



***Facultad
de
Ciencias***

**DETECTION OF THE KINEMATIC
DIPOLE WITH THE LARGE-SCALE
STRUCTURE OF THE UNIVERSE**
Detección del dipolo cinemático con la
estructura a gran escala del Universo

Trabajo de Fin de Grado
para acceder al

GRADO EN FÍSICA

Autor: Ixaka Labadie Garcia

Director: Patricio Vielva

Septiembre-2020

Abstract

The Universe according to the cosmological standard model is homogeneous and isotropic at very large scales. Since the Solar System is in motion with respect to the rest frame of the large-scale structure of the Universe a dipole signal is expected. The most paradigmatic case is that of the cosmic microwave background (CMB). The CMB dipole has been studied in many ways over the past decades and has been estimated with high accuracy, notably by the ESA Planck mission. However, tracers of the large-scale structure, as radio galaxy surveys can give a complementary evaluation of the dipole from the point of view of additional cosmological probes, helping, both, to corroborate the CMB findings or, eventually, to spot possible departures from isotropy, as reported by some observations. The aim of this project is to estimate the power of the Square Kilometer Array (SKA) telescope, to be operational around 2028, to detect the kinematic dipole. A modulation equivalent to that seen in the CMB has been applied to a set of simulations of the large-scale structure and a method for its measurement is proposed. The dipole signal obtained with SKA in the best case scenario has an amplitude of $(4.82 \pm 0.23) \cdot 10^3$ in the direction $(l, b) = (268.47 \pm 5.84, 40.93 \pm 3.41)^\circ$. The amplitude and the longitude are in good agreement with the one determined with the CMB, but the latitude shows a deviation. This is caused by the specific geometry of the mask accounting for the expected sky observed by SKA, and, in principle, it is a systematic effect that can be corrected. Still, the precision of the measurements exceeds others obtained with current surveys proving that SKA can become a very important cosmological probe.

Keywords: cosmology, kinematic dipole, radio galaxy surveys, CMB

Resumen

El universo según el modelo cosmológico estándar es homogéneo e isótropo a grandes escalas. Debido a que el sistema solar esta en movimiento respecto al sistema en reposo de la estructura a gran escala del universo se espera encontrar una señal dipolar. El caso más paradigmático es el del fondo cósmico de microondas (CMB). El dipolo del CMB ha sido estudiado de distintas maneras durante las ultimas décadas y se ha logrado medir con gran precisión, especialmente por la misión Planck de la ESA. Trazadores de la estructura a gran escala le pueden dar otro enfoque al tema desde el punto de vista de otros observables. Este proyecto pretende estimar la capacidad de detectar el dipolo cinemático de el telescopio Square Kilometer Array (SKA), previsto para hacia 2028. Se ha aplicado una modulación equivalente a la observada en el CMB a un conjunto de simulaciones de la estructura a gran escala y se propone un método para su medida. La señal dipolar obtenida con SKA en el mejor caso posible tiene una amplitud de $(4,82 \pm 0,23) \cdot 10^3$ en dirección a $(l, b) = (268,47 \pm 5,84, 40,93 \pm 3,41)^\circ$. La amplitud y la longitud son compatibles con los valores obtenidos del dipolo del CMB, pero la latitud muestra una desviación. La causa de esto es la geometría de la mascara que representa el cielo observado por SKA y, en principio, es un efecto sistemático que puede ser corregido. Aun así, la precisión de las medidas es considerablemente mejor que las tomadas con catálogos actuales, demostrando que SKA se puede convertir en un instrumento cosmológico muy importante.

Palabras clave: cosmología, dipolo cinemático, catálogo de radio-galaxias, CMB

A Patricio, familia y amigos

Contents

1. Introduction	1
1.1. Dipole from the CMB	2
1.2. Kinematic dipole from the large-scale structure	4
1.3. Square Kilometer Array	5
1.3.1. SKA1-MID	7
1.3.2. SKA1-LOW	7
1.3.3. Cosmological surveys	7
2. Simulations	9
2.1. Power spectra, source distribution and galaxy bias	9
2.2. FLASK	14
2.3. Mask and HEALPix	16
3. Optimization	18
3.1. Nelder-Mead	19
3.2. Newton's Conjugate Gradient	20
3.3. L-BFGS-B	21
3.4. Our minimization approach	22
4. Results and Analysis	25
5. Conclusions and future work	34
References	36
A. Additional figures	38

List of Figures

1.1. Exaggerated modulated dipole map	5
2.1. Angular power spectrum	11
2.2. Polynomial fits	12
2.3. Clustering bias $b(z)$ over redshift	12
2.4. HEALPix pixel resolution examples	17
3.1. Optimization method flow-chart	22
3.2. $N_{side} = 4$ pixel direction estimations	23
4.1. Simulation map examples	26
4.2. Amplitude Histograms masked-sky	29
4.3. Direction plots masked-sky: 10 μJy , 22.8 μJy	30
4.4. Difference angle θ histograms masked-sky	31
4.5. Correlation plots $S > 22.8 \mu\text{Jy}$	32
A.1. Direction plots full-sky	38
A.2. Direction plots masked-sky: 1 μJy , 5 μJy	39
A.3. Amplitude Histograms full-sky	39
A.4. More $N_{side} = 4$ pixel direction estimations	40
A.5. Longitude l histograms masked-sky	41
A.6. Latitude b histograms masked-sky	42
A.7. Longitude l histograms full-sky	43
A.8. Latitude b histograms full-sky	44
A.9. Difference angle θ histograms full-sky	45
A.10. All Correlation plots	46

Chapter 1

Introduction

The Universe is characterized by cosmological observables that give information about its origin and evolution, as well as its energetic content and of its structure. The cosmic microwave background (CMB) radiation and the large-scale structure of the Universe are prominent observables and have given shape to the standard cosmological model, the so-called Λ CDM (Lambda Cold Dark Matter) model. It describes a universe with flat spacial geometry and in with a current accelerated expansion, formed by three components. The biggest fraction of the energy density is that of the vacuum energy or dark energy, accounting for 69%. Dark energy is believed to be responsible of the accelerated expansion and even if its nature is still unknown, it is seen as a "fluid" with an equation of state of pressure with opposite sign to its density ($p \propto \rho$). Most recent data has proved dark energy to be compatible with a cosmological constant Λ (i.e., $p = -\rho$) which is a key part of the model. The remaining 31% is divided between baryonic matter or "ordinary" matter (5%) and a weakly interactive dark matter (26%). The latter is called cold dark matter (CDM) in the model because it is hypothesized to have low velocity; it is the major responsible for the gravitational potential that dominates the formation of the large-scale structure of the Universe.

The origin of the Universe is given by the cosmological inflation theory. It considers an almost exponential expansion at early times until becoming more stable and reaching the actual state of the Universe. Inflation creates the initial perturbations in matter density that are the seed of the large-scale structure. The significance of this theory is great, because its standard variant implies in a simple way that the Universe must be isotropic and homogeneous, hence, the cosmological principle would be a consequence of the inflationary mechanism and not to be taken just as a guess.

The fulfillment of the cosmological principle is a condition for the metrics that rule Λ CDM to be valid; It states that the Universe is homogeneous and isotropic at large scales, i.e. that it has the same properties at any point and that it has no preferred directions. If this requirement is met, the effect of all forces would be expected to be the same throughout the Universe, and an observer at rest in the co-moving coordinate system of the Universe would see the same distribution in the sky at any frequency. Until now most observations have shown agreement with this principle and corroborate the current standard model, but an opposite result could mean a big change in the model is needed. In fact, there have been some claims of anomalies in the analysis of

the CMB at large scales that, with a significance of 3σ , could mean a violation of the isotropy, e.g. [Planck Collaboration et al. \(2020\)](#). There have been many studies that analyze the solutions of Einstein’s field equations for an anisotropic universe. Instead of the Friedmann–Lemaître–Robertson–Walker metrics used in Λ CDM, the premise of anisotropy leads to different solutions called Bianchi metrics and other cosmological models.

The goal of this thesis is to determine the power of tracers of the large-scale structure, in particular, radio catalogues that could be obtained by the incoming SKA telescope to estimate the cosmic dipole and to compare its precision it with results from the CMB dipole. For that, different optimization methods have been tested with simulations of different radio sources catalogues. In particular, the effect of the sky coverage, the flux threshold and the local dipole induced by the nearby structure has been studied to see how they influence the results and how they could be more accurate. Since the simulations follow the standard model of cosmology the values obtained from them are to be expected of the estimations from SKA surveys, thus, corroborating the model if the results are compatible or refuting it and making way for new models if otherwise.

1.1. Dipole from the CMB

The cosmic microwave background radiation (CMB) is one of the most important instruments to obtain information about the models that characterize our universe and has recently been used to define the parameters of Λ CDM with great precision, see [Planck Collaboration et al. \(2018b\)](#). Also, it is one of the cosmological observables where the isotropy is most clearly seen and because of this the CMB rest frame is the favored frame of reference in cosmology. The CMB spectrum is considered to be that of a *black body*, with temperature $T = (2.72548 \pm 0.00057)$ K, determined by the analysis of various experiments ([Fixsen, 2009](#)). This signal was created around 380,000 years after the big bang when photons became free to travel through the Universe, this is known as the time of last scattering.

Before last scattering, there were inflation fluctuations, and because of that some anisotropies are expected in the CMB; these are called primary fluctuations. Some reasons are the Doppler effect due to velocity changes in the plasma at last scattering and *Sachs-Wolfe effect*. There are other anisotropies with origin in the later universe, these are secondary fluctuations. They come from the interaction, be it gravitational, lensing or scattering with matter, between the CMB radiation and the large-scale structure of the Universe. For more information on this topic see, e.g. [Weinberg \(2008\)](#). Both primary and secondary anisotropies are of the order of $\Delta T/T = 10^{-5}$ or smaller, however, there is another source much stronger than the previously mentioned, of the order of 10^{-3} , which is a dipole signal created by the movement of the Earth (or the Solar System) respect to the rest frame of the CMB. The dipole that will be studied in this work is the kinematic dipole but there are other contributions to the observed dipole. One of them is the random intrinsic cosmological dipole. The cosmological dipole is believed to be of the same order than the rest of the anisotropies and, therefore, smaller than the kinematic. Finally, it could be also possible to have an

intrinsic cosmic dipole in the context of an anisotropic universe, with an amplitude that would depend on the specific model. The kinematic dipole is generated by Doppler boosting of the CMB monopole and amplifies the signal in its direction and reduces it in the opposite way. Additional dipole effects on the CMB anisotropies are also generated by the movement of the Solar System with respect to the CMB rest of frame, in particular a *dipole modulation* of the CMB anisotropies and a *dipole aberration*, which changes the arrival direction of CMB photons towards the direction of our motion. The best measurements for the CMB dipole is obtained from the direct estimation of the dipole modulation of the CMB monopole, reporting an amplitude $A_{CMB} = (3362.08 \pm 0.99) \mu\text{K}$ towards the direction $l = (264.021 \pm 0.011)^\circ$ and $b = (48.253 \pm 0.005)^\circ$ in galactic coordinates, which implies a velocity $v = (369.82 \pm 0.11) \text{ km s}^{-1}$ (Planck Collaboration et al., 2018a). *Planck* also reported complementary estimates of the kinematic dipole from the *modulation* effect with $v = (384 \pm 78) \text{ km s}^{-1}$ towards a direction compatible with the previous ones, but with an uncertainty of many degrees Planck Collaboration et al. (2014).

Complementary measurements from other observables such as the modulation of current large-scale structure tracers do not give more precise results, but, in principle, it could provide an extra piece of valuable information to confirm our global picture of the standard cosmological model. Our study focused, precisely, in forecasting the capabilities of future large-scale tracers (as SKA) to provide this information.

Following Rubart & Schwarz (2013), we define the dipole vector:

$$\mathbf{d}_{CMB} = \mathbf{d}_{motion} + \mathbf{d}_{primordial} + \mathbf{d}_{ISW} + \mathbf{d}_{foregrounds} + \mathbf{d}_{noise}, \quad (1.1)$$

where, as explained earlier, $\mathbf{d}_{primordial}$ is 10^2 times smaller than \mathbf{d}_{motion} ; \mathbf{d}_{ISW} , which is an effect of gravity on radiation is of the order of $\Delta T/T \approx 10^{-4}$; the foregrounds (caused from physical processes originated in the Milky Way) are considered to be controlled; and for moderns experiments like *Planck*, their sensitivities are good enough to guarantee that the noise is negligible at very large angular scales. This would only leave the contribution of the Doppler boosting. Even if the most recent results seem to agree with the model it is not yet clear whether the dipole has other contributions or not so other types of surveys could be helpful for a deeper understanding, for example of the dipole anomalies at large scales of the CMB with *Planck*.

The cosmological principle suggests that the same dipole as in the CMB should be seen in the matter distribution of the Universe, at sufficiently large scales, i.e., when a $\mathbf{d}_{structure}$ sourced by the nearby Universe is removed. In this case it would have the form:

$$\mathbf{d}_{radio} = \mathbf{d}_{motion} + \mathbf{d}_{structure} + \mathbf{d}_{foregrounds} + \mathbf{d}_{noise}. \quad (1.2)$$

The structure dipole has its origin in the structure of the Universe. The distribution of sources follows a Poisson distribution, for this reason, a dipole contribution due to Poisson noise or shot noise is expected, and it should be random. This effect should be more important in the local structure because it is the closest to us and has had more interactions. Since at $z \ll 1$ the galaxy distribution is dominated by the large-scale structure, radio surveys, with $z \sim 1$, are a very good option to study the differences between the CMB and matter distribution dipoles. Here the structure contribution is lowered as the mean redshift increases so it can be neglected as well as the foregrounds and noise, for the same reason as for the CMB.

