of the SKA but this is limited to lower redshifts  $\sim 2$  and would only detect a limited sample of the SFGs. Thus, it could provide a more reliable redshift distribution for the SFGs but would contribute little to the AGN number distributions.

### 4.3 Fisher analysis

Fisher matrix analyses (Tegmark, Taylor & Heavens 1997) allow the determination of the minimum errors on parameters that are obtainable using a particular method and experiment. This is possible because the precision to which a parameter can be determined depends on the dependence of the model on the parameter (the greater the model dependence, the greater the constraining ability) and on the covariance of the observed data (the greater the independence of the observed data, the more information can be extracted). In order to obtain the Fisher matrix for multiple tracers,

**Table 1.** Fiducial values chosen for the parameters in this model. The masses are in units of  $M_{\odot} h^{-1}$ , where  $h = H_0/100$ .

Parameter	Fiducial value	
$\Omega_{\text{CDM}}$	0.26	
$H_0$	67.74	
$\ln(10^{10} A_s)$	3.064	
$f_{\text{NL}}$	0	
	Hale halo masses	$S^3$ halo masses
$M_{\text{cent}}^{\text{SFG}}$	$4 \times 10^{12}$	$1.1 \times 10^{12}$
$M_{\text{cent}}^{\text{MLAGN}}$	$3.5 \times 10^{13}$	$1 \times 10^{13}$
$M_{\text{cent}}^{\text{HLAGN}}$	$1.5 \times 10^{13}$	$1 \times 10^{14}$
$M_{\text{cent}}^{\text{AGN}}$	$3.46 \times 10^{13}$	$1.01 \times 10^{14}$

in different redshift bins, we use the following expression:

$$F_{\alpha\beta} = \sum_{l=l_{\text{min}}}^{l_{\text{max}}} \frac{2l+1}{2} f_{\text{sky}} Tr \left( \frac{\partial \mathbf{C}(l)}{\partial \theta_{\alpha}} (\Gamma_l)^{-1} \frac{\partial \mathbf{C}(l)}{\partial \theta_{\beta}} (\Gamma_l)^{-1} \right), \quad (8)$$

where  $\mathbf{C}(l) = [C_l^{ij}]$  (defined in equation 3) and  $\theta_{\alpha}$  and  $\theta_{\beta}$  are any two model parameters.  $f_{\text{sky}}$  is the fraction of the sky expected to be covered by the radio survey, for this analysis we use a survey size of  $2\pi$  steradian as we expect the SKA to cover at least half of the sky and therefore we use a value of  $f_{\text{sky}} = 0.5$ .  $\Gamma$  is the covariance matrix and if it is assumed that the cosmic variance and shot noise are the dominant contributors to the noise (the experimental systematics are subdominant) then it is given by

$$\Gamma_l^{ij} = C_l^{ij} + \delta^{ij} N^{ii}, \quad (9)$$

where  $N^{ii}$  is the noise power spectrum for a given redshift bin and galaxy population  $i$  and is given by  $1/n^i$ , where  $n^i$  is the number of sources per steradian. These values of  $n^i$  were obtained using the scaled redshift distributions discussed above.

The bias and number distributions were then used to calculate the auto- and cross-correlation power spectra using equations (3) and (4). For this computation, a modification of CAMB sources (Challinor & Lewis 2011) presented by Fonseca et al. (2015) was used. This code allows the user to input the redshift distributions and biases of the various tracers. In the most ideal case, we split the redshift distribution into top-hat redshift bins of width 0.25 (a conservative size as we expect the photo- $z$  estimates/spectroscopic redshifts to be much more precise than this) leading to a total of 20 windows for each of the three galaxy populations. The fiducial parameters were then inputted and power spectra were obtained with  $l_{\text{min}} = 2$  and  $l_{\text{max}} = 200$  (as in Ferramacho et al. 2014). The cosmological parameters that are most degenerate with the non-Gaussianity signal are dark matter density  $\Omega_{\text{CDM}}$ , the Hubble parameter  $H_0$ , and the amplitude of the primordial fluctuations  $A_s$  and therefore these are included in the Fisher analysis. We use the parameters obtained by Planck 2015 (Planck Collaboration XVII 2016) with CMB temperature and polarization power spectra as well as lensing and external (baryon acoustic oscillations, supernova joint light-curve analysis and  $H_0$ ) measurements. The other cosmological parameters were used in producing the angular power spectra. The parameters used for this analysis (the cosmological parameters and the estimated central masses of the galaxy populations) and their fiducial values are given in Table 1.

The next step was to obtain the derivatives of these power spectra with respect to each parameter. Some of the derivatives were computed analytically while others were computed numerically

using a five-point stencil method. The covariance matrix,  $\Gamma$ , was then computed using the  $C_l$  results with the fiducial cosmology according to equation (9). The Fisher matrix was then computed using equation (8). The constraint on the  $f_{\text{NL}}$  parameter was obtained by inverting the Fisher matrix and extracting the  $f_{\text{NL}} \times f_{\text{NL}}$  component.

## 5 SKA FORECASTS – RESULTS AND ANALYSIS

In order to understand the relationships between the constraints that can be obtained with differently sized redshift bins (corresponding to photometric redshift quality) and different redshift ranges (which will depend on the performance of the SKA as well as the availability of multiwavelength data with photometric redshifts to cross-match to the radio data), we considered a number of cases:

- (i) Redshift bins of width 0.25 and a range of  $0 < z < 5$ .
- (ii) Redshift bins of width 1 and a range of  $0 < z < 5$ .
- (iii) Redshift bins of width 0.25 and a range of  $0 < z < 2$ .

The first of these cases represents the most optimistic view, implying that there will be complete and accurate photometric redshift coverage to match to the radio galaxies up to a redshift of  $z \sim 5$ . This is unlikely to be the case due to obstacles such as dust obscuration at high redshifts, which could introduce artificial selection effects (with brighter objects more likely to be detected) and impact our measurements of  $f_{\text{NL}}$ . In addition, there are limited spectroscopic redshifts at high redshifts, making the training of machine learning photometric redshift estimation methods as well as the redshift distribution estimation methods more difficult and less accurate. For these reasons, we also evaluated the case in which we are able to get redshifts up to 5 but we have much lower photometric redshift accuracy and therefore we use much larger bins of width 1. Finally, we also consider the much more realistic case of obtaining galaxies with photometric redshifts only up to a redshift of 2. Below this redshift it is more likely that multiwavelength data, and correspondingly photometric redshifts will be available for cross-matching to most of the radio sources. In addition, in this redshift range it is also possible to obtain some redshifts spectroscopically from HI line measurements. Photometric redshift accuracy is also expected to be higher in this redshift range, making the bin width of 0.25 a conservative choice. In addition to obtaining better redshifts below  $z \sim 2$ , the bias functions measured/simulated are also more accurate below these redshifts. As a result, constraints obtained for this case of restricted redshift are much more realistic than other cases.

For each of these cases, we evaluated the constraints on  $f_{\text{NL}}$  with the different combinations of either the  $S^3$  or Hale et al. (2018) biases and the  $S^3$  or T-RECS  $N(z)$  distributions in order to obtain a range of constraints depending on the range of possibilities for the radio source populations. We note that combining the  $N(z)$  and the biases, from different studies and techniques, in this way could be viewed as internally inconsistent. However, we believe that this gives an indication of how the different assumed distributions affect the results. We also compared the result of having three independent distributions: SFG + SB, MLAGN/FR I, and HLAGN/FR II to the case in which MLAGN/FR I and HLAGN/FR II are not distinguished and we have two populations: SFG + SB and AGNs. The morphology of radio galaxies is not a discrete distribution of FR I and FR II objects, but a continuous distribution of morphologies, with many objects that are difficult to classify as either an FR I or an FR II. This distinction also becomes increasingly difficult as redshift

**Table 2.** Forecasts on  $f_{\text{NL}}$   $1\sigma$  errors obtained using a multitracer Fisher analysis with redshift bins of size  $\Delta z = 0.25$  and a range of  $0 < z < 5$ .