Many studies have estimated the radio dipole with different surveys and the results are unclear. With the NRAO VLA Sky Survey (NVSS), [Blake & Wall \(2002\)](#) finds the dipole compatible with that of the CMB, but [Singal \(2011\)](#) found it with an amplitude ~ 4 times larger while in the same direction. [Rubart & Schwarz \(2013\)](#) had the same result as the latter, both with NVSS and with the Westerbork Northern Sky Survey (WENSS). The result of [Gibelyou & Huterer \(2012\)](#) gave a different direction and an amplitude ~ 6 times larger. A study presented in [Bengaly et al. \(2018\)](#) obtained again the same direction, however the amplitude was ~ 2 times larger with NVSS and ~ 5 times larger with TIFR GMRT Sky Survey (TGSS). Finally, [Colin et al. \(2017\)](#) merged the NVSS and the Sydney University Molonglo Sky Survey to construct NVSUMSS and found an amplitude ~ 4 times bigger and a direction a bit off the expected one. It must be noted that different estimators have been used in each work and that they could have a relevant impact on the results. These results infer that the origin of the dipole might not be purely kinetic, but the fact that other contributions seem to be in the same direction makes the problem even more confusing. A chance to solve this issue might be given by a new cutting-edge radio telescope such as SKA (Section 1.3) planned to be operational by the end of the 2020s. Among many other goals, it is designed to obtain a much more extensive survey than others before it, which would be perfect for the study of the cosmic radio dipole.

1.2. Kinematic dipole from the large-scale structure

The effect of the kinematic dipole to the observed dipole was given by [Ellis & Baldwin \(1984\)](#). The radio source population is assumed to have a "straight" spectral energy distribution following the power-law:

$$S \propto \nu^{-\alpha}, \quad (1.3)$$

with S the flux and ν the frequency. The number of galaxies per solid angle above a certain flux threshold is

$$\frac{dN}{d\Omega}(>S) \propto S^{-x}. \quad (1.4)$$

Where x and α are astrophysical parameters related to the nature of the galaxies and are usually fixed to $\alpha = 0.75$ and $x = 1$ (e.g. [Ellis & Baldwin \(1984\)](#)). An observer at speed \mathbf{v} with respect to the rest frame of the galaxy distribution will see the radiation coming from the source with a Doppler-shift in the frequency like

$$\nu_{\text{obs}} = \nu_{\text{rest}} \delta, \quad (1.5)$$

where

$$\delta = \frac{[1 + (v/c) \cos \theta]}{(1 - v^2/c^2)^{1/2}}. \quad (1.6)$$

Here $v = |\mathbf{v}|$ and $\cos \theta = (\mathbf{n} \cdot \mathbf{v})/v$, i.e. θ is the angle between the velocity of the observer and the direction of the radiating source, \mathbf{n} . From Eq. (1.3), for a fixed observer:

$$S_{\text{obs}} \propto \delta \nu_{\text{rest}}^{-\alpha} \propto \delta^{1+\alpha} \nu_{\text{obs}}^{-\alpha} \propto S_{\text{rest}} \delta^{1+\alpha}, \quad (1.7)$$

and the relation of the solid angle in the direction of θ is

$$d\Omega_{\text{obs}} = d\Omega_{\text{rest}} \delta^{-2}. \quad (1.8)$$

The number of galaxies per solid angle is therefore

$$\left(\frac{dN}{d\Omega}(>S, \mathbf{n}) \right)_{\text{obs}} = \left(\frac{dN}{d\Omega}(>S, \mathbf{n}) \right)_{\text{rest}} \delta^{2+x(1+\alpha)}. \quad (1.9)$$

After a first order approximation of δ and assuming $v/c \ll 1$, which is the case with the Solar System:

$$\delta \approx [1 + (v/c) \cos \theta], \quad (1.10)$$

and Eq. (1.9), in turn, since $(1+x)^y \approx 1+xy$ for $x \ll 1$ becomes

$$\left(\frac{dN}{d\Omega}(>S, \mathbf{n}) \right)_{\text{obs}} = \left(\frac{dN}{d\Omega}(>S, \mathbf{n}) \right)_{\text{rest}} \left[1 + [2 + x(1+\alpha)] \left(\frac{v}{c} \right) \cos \theta \right]. \quad (1.11)$$

The amplitude A of the dipole is the second addend between brackets, this way, the maximum change in the observed number of galaxies comes when $\cos \theta = 1$. The dipole is defined by an amplitude and a direction. The velocity obtained from the CMB dipole, with the proposed values of x and α , gives a kinematic amplitude of $A = 0.00462$. The measurements of the radio dipole are expected to give an amplitude compatible with the one of CMB. In Figure 1.1 an exaggeration of Doppler boosting is shown, with an amplitude of 0.02. In the top-right part an increase in the number of galaxies is seen while in the bottom-left that number is decreased. The same happens with the dipole amplitude of the CMB but it is not as clearly seen.

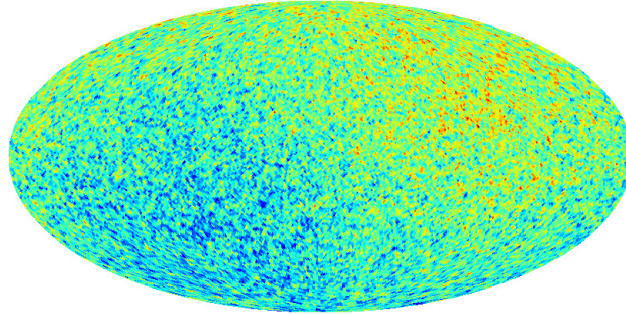


Figure 1.1: Exaggerated map showing the Doppler boosting effect on source counts in mollweide. The redshift and blueshift are clearly seen in opposite direction to each other.

1.3. Square Kilometer Array

The Square Kilometer Array¹ (SKA) is a project compromising many countries with the objective to build a state-of-the-art radio telescope, which would be the world's largest. It is planned to have a collecting area of 10^6 metres, hence its name.

¹<https://www.skatelescope.org/>

Fourteen countries are members of the SKA Organisation, with many more involved as partners. In June 2018 Spain became the eleventh country in the organisation and many researchers and technology companies are involved in the project. The idea of SKA first came to be in 1993, but it moved slowly until 2011, when the non-profit SKA Organisation was created and in 2019 the SKA Observatory Convention was signed, confirming the start of the project. Construction of the telescope is due to start in 2021 and the first results are expected as soon as 2028.

There are two phases of the project: SKA1 and SKA2. SKA1 will be a fully operational telescope which will be able to produce results that are not possible with other telescopes, even so, SKA2 will be an expansion on the previous SKA1 immensely increasing its capabilities and finally achieving the full collecting area.

The telescope will have two locations, South Africa and Australia. Details about SKA2 are not totally confirmed yet, so this work will focus only on SKA1. Each location corresponds to a different frequency range, with SKA1-MID in the Northern Cape province, South Africa, and SKA1-LOW in Western Australia. It has to be pointed out that SKA is not a project made from scratch, there are many precursors and pathfinders recognised by SKA. Pathfinders are facilities located in planned SKA areas and that will be part of the final array. Two of them are in South Africa, MeerKAT and HERA (Hydrogen Epoch of Reionization), and another two in Australia, ASKAP (Australian SKA Pathfinder) and MWA (Murchison Widefield Array). Precursors are facilities that contribute in technology, science or operation activities. Out of seventeen pathfinders LOFAR in the Netherlands, the Arecibo Observatory in Puerto Rico and EVLA in the United States of America stand out.

The scientific work that SKA will enable is extensive and variate. The telescope will observe emission from different sources: on one hand is the synchrotron emission from electrons moving in the magnetic field of galaxies and in the other the 21 cm line emission due to the spin-flip transition between hyperfine states of neutral hydrogen (HI) (SKA Cosmology Science Working Group et al., 2020). Detecting the HI line not only will allow spectroscopic galaxy surveys but also to use a recently conceived technique, the intensity mapping of neutral Hydrogen (HI IM). It measures the intensity of the 21 cm line thus obtaining the density field at large enough scales. With this technique multiple galaxies will be gathered in the same pixel so the tracking of individual galaxies will not be possible. Furthermore, as the 21 cm line is well known, the correspondence between frequency and redshift will allow a high redshift resolution. The redshift information of sources emitting synchrotron radiation will not be as precise as the spectroscopic redshifts of the 21cm line, but they will still be possible to obtain with photometric or statistical data. Three main surveys (Section 1.3.3) will enable studies on various fields, for example: the expansion of the Universe by measuring the equation of state of dark energy; the nature of gravity and the precision of General Relativity in strong-fields such as the ones of pulsars and black holes; the origin and evolution of cosmic magnetism and its influence on galaxies, stars and planets; the evolution of the Universe in the Dark Ages and the epoch of reionisation; and the search of organic molecules and weak radio signals that could lead to the discovery of extra-terrestrial life. Apart from these and other goals there will probably be many other uses for SKA.

1.3.1. SKA1-MID

SKA1-MID will have a frequency range from 350 MHz to 15.3 GHz (the goal for SKA2 is set at 24 GHz). The frequency will be split in 5 bands but SKA1 only has funds for bands 1,2 and 5 in the current configuration; Band 1 goes from 0.35 to 1.05 GHz, which is equivalent to $0.35 < z < 3$ for HI IM, Band 2 from 950 MHz to 1.75 GHz and $0 < z < 0.5$, and Band 5 from 4.6 to 15.8 GHz. It will consist of 197 dishes, 64 of which will come from the previously present MeerKAT telescope. The SKA dish diameter will be of 15 m while the ones of MeerKAT have 13.5 m of diameter.

The configuration of the dishes will be in three logarithmically distributed spiral arms instead of in clusters. The distance between the dishes will be very accurately calculated with radio signals and then computers can combine the arriving signals to obtain the equivalent of a single dish with the size of the distance between the most separated dishes. Telescopes using this configuration and technique compose what is called an *interferometer array*. This is easier than creating a really huge single dish telescope, but besides that it also give the option to use each dish or a few of them individually. Therefore, the resolution increases considerably having 0.3 arc seconds at 1.4 GHz. The maximum baseline, i.e. the maximum distance from the center of the spiral to the farthest dish, is 150 km. Two surveys will be made with SKA1-MID.

1.3.2. SKA1-LOW

Instead of dishes as in SKA1-MID, SKA1-LOW will consist of many small, fixed dipole antennas forming what is called an aperture array. It will have a similar spiral configuration, but the antennas will be placed in 512 stations, with 256 antennas each; again interferometry will be used to create a signal "beam", with the particularity that multiple beams can be created to obtain a wider field of view. The core of the spiral, with a diameter of 1 km, will have 224 of the stations and is responsible the high sensibility of the array. The diameter of each station in 40 m, and the area per antenna 3.2 m^2 at a frequency of 110 MHz. SKA1-LOW will cover a single frequency band, from 0.05 to 0.35 GHz. Even though HERA is being constructed in the South African site it will contribute to SKA1-LOW with 350 14m parabolas that will observe between 0.05 and 250 MHz, Furthermore, ASKAP will provide 36 antennas and MWA another 16.

1.3.3. Cosmological surveys

Three surveys are proposed with the two instruments above: *Medium-Deep Band 2 Survey*, *Deep SKA1-LOW Survey* and *Wide Band 1 Survey*. The latter is the most interesting to study the cosmic dipole, due to its big sky coverage and frequency range. Briefly, the *Medium-Deep Band 2 Survey* covers $5\,000 \text{ deg}^2$ in Band 2 of SKA1-MID. It is conceived to make a continuum weak lensing survey and an HI galaxy survey at low redshifts. The *Deep SKA1-LOW Survey* will cover 100 deg^2 in frequencies from 200 to 350 MHz, which is equivalent to a high redshift range between 3 and 6.

The SKA radio continuum survey proposed in [SKA Cosmology Science Working](#)

[Group et al. \(2020\)](#) is based in the *Wide Band 1 Survey*. It covers around 20 000 deg^2 which is less than 50% of the sky. The proposed survey will have a sky coverage of $f_{\text{sky}} \approx 0.52$ after the removal of the galactic plane ($|b| < 10^\circ$, the exact values of the field of view used in this work are presented in Section 2.3 about the mask or sky coverage). It will observe in Band 2 of SKA1-MID and will be used for a wide continuum galaxy survey and for HI IM at medium range redshifts. The mean redshift of the galaxy survey is expected to be around 1, avoiding the presence of large-scale structure and making it appropriate for a study of the cosmic dipole.

Chapter 2

Simulations

The radio continuum sky has been simulated following [SKA Cosmology Science Working Group et al. \(2020\)](#) and [Bengaly et al. \(2019\)](#), to emulate the results expected with SKA1 *Wide-Band 1 Survey*. The simulations have been made using the lognormal code FLASK (*Full-sky Lognormal Astro-fields Simulation Kit*) ([Xavier et al., 2016](#)).

Four different flux thresholds have been considered, $1 \mu\text{Jy}$, $5 \mu\text{Jy}$, $10 \mu\text{Jy}$ and $22.8 \mu\text{Jy}$. The latter is the expected one for SKA1 assuming a 10σ detection threshold, meaning the certainty that what is observed is indeed a radio source. Less strict detection criteria could give a smaller flux threshold such as $10 \mu\text{Jy}$, something that in a good scenario SKA could achieve. The two lowest flux thresholds have been considered in order to see the behaviour of the estimations, since with lower frequency catalogues or distant future ultra sensible experiments $1 \mu\text{Jy}$ or $5 \mu\text{Jy}$ could be seen as representative catalogues in the sense of the number of sources. For each of the flux thresholds 100 mock catalogues have been made. FLASK needs the redshift distribution of sources $N(z)$ and the theoretical angular power spectrum as input data, the angular selection function can also be included but if none is applied the complete sky is simulated, and a mask can be included later to account for it.

2.1. Power spectra, source distribution and galaxy bias

The angular power spectrum has been obtained using CAMB SOURCES ([Challinor & Lewis, 2011](#)), which calculates CMB, lensing, and galaxy counts angular power spectra. A representation of the spectra is shown in Figure 2.1. CAMB needs cosmological parameters to make calculations, the ones used in this work are from the best-fit flat ΛCDM model presented in [Planck Collaboration et al. \(2018b\)](#). The output has been set to be divided in twelve redshift windows, both for source counts and lensing, and in each bin its respective clustering bias $b(z)$ and the magnification bias α_{mag} values have been included. $b(z)$ is the relation between matter density, including dark matter, and the galaxy density; this reflects the importance of dark matter in the formation of large-scale structure. α_{mag} quantifies the effect of gravitational lensing on the observation of galaxies. How these have been obtained is explained below. FLASK uses the

C_ℓ s and an ".ini" file with the most important information of the calculations, which is the one usually used to run CAMB.