	$S^3 N(z)$	T-RECS $N(z)$
$S^3$ biases	1.90	1.88
Hale biases	1.54	1.56

increases and the images obtained resolve less structure. Therefore, for the most conservative case we assume that we are not able to make this distinction. We also observed that the number density of the HLAGNs in both redshift distributions was relatively low in comparison to the other populations and therefore we decided to compare the constraints obtained when we use just SFG + SB and MLAGN + SB and AGNs (where the HLAGNs is included in the AGNs).

The constraints measured for each of the cases presented above are shown in Tables 2, 3, and 4. We found that for all cases (of redshift bin size and range as well as combinations of galaxy types) the biases from Hale et al. (2018) lead to significantly tighter constraints than those from  $S^3$ , while the different redshift distributions of the two simulations lead to less notable differences. This means that if the biases calculated by Hale et al. (2018) are nearer to the truth, tighter constraints than previously expected (Ferramacho et al. 2014) will be possible with this method. The Hale et al. (2018) biases may have caused this effect because the separation in the biases of the FR I and SFGs is much greater than for  $S^3$  (compare Figs 1 and 2) and the numbers of SFG and FR I are much greater than the number of FR II present. Therefore, although the difference between the bias functions of the populations is greater for  $S^3$ , with the FR II galaxies having very high biases, the very limited number of FR IIs and larger numbers of FR Is and SFGs lead to the Hale et al. (2018) biases producing tighter constraints. We note that ideally, the  $N(z)$  distribution parameters would be marginalized over in our analysis. However, the differences in the constraints caused by the  $N(z)$  distributions is subdominant to the differences due to the biases using the  $S^3$  simulation and Hale et al. (2018). Therefore, the fact that the number distributions have substantially different distributions (see Figs 3 and 4) implies that small variations in the number distributions will not result in notable differences in constraints.

We observed that the effect of the  $S^3$  versus the T-RECS redshift distributions varied across the redshift bin size and range. For the case with  $z < 2$  (Table 4), the T-RECS number distributions produced significantly tighter constraints while for the case of larger redshift bins but a range of  $0 < z < 5$  (Table 3), the  $S^3$  number distributions resulted in tighter constraints. This pattern is likely because the numbers of MLAGN in the T-RECS simulation tapers off to much lower numbers beyond  $z = 2$ . This causes the MLAGN in the  $S^3$  simulation to contribute more to the  $0 < z < 5$  measurement than those in the T-RECS simulation, while the MLAGN in the T-RECS simulation contributes more to the  $z < 2$  measurement than those in the  $S^3$  simulation. On the other hand, in the most ideal case with small redshift bins and a large redshift range, the differences induced by the different redshift distributions are insignificant.

Excluding the HLAGN (see Table 3) and joining the MLAGN and HLAGN (Tables 3 and 4) both reduced the constraints (increased the  $1\sigma$  errors) by small but notable amounts with the combined MLAGN and HLAGN performing marginally better than removing the HLAGN entirely. This is expected as the HLAGN population is

**Table 3.** Forecasts on  $f_{\text{NL}}$   $1\sigma$  errors obtained using a multitracer Fisher analysis with redshift bins of size  $\Delta z = 1$  and a range of  $0 < z < 5$ .

		SFG + SB MLAGN HLAGN	SFG + SB MLAGN	SFG + SB AGNs
$S^3$ biases	$S^3 N(z)$	2.23	2.27	2.26
	T-RECS $N(z)$	2.35	2.47	2.46
Hale biases	$S^3 N(z)$	1.83	1.84	1.84
	T-RECS $N(z)$	1.98	1.99	1.98

**Table 4.** Forecasts on  $f_{\text{NL}}$   $1\sigma$  errors obtained using a multitracer Fisher analysis with redshift bins of size  $\Delta z = 0.25$  and a range of  $0 < z < 2$ .