The code used to generate the radio catalogues is different from the solutions adopted in similar studies to this one. In both [SKA Cosmology Science Working Group et al. \(2020\)](#) and [Bengaly et al. \(2019\)](#) the distribution has been generated with SKA Simulated Skies (S^3) simulations¹ ([Wilman et al., 2008](#)). This data-base gives the option to choose the flux threshold, the galaxy types, and the sky area among other properties. A request for a small query can be made via the web page, but for a big survey such as this an email has to be sent to obtain all the data.

In our work, the $N(z)$ was taken from the Table 1 of [Jarvis et al. \(2015\)](#). Those values are also obtained from S^3 simulations but that does not mean the results are the same as in other papers. [Jarvis et al. \(2015\)](#) gives the number density per square degree for 12 redshift bins (from 0 to 6 with steps of 0.5) and for the galaxy types star-forming galaxy (SFG), starburst galaxy (SB), radio-quiet quasar (RQQ), Faranoff-Riley Class I sources (FRI) and Faranoff-Riley Class II sources (FRII). The detection flux thresholds provided in this paper are 100 nJy, 1 μ Jy, 5 μ Jy and 10 μ Jy. Since the flux threshold of SKA1 is estimated to be 22.8 μ Jy a polynomial fit of order 3 has been made with the other four fluxes to approximate that data. An example is shown in Figure 2.2. The number of galaxies calculated fitting galaxy types separately in compatible with directly fitting the total number of each flux. The values of radio source counts that have been used in this project are shown in Table 2.1, considering an observation frequency of $\nu = 1$ GHz which is just the upper limit of the Band 1 of SKA1-MID. It has to be noted that the proposed surveys are sensitive to lower frequencies; a lower frequency means a higher number of galaxies per solid angle following Eq. (1.3) and Eq. (1.4).

Regarding the clustering bias $b(z)$, the Figure 2 in [SKA Cosmology Science Working Group et al. \(2020\)](#) has been used. The same plot is shown here in Figure 2.3. It has been obtained following [Alonso et al. \(2015\)](#), but, since in the *evolution* models the bias increases continuously with redshift and becomes unphysical at high redshift ([SKA Cosmology Science Working Group et al., 2020](#)), a cut-off value has been taken above which the bias is taken as constant, following [Raccanelli et al. \(2012\)](#), offering a more conservative scenario. Then, with the values of Table 2.1, the combined galaxy bias (taking into account all the galaxy types) $b(z)$ is obtained through a weighted mean as

$$b(z) = \sum_i \frac{n_i(z) \cdot b_i(z)}{N(z)}, \quad (2.1)$$

where i goes over all five types of galaxies and, here, $N(z) = \sum_i n_i(z)$. The result of this weighted-average bias is shown in Table 2.2.

The magnification bias α_{mag} is more difficult to obtain, so it has been taken directly from the Table 3 of [SKA Cosmology Science Working Group et al. \(2020\)](#), which is for $S > 22.8 \mu$ Jy, but still it has been used the same for all fluxes. Although we have been forced to adopt this approximation, the impact on the results present in this work are negligible, since they rely on the radio catalogues, and not on the lensing potential maps. A better solution should be adopted in a possible future extension that make

¹<http://s-cubed.physics.ox.ac.uk/>

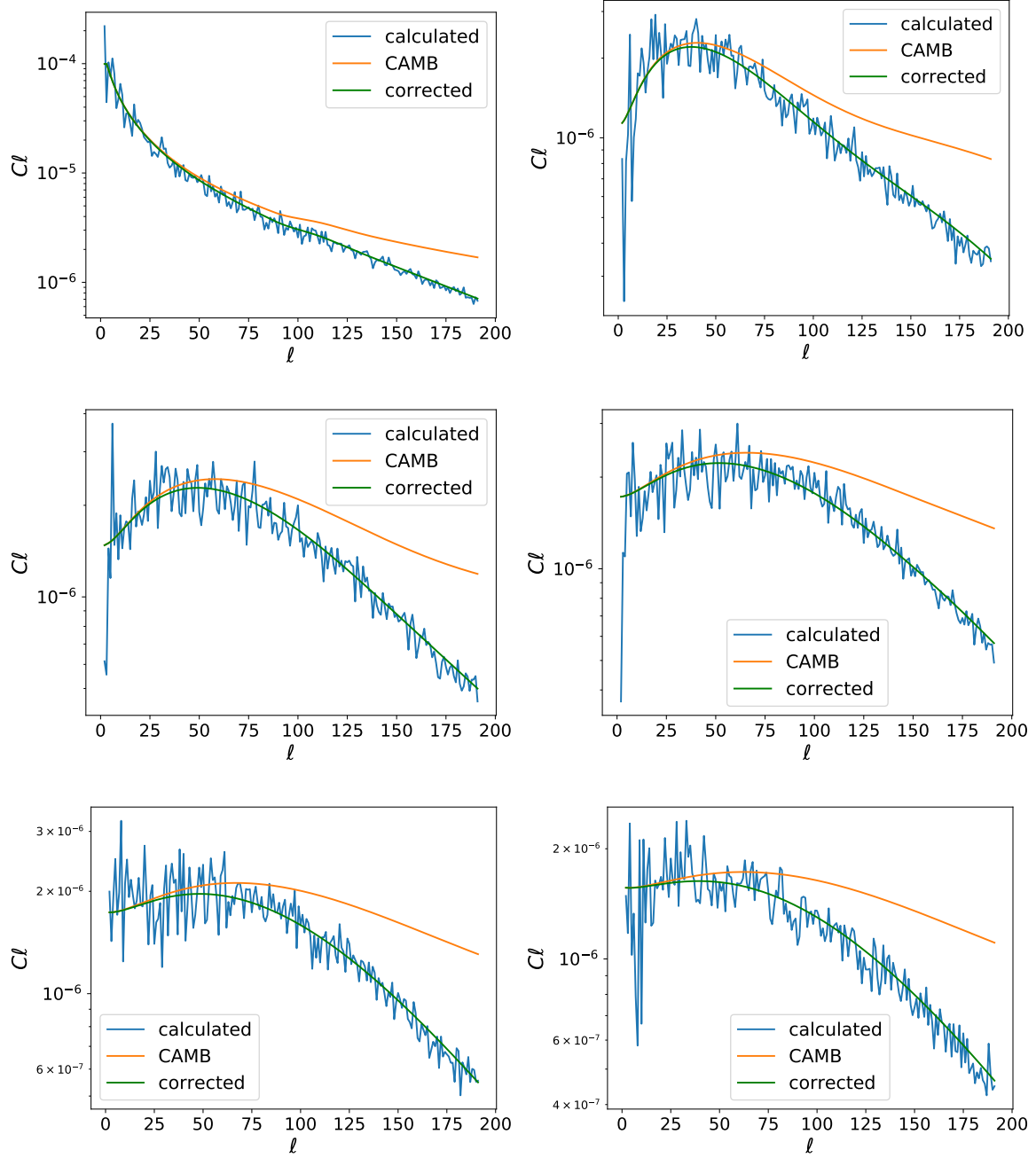


Figure 2.1: Angular power spectra for six redshift bins, starting with $0 < z < 0.5$ and skipping one bin of 0.5 every two bins, from left to right from top to bottom. Auto-spectra calculated from FLASK maps (blue) is displayed along with theoretical values from CAMB (orange) and after applying a the pixel window function correction (green). There are fluctuations in the calculated power spectra which are intrinsic in the matter density distribution of the Universe.

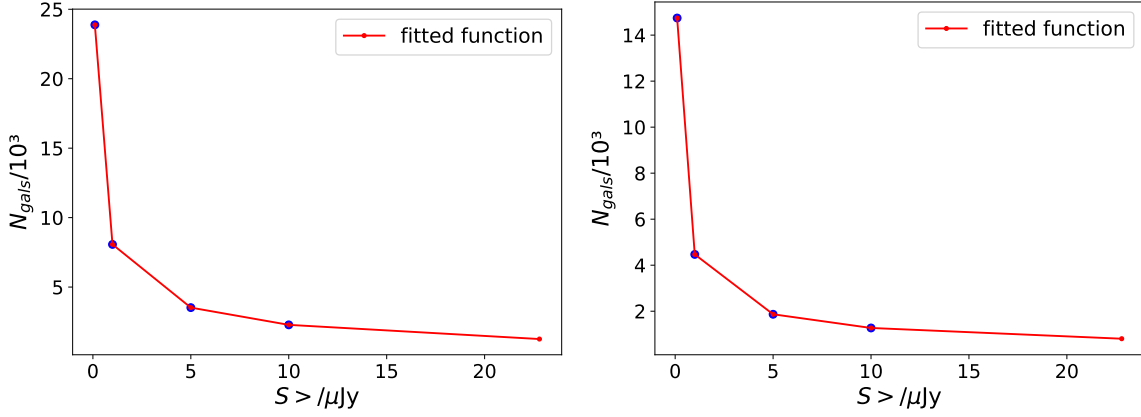


Figure 2.2: Example of polynomial fits made to determine the number of galaxies for a flux threshold of $22.8 \mu\text{Jy}$. The plot on the left is for RQQ type galaxies and the one on the right for FRI galaxies.

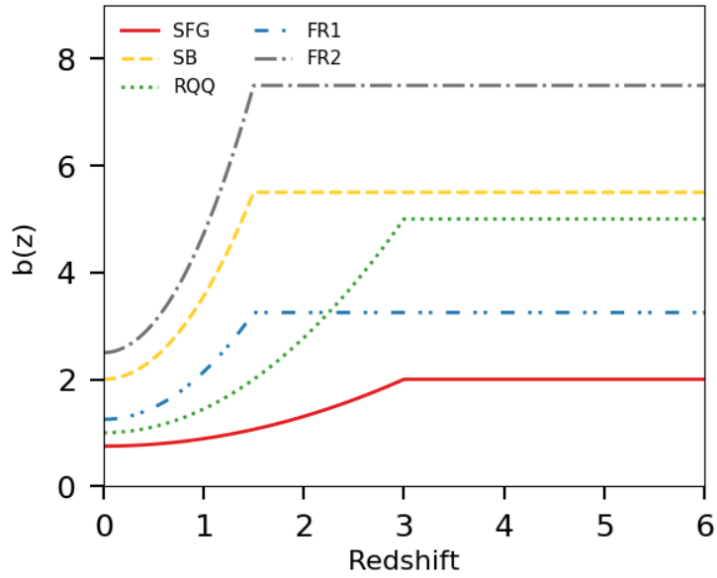


Figure 2.3: Clustering bias $b(z)$ over the redshift interval, for different galaxy types. From [SKA Cosmology Science Working Group et al. \(2020\)](#).

Redshift bins	1 μ Jy N/deg^2						5 μ Jy N/deg^2					
	SFG	SB	RQQ	FR I	FR II	Total	SFG	SB	RQQ	FR I	FR II	Total
0.0< z <0.5	3419	37	1710	279	0	5445	1761	40	701	140	0	2642
0.5< z <1.0	7759	260	1964	1004	0.2	10987	2894	267	913	425	0.2	4499
1.0< z <1.5	9181	666	1466	857	0.8	12171	2973	368	713	351	0.5	4406
1.5< z <2.0	8042	719	967	598	1.2	10327	2350	317	468	245	1.2	3381
2.0< z <2.5	5844	697	664	449	1	7655	1462	218	290	184	1	2155
2.5< z <3.0	4141	525	438	328	1.6	5434	930	162	174	136	1.2	1403
3.0< z <3.5	2970	369	281	249	0.8	3870	605	118	103	105	0.7	932
3.5< z <4.0	2213	302	204	209	0.3	2928	407	91	66	84	0.3	648
4.0< z <4.5	1676	232	145	163	0.3	2216	274	69	39	66	0.4	448
4.5< z <5.0	1259	184	102	130	0.2	1675	185	53	26	53	0.1	317
5.0< z <5.5	1005	154	75	110	0	1344	130	42	17	42	0	231
5.5< z <6.0	803	126	54	93	0	1076	91	35	10	36	0	172
Redshift bins	10 μ Jy N/deg^2						22.8 μ Jy N/deg^2					
	SFG	SB	RQQ	FR I	FR II	Total	SFG	SB	RQQ	FR I	FR II	Total
0.0< z <0.5	1117	40	445	99	0	1701	540	38	243	66	0.0	888
0.5< z <1.0	1594	208	636	293	0.2	2731	670	113	390	191	0.2	1364
1.0< z <1.5	1588	225	495	239	0.5	2548	679	99	297	151	0.7	1226
1.5< z <2.0	1206	192	309	167	1.2	1875	498	99	166	107	1.2	872
2.0< z <2.5	693	129	179	124	1	1126	256	75	87	76	1.0	495
2.5< z <3.0	405	95	97	92	1.2	690	129	54	39	57	1.5	280
3.0< z <3.5	239	67	53	71	0.7	431	62.7	33	18	43	0.8	158
3.5< z <4.0	145	50	31	56	0.3	282	31.0	23	9.3	34	0.3	98
4.0< z <4.5	87	38	17	44	0.4	186	15.1	18	4.8	27	0.3	65
4.5< z <5.0	53	28	9	35	0.1	125	7.6	12	1.5	21	0.2	42
5.0< z <5.5	34	21	5	28	0	88	4.2	8.0	0.6	18	0.0	30
5.5< z <6.0	21	17	3	24	0	65	2.1	5.9	0.4	15	0.0	24

Table 2.1: Number density per square degree N/deg^2 for different flux thresholds, redshift bins and galaxy types. SFG=star-forming galaxy, SB=starburst galaxy, RQQ=radio-quiet quasar, FRI=Fanaroff-Riley Class I sources, FRII=Fanaroff-Riley Class II sources. The data of the first three flux thresholds, 1 μ Jy, 5 μ Jy and 10 μ Jy, comes directly from [Jarvis et al. \(2015\)](#) who has used the SKA S³ simulations data-base ([Wilman et al., 2008](#)). The remaining one, of $S > 22.8$ μ Jy, has been approximated with a polynomial fit of other fluxes.

use of this extra piece of information.