		SFG + SB MLAGN HLAGN	SFG + SB AGNs
$S^3$ biases	$S^3 N(z)$	6.53	6.58
	T-RECS $N(z)$	5.39	5.82
Hale biases	$S^3 N(z)$	4.61	4.63
	T-RECS $N(z)$	4.07	4.08

very small and therefore does not contribute much to the constraints when added to the MLAGN. The addition of the HLAGN does increase the numbers to a small extent, which decreases the shot noise, while also contributing a greater ratio of higher redshift galaxies, these effects lead to the marginal improvement observed. On the other hand, removing the HLAGN as an independent tracer does decrease the constraints notably as there are now only two tracers with different biases to be compared. This reduction in constraining power was found to be much larger for the combination of the  $S^3$  biases and the T-RECS  $N(z)$  distributions than for the other cases and this is due to the very large bias values assigned to the HLAGN by the  $S^3$  simulation coupled with the larger numbers of HLAGN predicted by the T-RECS simulation. Another important observation is that the increases in errors imposed by the removal of the HLAGN as an independent tracer are much less significant than the variation between the bias and number distribution configurations. This is clearly observed in Tables 3 and 4 and proves that the SFG and MLAGN populations are contributing the majority of the constraining power. This implies that if image resolution or multiwavelength data are not sufficient for distinguishing between MLAGN/FR I and HLAGN/FR II then this is not a major problem as it will not significantly decrease the possible constraints.

When comparing the cases with different redshift bin widths, we find that increasing the bin size increased the  $1\sigma$  errors on  $f_{\text{NL}}$  on all the bias and  $N(z)$  configurations by a factor less than 1.5 (compare Table 2 to Table 3) while reducing the redshift range to  $z < 2$  had a much more significant impact with a factor nearer to 3 (see Table 2 and Table 4). This implies that while having accurate photometric redshift estimates that allow smaller bins is important it is not as vital to adding constraining power as having deeper surveys with galaxies and cross-matched photometric redshifts going to higher redshifts. The reduced redshift led to more significant reductions in constraining power because the volume of the survey is decreased, thus reducing the number of large-scale measurements contributing to the angular power spectra. Despite this, as previously discussed, we do not expect to have these data at  $z \gtrsim 2$  in the near future, and the bias functions have not been measured at higher redshifts therefore we take the results in Table 4 as the expected limits of constraints with the SKA Phase 1 continuum surveys. Note here that our choice of  $f_{\text{sky}} = 0.5$  was conservative for the SKA and it is possible that greater coverage will be obtained, as the errors scale with  $\sqrt{f_{\text{sky}}}$  this will lead to tighter constraints. Also note that these

Fisher matrix constraints are the minimum possible errors and that their calculation assumes that any systematic errors inherent in the experiment have a much smaller impact than the cosmic variance or shot noise contributions.

It should be noted that this was a conservative analysis, using only the three or two galaxy populations that are most likely to be correctly distinguished. If the multiwavelength data available become sufficient to distinguish SBs from the other SFGs and to similarly differentiate the radio quiet quasars then additional tracers can be utilized and the results will be further improved, as evaluated in Ferramacho et al. (2014).

## 6 CONCLUSIONS

In this paper, we have presented an analysis of the constraints on PNG that could be obtained using the SKA radio continuum surveys. This paper serves as an update to the analysis undertaken by Ferramacho et al. (2014). We investigated the possibility of using multiple radio galaxy populations, with different halo mass and bias properties to perform a multitracer analysis, thus reducing the effect of cosmic variance and increasing the constraining power that can be obtained. The galaxy populations we considered were SFGs, MLAGNs (or FR Is), and HLAGNs (or FR IIs). We consider the MLAGNs and HLAGNs to be similar enough to the FR I and FR II populations that we can use the classifications interchangeably based on the simulation/observations utilized. In contrast to Ferramacho et al. (2014), we do not distinguish SFGs and SBs as this may not be observationally feasible at the required level of precision.