Nevertheless, small differences on the $b(z)$ and α_{mag} functions are not very critical, since they appear as integrated quantities in the computations made by CAMB. $N(z)$, on the other hand, has a more important impact, since it appears as a proportional function on the computations made by FLASK. They have a direct role to the whole signal and, also, if sufficiently large (as it is the case) to make the shot-noise a negligible term, as it is the case for SKA.

Redshift bins	1 μ Jy $b(z)$	5 μ Jy $b(z)$	10 μ Jy $b(z)$	22.8 μ Jy $b(z)$
0.0< z <0.5	0.90	0.89	0.91	0.94
0.5< z <1.0	1.02	1.11	1.17	1.24
1.0< z <1.5	1.39	1.53	1.59	1.65
1.5< z <2.0	1.77	1.95	2.05	2.20
2.0< z <2.5	2.12	2.29	2.43	2.69
2.5< z <3.0	2.45	2.68	2.86	3.19
3.0< z <3.5	2.60	2.86	3.07	3.39
3.5< z <4.0	2.62	2.91	3.15	3.52
4.0< z <4.5	2.62	2.94	3.25	3.69
4.5< z <5.0	2.63	3.00	3.32	3.73
5.0< z <5.5	2.64	3.05	3.38	3.69
5.5< z <6.0	2.64	3.12	3.49	3.72

Table 2.2: Bias of all galaxies combined $b(z)$ for each flux threshold and redshift bin coming from the weighted mean of values in Table 2.1 and Figure 2.3.

2.2. FLASK

With all the input parameters and functions described above, FLASK is able to create random simulations following a multivariate log-normal distribution. The supposition that the matter distribution is Gaussian has been proven wrong by many N-body simulations as well as large-scale surveys, that show that the galaxy distribution is not Gaussian either. Notice that, although the initial fluctuations generated during the inflation period are Gaussian, and that, indeed, the CMB fluctuations are also Gaussian (as proved by *Planck*), log-normal models (Kayo et al., 2001; Coles & Jones, 1991) seem to adequate to describe the structure of the Universe due to the non linear interaction of gravity. A log-normal distribution is such that its logarithm is a normal distribution. It could also be seen as an exponential of a Gaussian distribution. This type of model is the one used by FLASK to represent matter densities and, therefore, galaxies catalogues and weak lensing maps.

FLASK is a free software which can be easily downloaded² or cloned from a github repository³. It needs a configuration file to run, with keywords defining cosmological and astrophysical parameters and input and output data. The list of keywords is explained in detail in Xavier (2015), but the most important in relation to this project are described below, with the options chosen and a brief explanation.

The outcome of the run are various files, which also controlled by a set of keywords. These output files include various HEALPix⁴ (see section 2.3) maps with the given fields (*counts* and *lensing*) and after applying the selection function the number of galaxies per pixel, all for each redshift bin. In addition, many types of angular power spectra and correlation functions can be obtained, as well as a catalogue with the position, redshift, type and other properties of every galaxy in the simulation. Next

²<http://www.astro.iag.usp.br/~flask/#download>

³<https://github.com/hsxavier/flask>

⁴<http://healpix.sf.net>

some keywords are explained, with the introduced value or word following the colon in each keyword:

- **DIST: LOGNORMAL.** Specifies the distribution that the random fields will follow.
- **FIELDS_INFO:** Asks for a string with the path leading to the fields information file. This field consists of seven columns: Field number, redshift bin number, mean, fields shift parameter, field type (1 for CMB, matter density or galaxies and 2 for weak lensing convergence), minimum redshift of the bin and finally the maximum. The field and redshift numbers can be random, but the input angular power spectra (C_ℓ s) and the output will be named accordingly. The PYTHON script "camb2info.py" is included with FLASK, which given the CAMB ".ini" file creates a fields info file ready for FLASK. It has to be noted that the field number for bins with different bias is different with this script so it has to be changed by hand.
- **CL_PREFIX:** Asks for a string with the prefix of the path leading to the C_ℓ s for each field in the fields information file. The files must be named as $[prefix]f[f_i]z[f_i]f[f_j]z[f_j].dat$ with f_i and f_j fields included in the fields information file or all spectra can be included in a single file in which case the string will not be a prefix but the complete path ending in ".dat"; the former has been used in this work. Again, there is a PYTHON script, "prepCambInput.py", given by FLASK that transforms the output from CAMB into multiple files with the required structure and path ending.
- **SELEC_SEPARABLE:** 1. Indicates whether the selection function can be divided between radial and angular parts (1 or 2) or not (0).
- **SELEC_PREFIX:** 0. If the selection function is separable, takes a string with the path to the angular part of the function, else this keyword is ignored. If this is set to 0 corresponds to a full-sky simulation. It is the case with our simulations, since the mask can be applied later and the full-sky is also interesting.
- **SELEC_Z_PREFIX:** Takes a string with the path to the radial selection function for each galaxy field (for us there is only one field). The file must have two columns, the redshift and the selection function. Here the data from Table 2.1 is included, in units of number of galaxies per unit redshift per square arcmin. The units can be changed with the next keyword, SELEC_TYPE.
- **SELEC_TYPE:** 0. Specifies the units of the selection function. If 0 it needs the number of galaxies per unit redshift per square arcmin (SELEC_SEPARABLE: 1 or 2) or the number of galaxies per square arcmin for each redshift slice (SELEC_SEPARABLE: 0). If 1 the fraction of galaxies at each angular position and redshift is asked.
- **NSIDE:** 64. Asks for an integer to specify the N_{side} parameter from HEALPix. Only takes powers of two. The number of pixels in each map with NSIDE: 64 is $12N_{side}^2 = 49152$. The maps can be easily degraded or upgraded to other N_{side} with HEALPix.

Many tests were made to make sure the results of the simulations are correct. Most of them consisted in the analysis of the output C_ℓ s. As explained in section 2.1 the bias and magnification bias have to be included in CAMB. Knowing this, a test that was made is to simulate two maps one with clustering biases than differ on a constant value, and see that the ratio of the C_ℓ s from the maps is the expected one. Also, spectra of different fields were compared.

2.3. Mask and HEALPix

SKA1-MID is located in South Africa, so not all of the sky is available from that location. Since this project wants to predict the result that would be expected with the current model using the radio survey from SKA, a mask emulating the field of view of SKA1-MID is applied. The observable sky area from this site is $f_{sky} = 0.63$, the telescope having a declination of 15° , and the mask has been constructed based in the figures from [SKA Cosmology Science Working Group et al. \(2020\)](#) and [Bengaly et al. \(2019\)](#). Some papers hint that decreasing the flux threshold or the observation frequency could lead to an increase in the field of view ([Jarvis et al., 2015](#); [Bengaly et al., 2019](#)), but that has not been considered in this work. Furthermore, the galactic plane of the Milky way has been removed due to its proximity and brightness. The mask that has been used for this is the one defined by the *Planck mission* to remove 80% of the most damaging part for the CMB. Notice that a more tailored galactic mask could be defined for this specific emissions, but the one offered by Planck provides a reasonable approach, providing an unobserved region caused by the morphology of the Milky Way as opposite to a simple cut at low latitudes like the ones taken in other studies. The final sky area with the full mask is $f_{sky} = 0.50$. It is half of the sky, adding more than 20 000 deg^2 . The sky mask is added in HEALPix, a code presented in [Górski et al. \(2005\)](#) and [Zonca et al. \(2019\)](#), in the following paragraph the most important features of this pixelization scheme will be explained. HEALPix (Hierarchical Equal Area isoLatitude Pixelization) is a multi-function instrument for the pixelization of data in a sphere. It was first created for the analysis of data from the CMB, but it has been used for other numerous purposes. It is available in C, C++ Fortran90 and more; in PYTHON it is implemented in the HEALPY package and that is the one used in this study. There are many properties make it a useful tool for the discretization of functions on a sphere at high resolution: The sphere is divided in curvilinear quadrilaterals. The lowest resolution is $N_{side} = 1$ with $N_{pix} = 12N_{side}^2$, and the resolution increases with each pixel being divided in 4 smaller pixels; note that N_{side} must be a power of two (1, 2, 4, 8...). Maps of different resolution showing the shape of pixels can be seen in [Figure 2.4](#)

The areas of the pixels are all the same. The pixels distributed in iso-latitude lines, making the integration over spherical harmonics scales much faster. HEALPix has two pixel numbering schemes, RING and NESTED. In both schemes the pixels are located in $4N_{side} - 1$ iso-latitude rings with the same azimuthal distance. All rings between the upper and lower rings have $64N_{side}$ pixels, the remaining pixels are located in both poles. In RING scheme the pixels are counted from north to south, going from ring to ring, on the other hand, in NESTED, the pixels are ordered in twelve tree-like

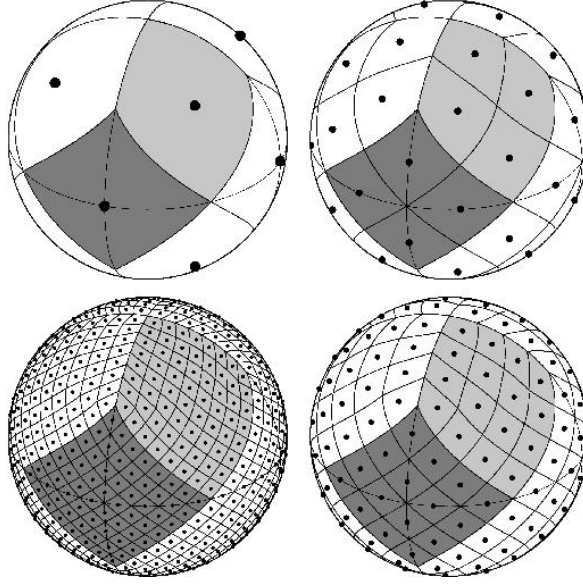


Figure 2.4: Examples of HEALPix pixel configurations with different resolutions. The pixel centres and the shapes of pixels are illustrated.

structures. The latter is more appropriate to use for neighbour search algorithms so it is the one used in this project.

In this work HEALPix has been used mainly to read and visualize galaxy distribution maps given by FLASK in ".fits" files and to create the SKA sky mask. The pixel indexation has also been used in some optimization methods to obtain the direction of the dipole. It also been useful in the first stages of the simulations to extract the angular -auto and cross- power spectra from maps. One example can be found in Figure 2.1. More functionalities include upgrading and downgrading maps, changing between Galactic, Ecliptic and Equatorial frames of reference and finding pixels within a disk. Summing up, it has been an essential part of this work.

Chapter 3

Optimization

Among the different estimators used in the literature to characterize the dipolar modulation of the large-scale structure, we have adopted the one given in Eq.(3.1) already presented in Bengaly et al. (2019). The minimization process is made by varying certain parameters, namely, the amplitude and the position coordinates (whichever they are, pixel index, Cartesian or spherical coordinates, see discussion below) to obtain the lowest possible value of the function, i.e. the minimum, for a given data:

$$\min \sum_p \frac{[N_p(\mathbf{n}, > S) - \bar{N}(> S) (1 + A \cos \theta_p)]^2}{\bar{N}(> S) (1 + A \cos \theta_p)}. \quad (3.1)$$

The term $N_p(\mathbf{n}, > S)$ is equivalent to $dN/d\Omega(> S, \mathbf{n})$ in Eq. (1.11), but we used the former to simplify the notation. The estimator goes over every pixel p and it assumes the approximation of replacing $N_{rest}(\mathbf{n}, > S)$ for $\bar{N}(> S)$, which is the average of the map for a flux threshold, thus, it ignores clustering. This is a reasonable and well-established approximation (Bengaly et al., 2019). The cosine of the angle θ_p depends on the pixel, but it can also be transformed to estimate the direction vector or directly the longitude and latitude. The use of each option depends on the optimization method and on whether there is a partial coverage or not, i.e., on the presence of the mask. Since $\cos \theta_p = (\mathbf{n}_p \cdot \mathbf{v})/v$, Eq. (3.1) can be written as

$$\min \sum_p \frac{[N_p(\mathbf{n}, > S) - \bar{N}(> S) (1 + A [x_p x + y_p y + z_p z])]^2}{\bar{N}(> S) (1 + A [x_p x + y_p y + z_p z])}, \quad (3.2)$$

where $\{x, y, z\}_p$ are the coordinates of the pixel and $\{x, y, z\}$ the ones defining the direction of the dipole. Some methods need to be provided with the gradient of the function; since it makes no sense trying to get it with Eq. (3.1) and they are more complicated for the longitude and latitude, only the gradient of Eq. (3.2) is shown

here:

$$\nabla f(A, x, y, z) = \begin{pmatrix} \bar{N} [x_p x + y_p y + z_p z] - \frac{[x_p x + y_p y + z_p z] N_p(\mathbf{n}, > S)^2}{\bar{N} (1 + A [x_p x + y_p y + z_p z])^2} \\ Ax_p \bar{N} - \frac{Ax_p N_p(\mathbf{n}, > S)^2}{\bar{N} (1 + A [x_p x + y_p y + z_p z])^2} \\ Ay_p \bar{N} - \frac{Ay_p N_p(\mathbf{n}, > S)^2}{\bar{N} (1 + A [x_p x + y_p y + z_p z])^2} \\ Az_p \bar{N} - \frac{Az_p N_p(\mathbf{n}, > S)^2}{\bar{N} (1 + A [x_p x + y_p y + z_p z])^2} \end{pmatrix}. \quad (3.3)$$

As mentioned above, different estimators have been used by other authors; a chi-square estimator was used in [Blake & Wall \(2002\)](#) to analyze the NVSS catalogue and a linear estimator in [Singal \(2011\)](#). A detailed study on this matter is made in ([Rubart & Schwarz, 2013](#)), including a characterization of the bias in the amplitude and direction, as well as corrections imposed by the mask. The estimator adopted in this work seems to provide good results.