We perform a Fisher matrix analysis following the formalism of Ferramacho et al. (2014), which utilizes the auto- and cross-correlation angular power spectra of the multitracers. These power spectra require the redshift and bias distributions of the radio galaxy populations. We use the redshift distributions of radio galaxies populations given by two SKA simulations: the  $S^3$  simulation and the more recent T-RECS simulation and we rescaled them to the appropriate area and flux limit. For the bias distributions, we used the average halo masses obtained from Hale et al. (2018) who use VLA COSMOS-field observations as well as those estimated by  $S^3$ . We then assumed a Gaussian distribution around this mass and calculated the bias using an analytical expression for the linear bias. We considered all of the combinations of the bias and redshift distributions as well as different redshift ranges (dependent on the depth



of the continuum survey as well as the availability of visible/near-IR data with reliable photometric redshifts that can be cross-matched to the radio sources) and redshift bin sizes (dependent on photometric redshift accuracy). In addition to treating the SFG, MLAGN/FR I, and HLAGN/FR II as independent populations, we also considered the combined MLAGN and HLAGN population to determine the effect on the constraints if the MLAGN and HLAGN cannot be differentiated. We take the case of small redshift bins (width = 0.25) and a small redshift range ( $0 < z < 2$ ) as the most realistic case for SKA observations and this provides a range in the  $1\sigma$  errors of 4.07–6.53 if the AGN populations are differentiated and 4.08–6.58 if they are not, the similarity between these is due to the fact that the HLAGNs are much rarer than the dominant AGN population at lower luminosities. The results with the observed biases (Hale et al. 2018) and the more recent simulations (T-RECS) provide the lower bounds of these ranges: 4.07 and 4.08 with three populations and two populations, respectively. These constraints will surpass the existing tightest constraints ( $\sim 5$  obtained with Planck 2015 and 2018 data; Planck Collaboration XVII, IX 2016, 2019) but will also provide an independent precise measure of  $f_{\text{NL}}$ , obtained from the large-scale structure instead of the CMB bispectrum. These results also indicate that if redshift information is available for the galaxy populations to higher redshifts, constraints will improve significantly, and with a redshift range of  $0 < z < 5$  the  $1\sigma$  error on  $f_{\text{NL}}$  will fall between 1.5 and 2, representing unprecedented precision on the measurement of  $f_{\text{NL}}$  and approaching the target of 1. On the other hand, some limitations of this analysis is that the masses of the galaxy populations might have non-Gaussian distributions or different standard deviations than the ones assumed in this analysis, in addition, the bias evolution used in this analysis is very uncertain at redshifts above  $z \sim 2$ . Moreover, the redshift distributions used are based on simulations, which are constrained by observations, but they still might not be very accurate, particularly at the higher redshifts where the lever arm for the constraints on  $f_{\text{NL}}$  are strongest.

It is clear that in order to use wide-area radio continuum surveys for constraining the influence of non-Gaussianity on the large-scale structure, then robust measurements of the distribution of biases for the different populations, along with their redshift distributions are needed. To some extent, this is underway, as we have used the observed constraints from the VLA-COSMOS survey in order to estimate the bias of the variety of radio source populations. However, VLA-COSMOS covers a relatively small area, and does not reach the depth of the planned ‘all-sky’ SKA continuum surveys. Therefore, the imminent MeerKAT International Giga-Hertz Tiered Extragalactic Exploration (MIGHTEE; Jarvis et al. 2016) survey will certainly provide much better information on the bias (through a clustering analysis covering many square degrees, as opposed to a single  $\sim 2$  deg field, and to a depth similar to that of the all-sky SKA survey. A key element of MIGHTEE is it also covers areas of the extragalactic sky with excellent multiwavelength data, thus more accurate redshift distributions will also be available based on the photometric redshifts in these fields (similar to VLA-COSMOS now).

The combination of these multiwavelength data, coupled with deep radio data may also allow the characterization of the sources into the populations discussed here using morphological information. Such an effort could potentially supply the necessary training sample for a variety of machine learning algorithms currently being tested in radio data (e.g. Lukic et al. 2018; Glaser et al. 2019), and then applied to wider field data where the data extent and quality are not comparable to in these deep fields.

Thus, although constraining  $f_{\text{NL}}$  requires large cosmological volumes, the information that can be gleaned from the deep narrow surveys will be crucial in planning the final strategy.

Furthermore, joint analyses with other multitracers probes (other than radio continuum) such as H I intensity mapping or low-redshift galaxy surveys would likely lead to further improvements on  $f_{\text{NL}}$  constraints.

## ACKNOWLEDGEMENTS

ZG is supported by a Rhodes Scholarship granted by the Rhodes Trust. MJJ acknowledges support from the South African Radio Astronomy Observatory through a visiting Professorship. CLH would like to acknowledge the support given from the Science and Technology Facilities Council (STFC) through an STFC studentship. SC is supported by the Italian Ministry of Education, University and Research (MIUR) through Rita Levi Montalcini project ‘PROMETHEUS - Probing and Relating Observables with Multiwavelength Experiments To Help Enlightening the Universe’s Structure’, and by the ‘Departments of Excellence 2018-2022’ Grant awarded by MIUR (L. 232/2016). JF acknowledges support from the South African Square Kilometre Array Project and National Research Foundation.

## REFERENCES

- Almosallam I. A., Jarvis M. J., Roberts S. J., 2016, *MNRAS*, 462, 726  
 Alonso D., Ferreira P. G., 2015, *Phys. Rev. D*, 92, 063525  
 Alvarez M. et al., 2014, preprint ([arXiv:1412.4671](https://arxiv.org/abs/1412.4671))  
 Ballardini M., Matthewson W. L., Maartens R., 2019, *MNRAS*, 489, 1950  
 Bartolo N., Komatsu E., Matarrese S., Riotto A., 2004, *Phys. Rep.*, 402, 103  
 Bonaldi A., Harrison I., Camera S., Brown M. L., 2016, *MNRAS*, 463, 3686  
 Bonaldi A., Bonato M., Galluzzi V., Harrison I., Massardi M., Kay S., De Zotti G., Brown M. L., 2018, *MNRAS*, 482, 2  
 Bruni M., Crittenden R., Koyama K., Maartens R., Pitrou C., Wands D., 2012, *Phys. Rev. D*, 85, 041301  
 Camera S., Santos M. G., Bacon D. J., Jarvis M. J., McAlpine K., Norris R. P., Raccanelli A., Röttgering H., 2012, *MNRAS*, 427, 2079  
 Camera S., Santos M. G., Ferreira P. G., Ferramacho L., 2013, *Phys. Rev. Lett.*, 111, 171302  
 Camera S., Maartens R., Santos M. G., 2015, *MNRAS*, 451, L80  
 Carbone C., Verde L., Matarrese S., 2008, *ApJ*, 684, L1  
 Carilli C., Rawlings S., 2004, *New Astron. Rev.*, 48, 979  
 Castorina E., Feng Y., Seljak U., Villaescusa-Navarro F., 2018, *Phys. Rev. Lett.*, 121, 101301  
 Castorina E. et al., 2019, *J. Cosmol. Astropart. Phys.*, 2019, 010  
 Challinor A., Lewis A., 2011, *Phys. Rev. D*, 84, 043516  
 Creminelli P., Zaldarriaga M., 2004, *J. Cosmol. Astropart. Phys.*, 2004, 006  
 Dalal N., Doré O., Huterer D., Shirokov A., 2008, *Phys. Rev. D*, 77, 123514  
 de Putter R., Doré O., 2017, *Phys. Rev. D*, 95, 123513  
 de Putter R., Gleyzes J., Doré O., 2017, *Phys. Rev. D*, 95, 123507  
 Dewdney P. E., Hall P. J., Schilizzi R. T., Lazio T. J. L. W., 2009, *IEEE Proc.*, 97, 1482  
 De Bernardis F., Serra P., Cooray A., Melchiorri A., 2010, *Phys. Rev. D*, 82, 083511  
 Doré O. et al., 2014, preprint ([arXiv:1412.4872](https://arxiv.org/abs/1412.4872))  
 Fanaroff B. L., Riley J. M., 1974, *MNRAS*, 167, 31P  
 Ferramacho L. D., Santos M. G., Jarvis M. J., Camera S., 2014, *MNRAS*, 442, 2511  
 Ferraro S., Smith K. M., 2015, *Phys. Rev. D*, 91, 043506  
 Fonseca J., Camera S., Santos M. G., Maartens R., 2015, *ApJ*, 812, L22  
 Fonseca J., Maartens R., Santos M. G., 2017, *MNRAS*, 466, 2780  
 Fonseca J., Maartens R., Santos M. G., 2018, *MNRAS*, 479, 3490  
 Gendre M. A., Best P. N., Wall J. V., Ker L. M., 2013, *MNRAS*, 430, 3086  
 Giannantonio T., Percival W. J., 2014, *MNRAS*, 441, L16