Three minimization methods have been tested; Nelder-Mead, Newton's Conjugate Gradient (NCG) and L-BFGS-B (Limited-memory Broyden-Fletcher-Goldfarb-Shanno Box-constrains). The main difference between methods is the execution time, whereas the results are equivalent in general, at least for full-sky coverage. However, some have advantages due to their functioning need to be discussed. These methods can be found implemented for PYTHON in the *optimize.minimize* module from SCIPY.

3.1. Nelder-Mead

The Nelder-Mead simplex method ([Nelder & Mead, 1965](#)) is used to minimize functions of n variables. Given starting values, it creates a simplex, this is, a geometrical object such as a triangle or a tetrahedron, or a more complicated object for large dimension. In fact, this simplex will have $(n + 1)$ vertices. A short explanation of the method is given below, following the notation of [Gao & Han \(2012\)](#) which is also used for the SCIPY algorithm.

For a given function $f : \mathbb{R}^n \rightarrow \mathbb{R}$, for which we want to obtain the minimum like

$$\min f(\mathbf{x}), \quad (3.4)$$

and an initial simplex Δ that is created with vertices $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{n+1}$; the dimension of \mathbf{x} is n . Then the function is evaluated for Δ to determine the best vertices as $f(\mathbf{x}_1) \leq f(\mathbf{x}_2) \leq \dots \leq f(\mathbf{x}_{n+1})$. This method uses four operations: *reflection* (α), *expansion* (β), *contraction* (γ) and *shrink* (δ) (shrink was not included in the original method in [Nelder & Mead \(1965\)](#)) and each has its scalar coefficient. Usually these coefficients are taken to be ([Gao & Han, 2012](#)):

$$\{\alpha, \beta, \gamma, \delta\} = \{1, 2, 1/2, 1/2\}.$$

Also, a centroid is defined as

$$\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i, \quad (3.5)$$

using the n best vertices, i.e., all but \mathbf{x}_{n+1} . $\bar{\mathbf{x}}$ is used in every iteration to obtain new vertices.

The method is known to lose precision when working with high dimensions. This is not a concern for this estimator as the highest dimension is four, one for the amplitude and, at most, three directional parameters (when working with Cartesian coordinates). In fact, Nelder-Mead can optimize the function for three types of variables (and in all the cases including the amplitude): pixel number; longitude and latitude; and $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ vector. The simplices are, from first to last, a triangle, a tetrahedron and a *5-cell*.

In principle, the best option to characterize the direction from the point of view of the optimization would be the pixel number of the HEALPix maps, because it is only one parameter. Notice that the gradient of the estimator with the pixel number cannot be obtained; on this matter, Nelder-Mead has an advantage against the other SCIPY methods because it is the only one that does not need the gradient to work properly. This makes it a very important tool because with the masked-sky the estimation in other coordinates suffers of some numerical instabilities.

3.2. Newton's Conjugate Gradient

Methods using the gradient of the function are in general more complex than Nelder-Mead, although they tend to be faster and more precise. This is the case with NCG and L-BFGS-B. Again, the goal is to solve Eq.(3.4).

Newton methods approximate the value of the function close to a point \mathbf{x}_k with the first three terms of its Taylor expansion (Polyak, 2007):

$$f(\mathbf{x}_k + h) \approx f_k(h) = f(\mathbf{x}_k) + (\nabla f(\mathbf{x}_k), h) + \frac{1}{2}(\nabla^2 f(\mathbf{x}_k)h, h). \quad (3.6)$$

Here we need the gradient $\nabla f(\mathbf{x}_k)$ and the Hessian $\nabla^2 f(\mathbf{x}_k)$. Taking as a condition that the minimum is found where $\nabla f(\mathbf{x}_k) = 0$ the iterations are obtained as follows:

$$0 = \nabla \left[f(\mathbf{x}_k) + (\nabla f(\mathbf{x}_k), h) + \frac{1}{2}(\nabla^2 f(\mathbf{x}_k)h, h) \right] = \nabla f(\mathbf{x}_k) + (\nabla^2 f(\mathbf{x}_k), h), \quad (3.7)$$

which gives:

$$h = -\frac{\nabla f(\mathbf{x}_k)}{\nabla^2 f(\mathbf{x}_k)}, \quad (3.8)$$

where h gives the direction and size of the *jump* of each iteration. The recursive approach to obtain the minimization is therefore

$$\mathbf{x}_{k+1} = \mathbf{x}_k + h. \quad (3.9)$$

Computing h is the difficult part of the problem, and is where the conjugate gradient method (CG) comes in. The CG is used to solve systems of linear equations, with the form

$$Ax = b, \quad (3.10)$$

where x and b are vectors and A is a square matrix. For the method to work A must be symmetric and positive-definite. Eq.(3.8) can be transformed into having the form

of the equation above, with the Hessian being the matrix and h the *unknown* vector x . The PYTHON module computes the Hessian on its own with many methods available. CG is derived from the method of *Conjugate Directions*, which in turn comes from the method of *Steepest Descent*. Explained briefly, following [Shewchuk \(1994\)](#), the iterative process of this method is as follows:

$$x_{i+1} = x_i + \alpha_i d_i, \quad (3.11)$$

where α is the step and d_i the search direction:

$$\alpha_i = \frac{r_i^T r_i}{d_i^T A d_i} \quad (3.12)$$

starting with

$$d_0 = r_0 = b - Ax_0 \quad (3.13)$$

The following residuals r_i are obtained with

$$r_{i+1} = r_i - \alpha_i A d_i \quad (3.14)$$

Finally, with β_{i+1} usually calculated with the *conjugate Gram-Schmidt process*

$$\beta_{i+1} = \frac{r_{i+1}^T r_{i+1}}{r_i^T r_i} \quad (3.15)$$

the recursive formula for the search directions is also found

$$d_{i+1} = r_{i+1} + \beta_{i+1} d_i \quad (3.16)$$

making the process less expensive by not having to store all directions and residuals. With this, Eq.(3.11) is complete and everything needed for NCG to work is set. Sometimes a scaling constant γ is added to Eq.(3.9) to improve the performance.

3.3. L-BFGS-B

The L-BFGS-B algorithm is what is called a quasi-Newton method and is an upgrade of the original BFGS, where the Hessian is not calculated but accurately approximated using the gradient of the function. In fact, in most implementations it is the inverse Hessian the object that is approximated; from now on it will be denoted as $H_k = (\nabla^2 f(\mathbf{x}_k))^{-1}$. Usually the initial guess H_0 is taken to be the identity matrix, which is clearly symmetric and positive definite. It uses Eq.(3.9) with the same purpose as NCG, but obtains h in a different way. The corresponding approach for BFGS is ([Nocedal, 1980](#)):

$$H_{k+1} = H + \frac{ss^T}{y^T s} \left[\frac{y^T H y}{y^T s} + 1 \right] - \frac{1}{y^T s} [s y^T H + H y s^T] \quad (3.17)$$

where

$$s = s_k = \mathbf{x}_{k+1} - \mathbf{x}_k \quad \text{and} \quad y = y_k = \nabla f(\mathbf{x}_{k+1}) - \nabla f(\mathbf{x}_k)$$

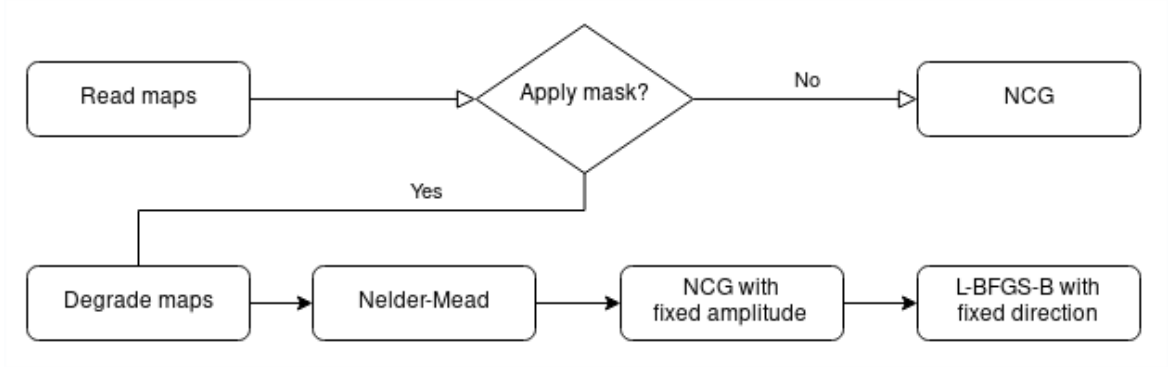


Figure 3.1: Flow-chart of the process used for the optimization of the estimator.

and the superscript T denotes the transpose. The Eq. (3.17) can also be written in the following two ways, each one of them useful to discuss some ideas:

$$H_{k+1} = H_k + U(s_k, y_k, H_k) \quad (3.18)$$

$$H_{k+1} = V_k^T H_k V_k + \rho_k s_k s_k^T \quad (3.19)$$

where $\rho_k = 1/y_k^T s_k$ and $V_k = I - \rho_k y_k s_k^T$. Now, in the original BFGS algorithm usually the new Hessian overwrites the previous one so that $n^2/2 + n/2$ storage spaces are needed for n iterations. Instead, the Limited-memory algorithm stores the corrections U_k individually (Nocedal, 1980). The algorithm has a maximum number m of correction matrices that can be stored, and after this is reached the stored values of the first corrections begin to get dropped. To take m into account a new number $j = \min\{k, m - 1\}$ is defined; the iterations are therefore (Liu & Nocedal, 1989):

$$\begin{aligned}
 H_{k+1} &= (V_k^T \cdots V_{k-j}^T) H_0 (V_{k-j} \cdots V_k) \\
 &+ \rho_{k-j} (V_k^T \cdots V_{k-j+1}^T) s_{k-j} s_{k-j}^T (V_{k-j+1} \cdots V_k) \\
 &+ \rho_{k-j+1} (V_k^T \cdots V_{k-j+2}^T) s_{k-j+1} s_{k-j+1}^T (V_{k-j+2} \cdots V_k) \\
 &\vdots \\
 &+ \rho_k s_k s_k^T
 \end{aligned} \quad (3.20)$$

To see an example, let's take $m = 6$ and, first, $k = 3$. Hence, the maximum storage capacity is not reached and $j = 3$. In the right part of the first line of the matrix (3.20) would be $(V_0 \cdots V_3)$, the results of all the previous iterations are used. If, alternatively, surpassing the maximum with $k = 8$ and $j = 5$ we see that some terms have been dropped in $(V_3 \cdots V_8)$. In fact, the same has happened in the other lines. With this algorithm many iterations can be made without having to store the terms used in the first ones.

This method is slightly faster than NCG but it is not so precise in some occasions. Even so, due to the fact that the Hessian is approximated it is less prone to numerical errors.

3.4. Our minimization approach

Each method has its pros and cons, this section will go over them and will present the procedure that we have defined to solve these issues.

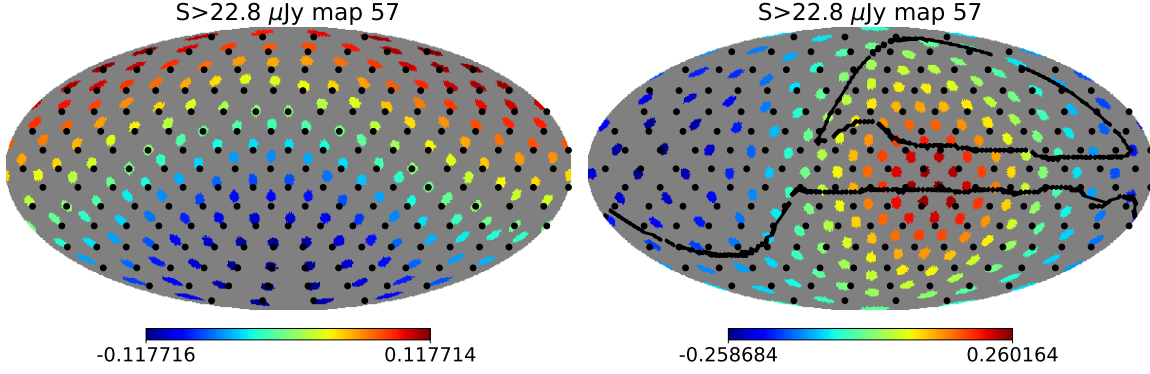


Figure 3.2: Relative errors of the measured amplitude for the same map, giving them the directions of all pixel centres of an $N_{side} = 4$ map. The left plot is for the full-sky and the right for the masked-sky, the border of the mask has been plotted in black. The coloured circles are in the direction estimated with the methods and the black points are the true directions.

First of all, a difference has to be made between maps with and without mask. Almost all methods work evenly without the mask, and independently on the selection made to characterize the direction of the monopole, i.e., the pixel index, the spherical coordinates, or the Cartesian coordinates. However, estimating the direction in Cartesian coordinates is the most precise, as their values are not delimited and are continuous (longitude and latitude go from 0 to 360 deg and from -90 to 90 respectively, the pixel number has sudden changes even in NEST ordering) and the computational time that it takes is not much more than the others. L-BFGS-B is the fastest method for this, but is not as precise as NCG, which has been the chosen one for full-sky optimization.

With a masked-sky the election is more complex, only the estimation of the direction in terms of the pixel index, at the same time as the amplitude works properly, in this case. However, as mentioned before, it is not very precise. The reasons with estimating the direction in this way, is that the direction is associated to the center of the HEALPIX pixel, which has a resolution of around 1 degree at $N_{side} = 64$. But both NCG and L-BFGS-B suffer of numerical errors with the mask so it is the only option.

Therefore, to overcome this problem we have developed an iterative approach. First, the maps are degraded to $N_{side} = 32$ and the amplitude and pixel number are estimated with Nelder-Mead. Next, and from now on with $N_{side} = 64$, the amplitude given in the previous step is fixed and the L-BFGS-B is used, so the only variables to compute are the directional ones x, y, z . L-BFGS-B is taken for this step because is slightly quicker than NCG and gives similar results. Finally, with the new directions fixed, the amplitude is estimated again with NCG, which is quicker in this case. More iterations have been tested, computing the direction and amplitude subsequently, but the results are the same as with the three steps described before. Therefore, this is the approach chosen to analyse the masked maps. A simple flow-chart showing this process can be found in Figure 3.1.