- Glaser N., Wong O. I., Schawinski K., Zhang C., 2019, *MNRAS*, 487, 4190
- Gomes Z., Jarvis M. J., Almosallam I. A., Roberts S. J., 2018, *MNRAS*, 475, 331
- Hale C. L., Jarvis M. J., Delvecchio I., Hatfield P. W., Novak M., Smolčić V., Zamorani G., 2018, *MNRAS*, 474, 4133
- Hardcastle M. J., 2004, *A&A*, 414, 927
- Hatfield P. W., Bowler R. A. A., Jarvis M. J., Hale C. L., 2018, *MNRAS*, 477, 3760
- Heckman T. M., Best P. N., 2014, *ARA&A*, 52, 589
- Huterer D., Knox L., Nichol R. C., 2001, *ApJ*, 555, 547
- Janssen R. M. J., Röttgering H. J. A., Best P. N., Brinchmann J., 2012, *A&A*, 541, A62
- Jarvis M., Bacon D., Blake C., Brown M., Lindsay S., Raccanelli A., Santos M., Schwarz D. J., 2015, Proc. Sci., Cosmology with SKA Radio Continuum Surveys. SISSA, Trieste, PoS# 18
- Jarvis M. et al., 2017, preprint ([arXiv:1709.01901](https://arxiv.org/abs/1709.01901))
- Jeong D., Schmidt F., Hirata C. M., 2012, *Phys. Rev. D*, 85, 023504
- Joudaki S., Doré O., Ferramacho L., Kaplinghat M., Santos M. G., 2011, *Phys. Rev. Lett.*, 107, 131304
- Karagiannis D., Lazanu A., Liguori M., Raccanelli A., Bartolo N., Verde L., 2018, *MNRAS*, 478, 1341
- Komatsu E. et al., 2003, *ApJS*, 148, 119
- Leistedt B., Peiris H. V., Roth N., 2014, *Phys. Rev. Lett.*, 113, 221301
- Li Y.-C., Ma Y.-Z., 2017, *Phys. Rev. D*, 96, 063525
- Lima M., Cunha C. E., Oyaizu H., Frieman J., Lin H., Sheldon E. S., 2008, *MNRAS*, 390, 118
- Lindsay S. N. et al., 2014, *MNRAS*, 440, 1527
- LSST Science Collaboration, 2009, preprint ([arXiv:0912.0201](https://arxiv.org/abs/0912.0201))
- Lukic V., Brüggen M., Banfield J. K., Wong O. I., Rudnick L., Norris R. P., Simmons B., 2018, *MNRAS*, 476, 246
- Lyth D. H., Ungarelli C., Wands D., 2003, *Phys. Rev. D*, 67, 023503
- Magliocchetti M., Popesso P., Brusa M., Salvato M., Laigle C., McCracken H. J., Ilbert O., 2017, *MNRAS*, 464, 3271
- Maldacena J., 2003, *J. High Energy Phys.*, 2003, 013
- Matarrese S., Verde L., 2008, *ApJ*, 677, L77
- Matthews D. J., Newman J. A., 2010, *ApJ*, 721, 456
- Ménard B., Scranton R., Schmidt S., Morrison C., Jeong D., Budavari T., Rahman M., 2013, preprint ([arXiv:1303.4722](https://arxiv.org/abs/1303.4722))
- Mueller E.-M., Percival W. J., Ruggeri R., 2019, *MNRAS*, 485, 4160
- Murray S. G., Power C., Robotham A. S. G., 2013, *Astron. Comput.*, 3, 23