To validate the approach and to see how the estimations are affected by the choice of the final method, several simulations have been modulated in many different directions. Directions of the 192 pixel centres of a HEALPix map of $N_{side} = 4$ have been chosen as modulation directions to have a complete coverage of the sphere, and after applying a modulation equivalent in amplitude to the CMB, the subsequent dipole has been estimated. Figure 3.2 shows the increase or decrease of the estimated amplitude with respect to the given one for each direction with and without mask. The color bar shows the relative error $(1 - A_{out}/A_{in})$ of the measured amplitude with respect to the original amplitude, and the position of the circles indicate the estimated directions. The black points are the true directions given to the modulation (i.e., the centers of the HEALPIX pixels at $N_{side} = 4$). The right plot shows the results for the masked-sky, and the border of the mask is displayed.

For the masked case, a clear bias in the amplitude estimation is seen, presenting a shape induced by the geometry of the sky coverage. Notice, in addition, that regarding the estimation of the direction, the masked sky shows larger displacements from the original directions, than for the full-sky case. The directions are slightly displaced towards the part with highest possible errors. This is because the structure of the universe makes its own contribution, which is random and increases the amplitude in its direction. However, the most important aspect is that the place where biggest the errors are located is changed by the morphology of the mask. Other maps with different fluxes can be found in Figure A.4 in the Appendix, and a similar behaviour is seen, meaning the this deviation is systematic.

Considering all these maps, a systematic effect is seen in the final method, seemingly because of the reduced observable sky. The change of direction happens in both maps; since the direction of this change is random for the full sky after 100 simulations the contributions are canceled and a accurate estimation is obtained. On the other hand, as the masked-sky changes direction of the deviations, the contribution to the amplitude is added up, so an increase of the amplitude will be seen in the average of 100 maps.

To sum up, it is expected that the analysis of real data (which will present incomplete sky coverage, either by the actual location of the experiment and, in any case, also for the need of masking the foreground contamination of the Galactic plane) will suffer, not only of larger uncertainties because the smaller available sky, but also some systematic biases due to the specific morphology of the mask. This is something that should be considered in a practical situation, but in this work we just acknowledge its presence.

Chapter 4

Results and Analysis

In this chapter the simulations that have been obtained are presented, as well as the dipole parameters estimated from them. Moreover, the simulations extend beyond the specific properties of the expected characteristics of SKA (such as the flux limit or the sky coverage), to have a more complete characterization of the capability of the proposed methodology to determine the kinetic dipole.

Altogether, 4800 galaxy number maps have been attained from FLASK, one for each of the 12 redshift bins, for 100 random simulations, and for 4 different flux thresholds. These are the maps that have been used to obtain the results, but they have been merged and modulated, applying Eq. (1.11), with the values of amplitude and direction derived from previous CMB analyses (Section 1), to obtain the maps that would be expected with a radio survey. To be accurate with the real galaxy distribution, the simulations should take into account the nature of local structure and for this reason the kinematic dipole should not be adequately seen at $z \ll 1$. There could be the possibility of removing the local structure in the real survey making use of optical or infra-red catalogues or through HI redshift measurements by SKA (SKA Cosmology Science Working Group et al., 2020), so this option has also been considered. To sum up, the data sets that have been used for the analysis are the following:

- 100 maps for each flux threshold; first adding all redshift bins except the first one ($0 < z < 0.5$) and modulating them, then adding the first redshift bin without modulation. This corresponds with a scenario close to the one expected for SKA surveys.
- 100 map for each threshold removing the first bin, $0 < z < 0.5$, i.e., the local structure.
- 100 maps for each threshold adding all redshift bins. Dipole modulation has not been applied, this way, the structure dipole can be measured.
- 100 maps for each flux threshold adding and modulating all redshift bins. It is not very probable that the SKA survey will be like these simulations, the kinematic dipole is not expected to be seen in the first bin because the local

structure dominates but still has been studied to see how having a very small structure dipole would change the estimations.

The mask specified in Section 2.3 has been applied to all maps, but the dipole has been estimated both for the full-sky and for the masked-sky adding up to 3200 maps to be analyzed, 100 simulations for the four samples and the four flux thresholds. The dipole in full-sky maps has also been estimated to see the impact of the sky coverage. Ideally, it would represent the case of having an extra telescope in the norther hemisphere. Some examples are shown in Figure 4.1 for each flux. The color bar under the maps represent the number of galaxies in each pixel, as expected, the number increases as the threshold is lowered.

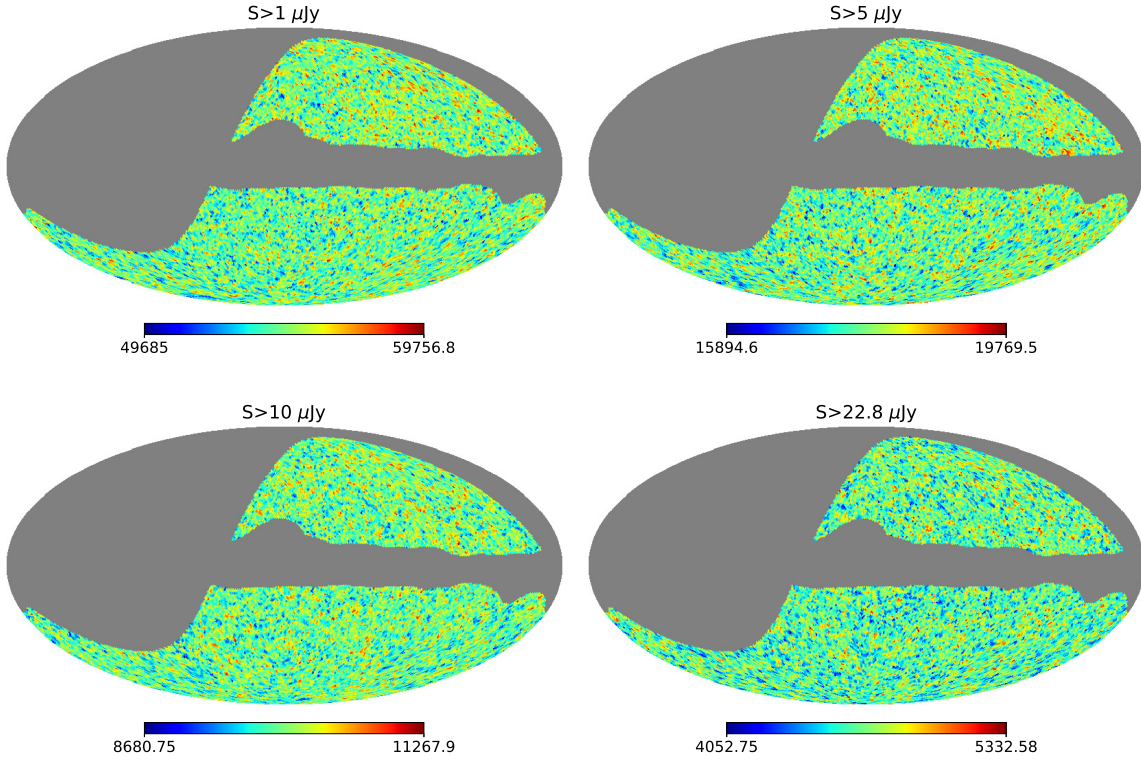


Figure 4.1: Examples of the number count distribution in mollweide projection, for each flux threshold $S >$, of the maps obtained with FLASK after adding all of the redshift bins. The pixels values represent the number of galaxies in that direction. The mask and the modulation from the CMB dipole are applied. The top-right part of the maps looks redder and the bottom-left part bluer if looked very closely.

Applying the method described in Section 3.4 the parameters of the dipole have been determined for each map. The results are presented in several tables and plots. In Table 4.1 the results for masked-sky maps are shown. There are four flux thresholds $S >$ and three date sets for each threshold as explained in the list above: with all 12 redshift bins (*full*), removing the first bin ($z > 0.5$), all bins without modulating the first one (*no Mod*) and all bins without modulation (*structure*). The values in the table are obtained by averaging over 100 simulations: the amplitude A , latitude l , longitude b , the angle between the estimated direction, and the CMB direction θ . In addition,

Masked maps						
$S > / \mu\text{Jy}$	Sample	$A/10^3$	l/deg	b/deg	θ/deg	$N_{gal}/10^7$
1	full	4.68 ± 0.51	266.41 ± 9.81	43.53 ± 6.34	9.35 ± 11.54	135.35
	$z > 0.5$	4.67 ± 0.18	267.11 ± 3.97	44.19 ± 2.20	5.37 ± 6.21	123.98
	no Mod	4.29 ± 0.50	266.62 ± 10.56	43.23 ± 6.41	10.01 ± 12.72	135.35
	structure	0.86 ± 0.41	—	—	—	135.35
5	full	4.83 ± 0.67	267.75 ± 13.49	43.54 ± 9.67	12.97 ± 15.17	44.13
	$z > 0.5$	4.76 ± 0.24	266.86 ± 4.35	43.17 ± 2.96	6.43 ± 7.56	38.62
	no Mod	4.26 ± 0.67	268.17 ± 15.23	42.84 ± 10.9	14.64 ± 17.46	44.13
	structure	1.25 ± 0.56	—	—	—	44.13
10	full	4.94 ± 0.92	267.46 ± 15.95	39.11 ± 10.74	16.51 ± 20.02	24.62
	$z > 0.5$	4.82 ± 0.23	268.47 ± 5.84	40.93 ± 3.41	9.0 ± 10.81	21.08
	no Mod	4.28 ± 0.91	267.72 ± 17.92	37.55 ± 12.46	18.96 ± 22.92	24.62
	structure	1.61 ± 0.65	—	—	—	24.62
22.8	full	4.91 ± 0.99	270.6 ± 18.16	39.73 ± 11.97	17.74 ± 21.22	11.53
	$z > 0.5$	4.87 ± 0.34	269.73 ± 5.19	38.84 ± 3.49	11.0 ± 12.44	9.67
	no Mod	4.17 ± 1.00	271.59 ± 21.18	38.2 ± 13.69	20.64 ± 25.58	11.52
	structure	1.78 ± 0.77	—	—	—	11.53
CMB		4.62	264.02	48.25	—	—

Table 4.1: Mean values of the amplitude A , galactic latitude l and longitude b , angle between the CMB dipole direction and the estimated dipole direction θ , and the total number of galaxies N_{gal} . The estimation has been done for 100 full-sky maps for each sample for four flux thresholds $S >$. The samples are all 12 redshift bins (*full*), removing the first bin ($z > 0.5$), all bins without modulating the first one (*no Mod*) and all bins without modulation (*structure*). CMB dipole parameters are also shown at the bottom of the table. The uncertainties are the standard deviations of each sample except for θ , which is the upper-tail probability.

the total number of galaxies N_{gal} is given. The standard deviation is also obtained from the simulations. Let us remark that in the case of θ , the upper-tail probability accounting for the 68% has been considered as a measurement of the uncertainty, due to the non-symmetric nature of the distribution. Table 4.2 shows the same parameters described above but having been obtained with the full-sky.

Table 4.2 and Table 4.1 give an initial all-round view of the obtained results, which can be analyzed in more detail looking at the histograms. Even if most values are compatible with the CMB in the masked-sky simulations results, there seem to be some systematic errors; the most important is present in the estimation of the latitude, showing some incompatibility for the sample without the local structure. Indeed, all samples give values that are substantially lower than the CMB latitude, and increasingly lower as the flux threshold rises. A similar error can be seen in both the amplitude and longitude, in this case the values are slightly higher than the expected ones but they are still within the uncertainty limit. Again the systematic error increases with the flux threshold. This is perceived in the big increase of θ from one table to the other. Considering that the estimations with full-sky simulations are in perfect agreement with the CMB dipole and the precision is quite high, this would imply a systematic bias caused by the morphology of the mask.

Full-sky maps						
$S > / \mu\text{Jy}$	Sample	$A/10^3$	l/deg	b/deg	θ/deg	$N_{gal}/10^7$
1	full	4.60 ± 0.42	263.59 ± 4.89	48.01 ± 3.94	3.93 ± 4.57	268.67
	$z > 0.5$	4.60 ± 0.12	264.11 ± 1.15	48.16 ± 0.87	0.83 ± 1.22	246.10
	no Mod	4.23 ± 0.42	262.8 ± 6.03	47.83 ± 4.41	4.87 ± 6.23	268.67
	structure	0.66 ± 0.28	—	—	—	268.67
5	full	4.68 ± 0.60	263.86 ± 9.54	48.08 ± 5.44	6.54 ± 7.23	87.60
	$z > 0.5$	4.64 ± 0.16	264.16 ± 2.16	48.62 ± 1.51	1.71 ± 2.18	76.66
	no Mod	4.09 ± 0.59	264.5 ± 10.05	48.23 ± 7.48	8.13 ± 10.32	87.60
	structure	0.99 ± 0.37	—	—	—	87.60
10	full	4.74 ± 0.76	263.03 ± 10.86	47.08 ± 6.24	8.22 ± 10.57	48.88
	$z > 0.5$	4.63 ± 0.18	264.03 ± 2.82	48.14 ± 1.75	2.12 ± 2.86	41.84
	no Mod	4.10 ± 0.80	262.49 ± 14.66	46.52 ± 8.57	11.41 ± 15.13	48.88
	structure	1.17 ± 0.49	—	—	—	48.88
22.8	full	4.73 ± 0.80	263.7 ± 13.18	48.0 ± 8.65	10.04 ± 11.47	22.86
	$z > 0.5$	4.65 ± 0.22	263.53 ± 3.01	48.24 ± 1.83	2.24 ± 2.97	19.19
	no Mod	3.97 ± 0.77	262.92 ± 15.39	48.52 ± 10.40	12.28 ± 14.29	22.86
	structure	1.30 ± 0.58	—	—	—	22.86
CMB		4.62	264.02	48.25	—	—

Table 4.2: Mean values of the amplitude A , galactic latitude l and longitude b , angle between the CMB dipole direction and the estimated dipole direction θ , and the total number of galaxies N_{gal} . The estimation has been done for 100 full-sky maps for each sample for four flux thresholds $S >$. The samples are all 12 redshift bins (*full*), removing the first bin ($z > 0.5$), all bins without modulating the first one (*no Mod*) and all bins without modulation (*structure*). The parameters of the CMB dipole are also shown at the bottom of the table. The uncertainties are the standard deviations of each sample except for θ , which is upper-tail probability.

Still, in both tables the overall behaviour of the results with the flux thresholds and with the samples is what was expected. Namely, the accuracy of the estimations is higher with a small flux threshold, i.e. high number of galaxies; the samples *full* and *no Mod* have a higher deviation than $z > 0.5$. The second sample ($z > 0.5$) is much more precise than the others, meaning that the intrinsic dipole of the matter distribution does affect the estimation of the kinematic dipole but, as thought, its influence comes mainly from low redshift galaxies $z < 0.5$.

From the *no Mod* sample a smaller amplitude than expected is always estimated, although it still is within the error bars. This could at first sight seem another systematic error, but it has to be taken into account that when not applying modulation to the first redshift bin, which is the one containing the local structure, the effect of the structure dipole is enhanced. Furthermore, the estimator looks at the variation in the number of galaxies in each pixel over or below the average of all the map. Having a part of the galaxies not following the dipole distribution causes an underestimation of the amplitude. Even so, the direction of the dipole does not change and is calculated also as precisely as for the *full* sample.

The amplitude of the structure dipole is shown in both tables. As expected, the uncertainty increases with the flux threshold; because the percentage of galaxies in the

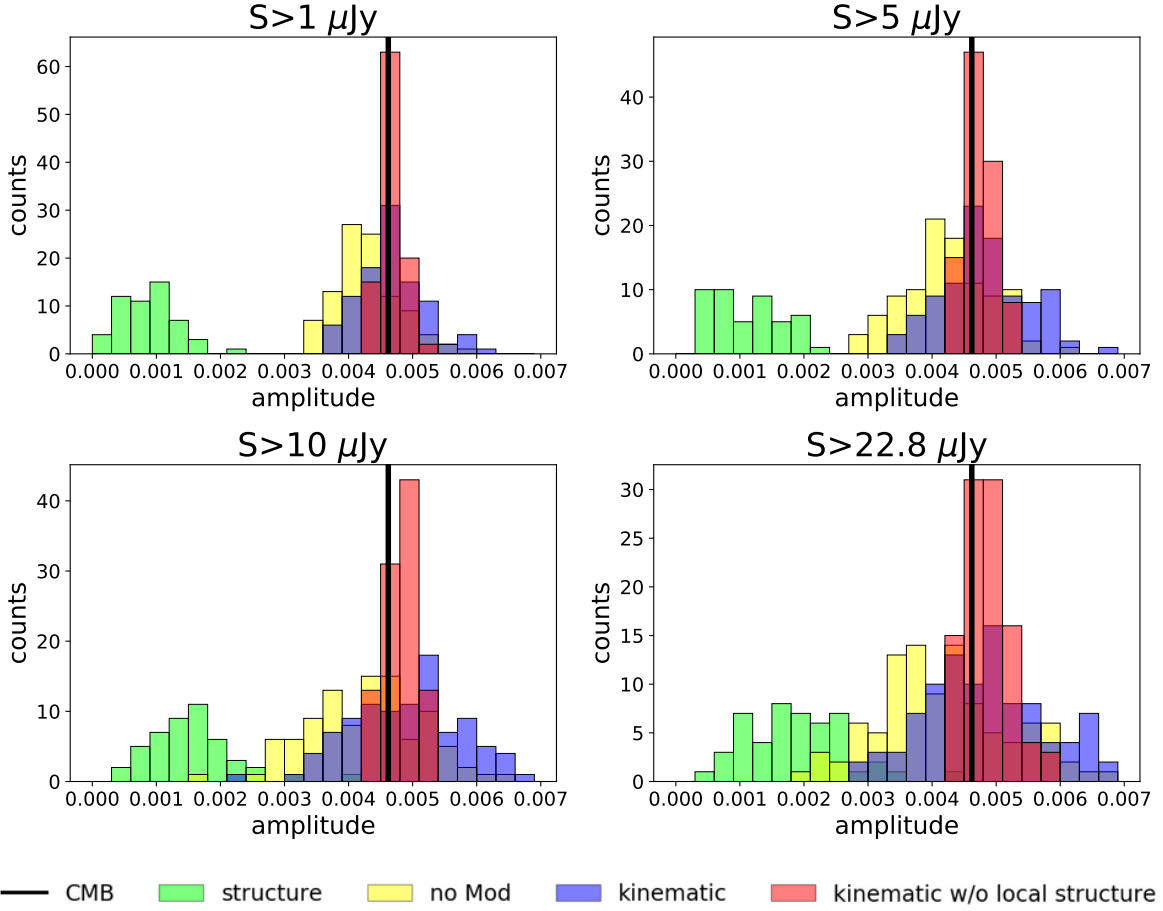


Figure 4.2: Histograms of the dipole amplitude for four flux thresholds $S >$, each plot showing 100 simulations with the masked-sky. The structure dipole (green), without kinematic dipole; the kinematic dipole (blue); kinematic dipole removing the local structure (red), without $z < 0.5$; kinematic dipole without modulating $z < 0.5$ (yellow) and the CMB dipole (black line) are presented.

first bin is higher for low fluxes (for increasing $S >$, $\sim 8\%$, $\sim 12\%$, $\sim 14\%$ and $\sim 16\%$; can be calculated from Table 2.1). Also, it is a bit larger for the masked-sky than for the full-sky, following the trend of the other samples. The mean of the direction parameters of the *structure* sample is not shown in the tables because the directions are randomly distributed, and should not be compared with the CMB dipole. Instead, these are shown in Figure 4.3 for a couple of fluxes with masked-sky and the rest are in Figures A.2, A.1 of Appendix A. The randomness of the directions is confirmed in these figures. The mean value of the latitude should be $l \approx 180^\circ$ and of the longitude $b \approx 0^\circ$, just at the middle of their respective intervals. Since there are only 100 simulations in each sample the mean is not exactly this, but very close and with a great standard deviation, the exact values are given in the histograms in Figures A.5, A.6 A.7, A.8 of Appendix A.

The estimation of the amplitude of all data sets is better seen in histograms, Figure 4.2 shows them for masked-sky simulations, for full-sky they can be found in the Appendix A. These histograms are quite simple and the results go according to the expectations. The distribution of all samples follows a symmetric distribution, meaning

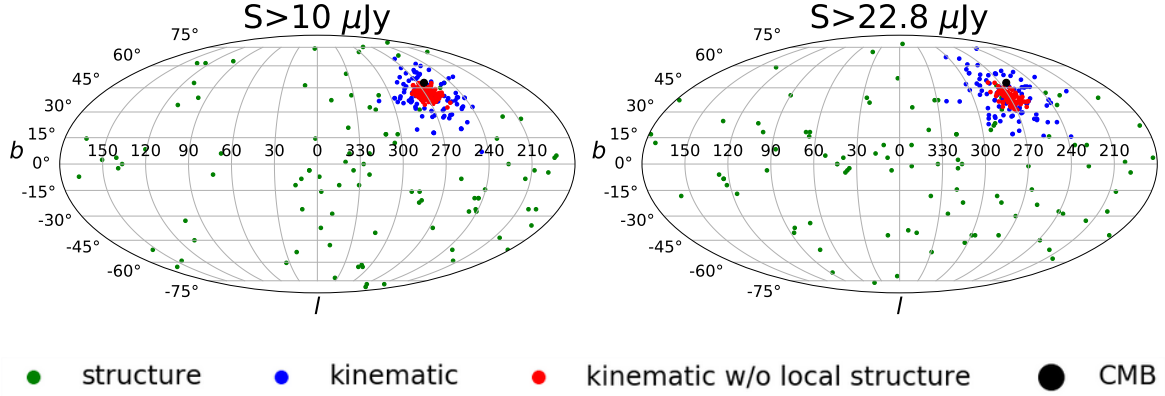


Figure 4.3: Direction of the estimated dipole for 100 simulations, for the masked-sky and flux thresholds $S > 10 \mu\text{Jy}$ and $S > 22.8 \mu\text{Jy}$. It is a mollweide projection in galactic coordinates with l the longitude and b the latitude. The different dots are the CMB dipole (black), structure dipole (green), kinematic dipole (blue) and kinematic dipole without the local structure (red), without $z < 0.5$.

the estimation is equally likely to be higher or lower than the mean. In all histograms the kinematic dipole without local structure (sample $z > 0.5$) has a higher number of counts in the central bin, meaning a smaller deviation. The kinematic (sample *full*) and the kinematic without modulating $z < 0.5$ (sample *no Mod*) have a wider distribution, the latter being displaced towards lower values for the reasons explained earlier. The histogram of the structure dipole widens and moves towards higher amplitudes as the flux threshold increases. Here a small shift with respect to the CMB amplitude (black line) can be seen, this gives the mean values from the tables. That shift does not appear in full-sky histograms so the result can be deduced to be better just by seeing the figures. Again, this is a reflection of the systematics induced by the morphology of the mask.

The direction plots indicate where the dipoles of 100 simulations of each sample point to, in galactic coordinates. The plots in Figure 4.3, show the results for simulations with flux thresholds $10 \mu\text{Jy}$ and $22.8 \mu\text{Jy}$ for masked-sky simulations while the other two fluxes and all full-sky plots can be found in the Appendix A. For the masked-sky the directions are more scattered than for the full-sky and a deviation can be seen towards lower latitudes, specially noticeable for the sample without local structure (red). The increase of the flux threshold leads to the points getting more spread, hence signalling a loss of precision in the measurements. This happens for every type of data sets. The direction of the structure dipole seems to be random; even if they are not homogeneously distributed in all plots, there is not a pattern that is repeated in every map.

As mentioned before, the results obtained with masked-sky simulations are not always compatible with the modulation applied to the maps, that is, with the CMB dipole. They show signs of systematic errors on all values, increasing the amplitude and longitude and decreasing the latitude, to the point that three values are not compatible with the CMB dipole. This error is considered systematic because the same simulations with full-sky give very accurate dipole components. In fact, all full-sky simulations

are compatible with the CMB parameters.

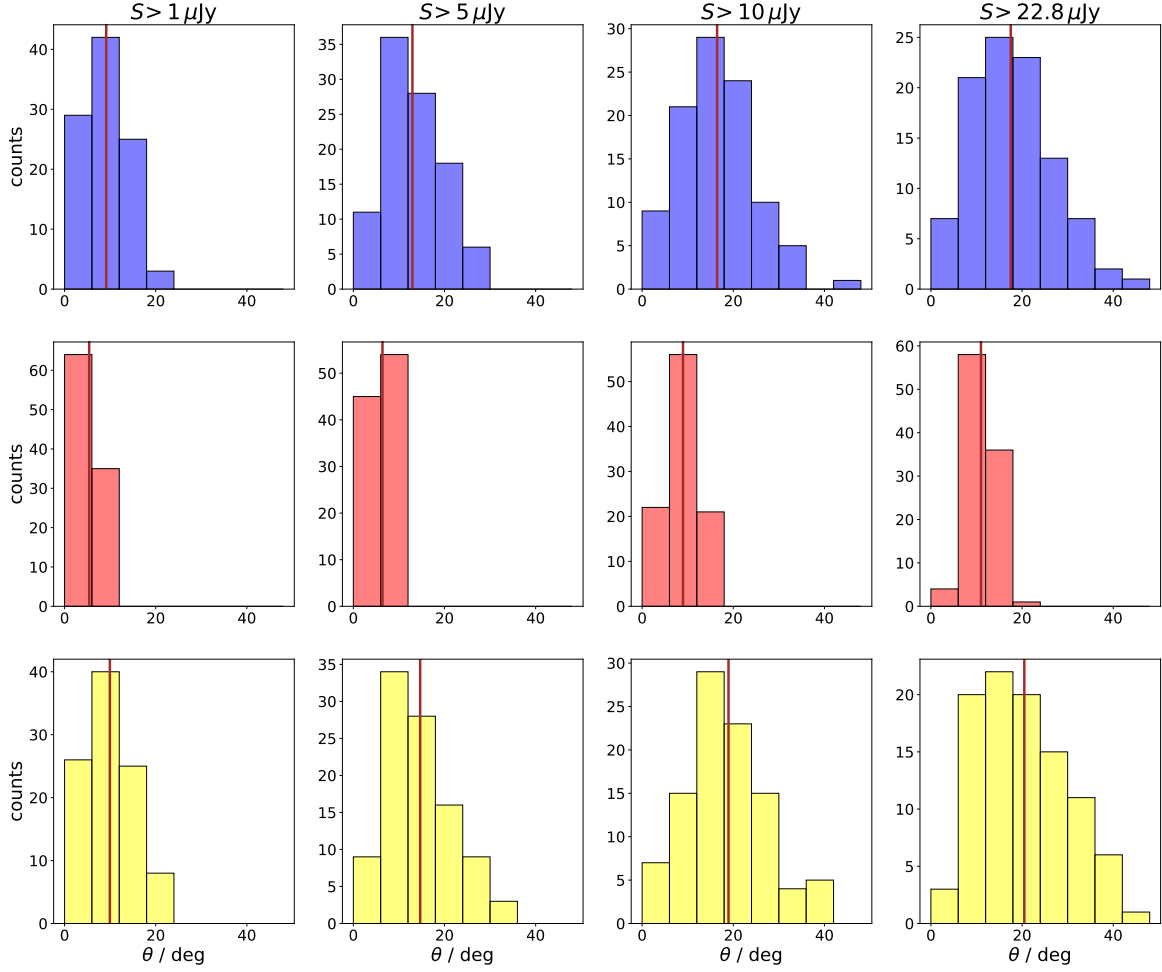


Figure 4.4: Angle between the CMB dipole and the estimated dipole θ from 100 masked-sky radio source simulations. Results for four flux thresholds are presented for three samples; *full* (blue), $z > 0.5$ (red) and *no Mod* (yellow). The mean of the distribution is shown as a brown line.