- Newman J. A., 2008, *ApJ*, 684, 88
- Planck Collaboration XVII, 2016, *A&A*, 594, A17
- Planck Collaboration IX, 2019, preprint ([arXiv:1905.05697](https://arxiv.org/abs/1905.05697))
- Quinn P., Axelrod T., Bird I., Dodson R., Szalay A., Wicenec A., 2015, Proc. Sci., Delivering SKA Science. SISSE, Trieste, PoS# 147
- Ramos Almeida C., Bessiere P. S., Tadhunter C. N., Inskip K. J., Morganti R., Dicken D., González-Serrano J. I., Holt J., 2013, *MNRAS*, 436, 997
- Ross A. J. et al., 2013, *MNRAS*, 428, 1116
- Sánchez C., Bernstein G. M., 2019, *MNRAS*, 483, 2801
- Santos M. et al., 2015, Proc. Sci., Cosmology from a SKA HI Intensity Mapping Survey. SISSE, Trieste, PoS# 19
- Schmidt S. J., Ménard B., Scranton R., Morrison C., McBride C. K., 2013, *MNRAS*, 431, 3307
- Seljak U., 2009, *Phys. Rev. Lett.*, 102, 021302
- Sheth R. K., Tormen G., 1999, *MNRAS*, 308, 119
- Slosar A., Hirata C., Seljak U., Ho S., Padmanabhan N., 2008, *J. Cosmol. Astropart. Phys.*, 8, 031
- Smolčić V. et al., 2017a, *A&A*, 602, A2
- Smolčić V. et al., 2017b, *A&A*, 602, A6
- Soo J. Y. H. et al., 2018, *MNRAS*, 475, 3613
- Square Kilometre Array Cosmology Science Working Group, 2018, preprint ([arXiv:1811.02743](https://arxiv.org/abs/1811.02743))
- Tasse C., Best P. N., Röttgering H., Le Borgne D., 2008, *A&A*, 490, 893
- Tegmark M., Taylor A. N., Heavens A. F., 1997, *ApJ*, 480, 22
- Tellarini M., Ross A. J., Tasinato G., Wands D., 2016, *J. Cosmol. Astropart. Phys.*, 6, 014
- The Dark Energy Survey Collaboration, 2005, preprint ([astro-ph/0510346](https://arxiv.org/abs/astro-ph/0510346))
- Uhlemann C., Pajer E., Pichon C., Nishimichi T., Codis S., Bernardeau F., 2018, *MNRAS*, 474, 2853
- Urry C. M., Padovani P., 1995, *PASP*, 107, 803
- Welling Y., van der Woude D., Pajer E., 2016, *J. Cosmol. Astropart. Phys.*, 8, 044
- Wilman R. J. et al., 2008, *MNRAS*, 388, 1335
- Witzemann A., Alonso D., Fonseca J., Santos M. G., 2019, *MNRAS*, 485, 5519
- Xia J.-Q., Viel M., Baccigalupi C., De Zotti G., Matarrese S., Verde L., 2010, *ApJ*, 717, L17
- Yamauchi D., Takahashi K., Oguri M., 2014, *Phys. Rev. D*, 90, 083520
- Yoo J., 2010, *Phys. Rev. D*, 82, 083508
- Zaldarriaga M., 2004, *Phys. Rev. D*, 69, 043508

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