Evidence of the effect can be found in histograms of the other parameters, i.e. longitude, latitude and θ . The first two are included in Appendix A as well as θ for the complete sky; in Figure 4.4 the histograms of θ for the masked sky are shown, from which more relevant information can be extracted. First, for l and b , all histograms seem symmetric (the structure dipole is not included in this discussion) as expected; this is good because it shows that there are not preferred values but if the distribution is not centered where it should be it means that there is a shift on the estimations equal for all simulations. The displacement of the direction parameters is also seen in Figure 4.4 for the angle between the CMB dipole direction and the estimated direction. In this case the distribution should not be symmetric, but centered close to zero and with a long tail towards upper values. This is because most estimations are expected to be close to the CMB directions while a few are more dispersed, meaning the optimization tends to look for the best result. Instead, masked-sky simulations show symmetric behaviour. This distribution can be interpreted like that even if there are parameters

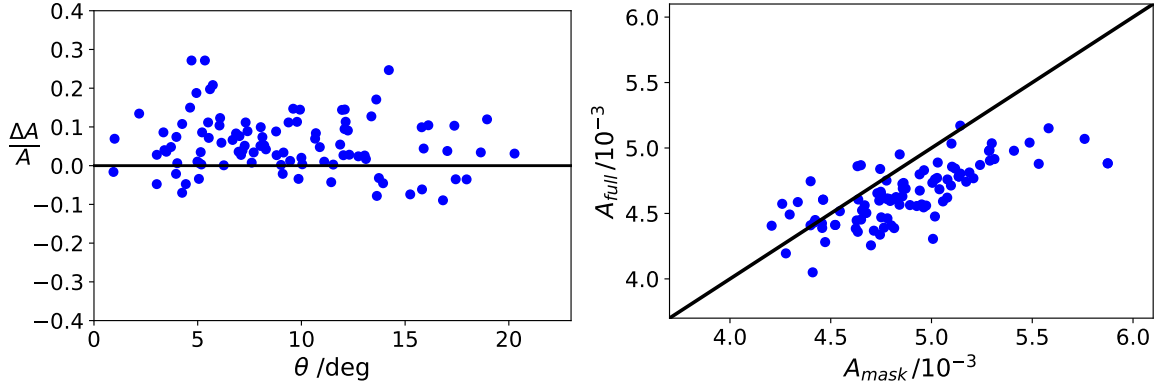


Figure 4.5: Correlation plots for the relative error of the measured amplitude $\Delta A/A$ against the angle θ between the estimated direction and the CMB direction with masked-sky (left) and the the amplitude measured for full-sky against the measured for masked-sky (right). The flux threshold for both is $22.8 \mu\text{Jy}$.

that could optimize the estimator better they are pushed towards certain preferred directions (i.e., a systematic effect caused by the morphology of the mask). The most clear is the shift towards low latitudes. On the other hand full-sky simulations follow the expected asymmetrical distribution, even if the peak is not so close to zero (it can be up to 10 deg) the tail is clearly seen. The difference between the masked-sky and full-sky is most clearly seen in the sample $z > 0.5$, but it also happens in the other samples.

In addition, Figures A.7, A.8, A.5, A.6 in the Appendix show the distribution of directions of the structure dipole. As mentioned earlier, the directions are random and so the distribution should have a similar value for all longitudes and latitudes. The mean should be in the center of the defined intervals of galactic coordinates, around 180 degrees for the longitude and 0 degrees for the latitude. The position of the mean shown in the labels of the figures follow the expectations considering the high deviation of the sample. Even so, less values seem to be found close to the limits of the intervals, the reason could simply be the actual closeness to the limits.

Correlation plots have been obtained to check for relations between the parameters, which are random variables. These are shown in Figure 4.5 and in Figure A.10 in Appendix A for the four flux thresholds. The one on the left correlates the measured amplitude and the angle θ . Due to the non-linear nature of the directional coordinates some correlation could be expected, but in the plots can be seen that the detection of A and θ is not related, this is the same for full-sky maps so the mask does not affect much on the relation between these parameters. The one on the right shows the amplitude measured for full-sky against the measured for masked-sky. The values are closely correlated meaning that the effect of the galaxy distribution is kept after reducing the observed sky area, but still, it can be seen that the values of the amplitude are a bit higher with the mask, confirming the systematic errors.

That removing the local structure improves the results has been seen above, but to study how deep this effect goes, the remaining bins have been merged in two big ones, each with approximately the same number of galaxies. The dipole has been estimated

	10 μJy		22 μJy	
	0.5< z <1.5	1.5< z <6	0.5< z <1.5	1.5< z <6
	Masked-sky			
$A \cdot 10^3$	5.13 \pm 0.47	4.86 \pm 0.29	5.12 \pm 0.47	4.86 \pm 0.29
l/deg	271.55 \pm 8.97	270.25 \pm 6.17	271.53 \pm 9.01	270.28 \pm 6.19
b/deg	33.47 \pm 5.8	38.5 \pm 3.92	33.58 \pm 5.73	38.49 \pm 3.94
θ/deg	17.14 \pm 20.31	11.73 \pm 13.32	17.05 \pm 20.31	11.76 \pm 13.39
	Full-sky			
	0.5< z <1.5	1.5< z <6	0.5< z <1.5	1.5< z <6
	Full-sky			
$A \cdot 10^3$	4.66 \pm 0.39	4.62 \pm 0.24	4.62 \pm 0.39	4.6 \pm 0.29
l/deg	263.85 \pm 4.87	264.12 \pm 2.99	264.24 \pm 4.68	263.35 \pm 3.58
b/deg	48.24 \pm 2.89	48.04 \pm 2.38	48.16 \pm 2.74	48.29 \pm 2.72
θ/deg	3.86 \pm 4.44	2.68 \pm 3.19	3.52 \pm 4.21	3.18 \pm 3.9

Table 4.3: Mean values of the dipole parameters, the amplitude A , galactic latitude l and longitude b and angle between the CMB dipole direction and the estimated dipole direction θ ; for masked-sky and full-sky and flux thresholds 10 μJy 22.8 μJy . The galaxies are divided in two redshift bins with the same amount of galaxies.

for maps with these new bins, for 100 simulations. Results of the optimization are presented in Table 4.3, similar to Table 4.1. Again the results are much better with the full-sky, as expected, but the main outcome of this analysis is that the values from the bin with highest redshifts are more accurate than the one with small redshifts in a significant way. The theory suggests that the properties of the standard model are better seen in the structure at cosmological distances rather than close to the Earth, the these results confirm that the effect of the structure extends at least until $z \approx 1.5$. Also indicates that radio galaxy surveys with highest mean redshift than SKA could give more precise results.

Chapter 5

Conclusions and future work

The cosmological principle underlying in the standard model of the universe states that the universe is homogeneous and isotropic at large scales. Due to our motion in respect to the rest frame of the CMB a dipole signal is expected to be seen on the large scale structure of the universe. We have made a set of simulations following the expected radio surveys from SKA and a method has been proposed for the detection of the kinematic dipole. Finally, it has been measured and the accuracy of the method has been analyzed.

The results obtained in our work are, in general, in good agreement with the expected amplitude and direction as shown in Table 4.2 and Table 4.1. Still, the masked-sky shows signs of systematic errors as explained in the previous section and thus the values predicted for SKA (highest fluxes and masked sky) are not fully compatible. Nevertheless, this systematic effect could be accounted for.

The precision of the results is high compared with the results from the study of the dipolar modulation of CMB fluctuations or with other radio surveys, specially after removing the local structure ($z < 0.5$). Indeed, the dipole anisotropy obtained with modulation and aberration effects of CMB fluctuations (Planck Collaboration et al., 2014) give an uncertainty of 20% of the value, which is the same as the obtained in this work for the highest flux threshold and without removing the local structure, removing it reduces the uncertainty to about one third, 7%. The errors in the direction are of the same order as for the CMB, that is, between 10-25 deg, but again are decreased by removing the local structure from the catalogues. Previous studies with other radio surveys give uncertainties from 20% up to 35% for the amplitude and from 10 to 30 deg in each directional parameter; these studies have been mentioned in Section 1.1.

Without doubt, the SKA telescope will be able to provide results that can corroborate previous estimations of the dipole and, if the local structure can be appropriately removed (as it expected using photometric information), they will be much more accurate. In addition, with the precision of the estimations that can be obtained from SKA, it could be verified whether other results, for example from NVSS, that give an amplitude ~ 4 times greater than that of the CMB, are correct or not. Furthermore, some light could be shed on the origin of anomalies claimed by studies of the CMB at large scales and in the nature of the observed dipole in general; with the prospect of detecting an intrinsic dipole anisotropy, other contributions not considered until now

or simply measuring the kinematic dipole with very high precision. With this, the strength of the cosmological principle will be evaluated and therefore the standard model of cosmology is expected to be validated, or, if the results are completely unexpected, refuted. This type of studies using simulations to forecast the accuracy of optimization methods are useful to give the certainty that when real estimations are made, the results are correct.

The final method for optimization might prove not to be the best, specially because of the systematic effects. A possible solution would be to define an optimal estimator based on a likelihood. Indeed, this approach could have the advantage of, using a suitable MCMC sampler, like, for instance the package EMCEE ([Foreman-Mackey et al., 2013](#)) provided in PYTHON, provide directly the estimations and the uncertainties from the samples, instead of performing simulations to derive them. As it is known, MCMC allows the user to marginalize over parameters, or to derive the right correlations, to incorporate prior information, or, even to obtain a posteriori probabilities that can be used on Bayesian model selection. Here, the problem is to find the right likelihood function. However, taking into account that the matter density follows a log-normal distribution (see, for instance, [Arnalte-Mur et al. \(2016\)](#)), we think that this is a good choice to be explored.

In addition to changing the estimator, further work could be to check the relation between the direction of the structure dipole from local structure and the estimation of the amplitude. Clearly, as it was mentioned in the Introduction there is a contribution from the local structure. It also has been indirectly seen by how the results improve after removing redshifts below 0.5 in some simulations and has been directly calculated in maps without modulation. Still, a further analysis of how the interaction between the structure dipole and the sky mask behaves could be helpful to know the reason of the systematic errors when applying the mask. This could be done making tests such as the ones shown in [Figure 3.2](#) and combining it with the information of the structure dipole.

References

- Alonso D., Salvador A. I., Sánchez F. J., Bilicki M., García-Bellido J., Sánchez E., 2015, [MNRAS](#), **449**, 670
- Arnalte-Mur P., Vielva P., Martínez V. J., Sanz J. L., Saar E., Paredes S., 2016, [Journal of Cosmology and Astroparticle Physics](#), **2016**, 005
- Bengaly C. A. P., Maartens R., Santos M. G., 2018, [Journal of Cosmology and Astroparticle Physics](#), **2018**, 031
- Bengaly C. A. P., Siewert T. M., Schwarz D. J., Maartens R., 2019, [MNRAS](#), **486**, 1350
- Blake C., Wall J., 2002, [Nature](#), **416**, 150
- Challinor A., Lewis A., 2011, [Physical Review D](#), **84**, 043516
- Coles P., Jones B., 1991, [MNRAS](#), **248**, 1
- Colin J., Mohayaee R., Rameez M., Sarkar S., 2017, [MNRAS](#), **471**, 1045
- Ellis G. F. R., Baldwin J. E., 1984, [MNRAS](#), **206**, 377
- Fixsen D. J., 2009, [The Astrophysical Journal](#), **707**, 916
- Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, [Publications of the ASP](#), **125**, 306
- Gao F., Han L., 2012, [Computational Optimization and Applications](#), **51**, 259
- Gibelyou C., Huterer D., 2012, [MNRAS](#), **427**, 1994
- Górski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, [Astrophysical Journal](#), **622**, 759
- Jarvis M., Bacon D., Blake C., Brown M., Lindsay S., Raccanelli A., Santos M., Schwarz D. J., 2015, in *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*. p. 18 ([arXiv:1501.03825](#))
- Kayo I., Taruya A., Suto Y., 2001, [Astrophysical Journal](#), **561**, 22
- Liu D. C., Nocedal J., 1989, *MATHEMATICAL PROGRAMMING*, **45**, 503
- Nelder J., Mead R., 1965, *Comput. J.*, **7**, 308

- Nocedal J., 1980, [Mathematics of Computation](#), 35, 951
- Planck Collaboration et al., 2014, [A&A](#), 571, A27
- Planck Collaboration et al., 2018a, arXiv e-prints, p. [arXiv:1807.06205](#)
- Planck Collaboration et al., 2018b, arXiv e-prints, p. [arXiv:1807.06209](#)
- Planck Collaboration et al., 2020, [A&A](#), 641, A7
- Polyak B., 2007, [European Journal of Operational Research](#), 181, 1086
- Raccanelli A., et al., 2012, [MNRAS](#), 424, 801
- Rubart M., Schwarz D. J., 2013, [A&A](#), 555, A117
- SKA Cosmology Science Working Group et al., 2020, [Publications of the ASA](#), 37, e007
- Shewchuk J. R., 1994, Technical report, An Introduction to the Conjugate Gradient Method Without the Agonizing Pain. USA
- Singal A. K., 2011, [Astrophysical Journal Letters](#), 742, L23
- Weinberg S., 2008, Cosmology. OUP Oxford
- Wilman R. J., et al., 2008, [MNRAS](#), 388, 1335
- Xavier H. S., 2015, FLASK: Usage and installation manual. https://github.com/hsxavier/flask/raw/master/doc/flask_manual.pdf
- Xavier H. S., Abdalla F. B., Joachimi B., 2016, [MNRAS](#), 459, 3693
- Zonca A., Singer L., Lenz D., Reinecke M., Rosset C., Hivon E., Gorski K., 2019, [Journal of Open Source Software](#), 4, 1298

Appendix A

Additional figures

With the purpose of making easier the reading of the thesis several figures have been included in this appendix. These figures are similar to the ones shown in the main text and do not provide much more relevant information, therefore, they have been moved here not to overfill the other part with figures. Below are the direction plots in mollweide projection for full-sky simulations and then the remaining two for the masked sky. After that the amplitude histograms like Figure 4.2 are presented.

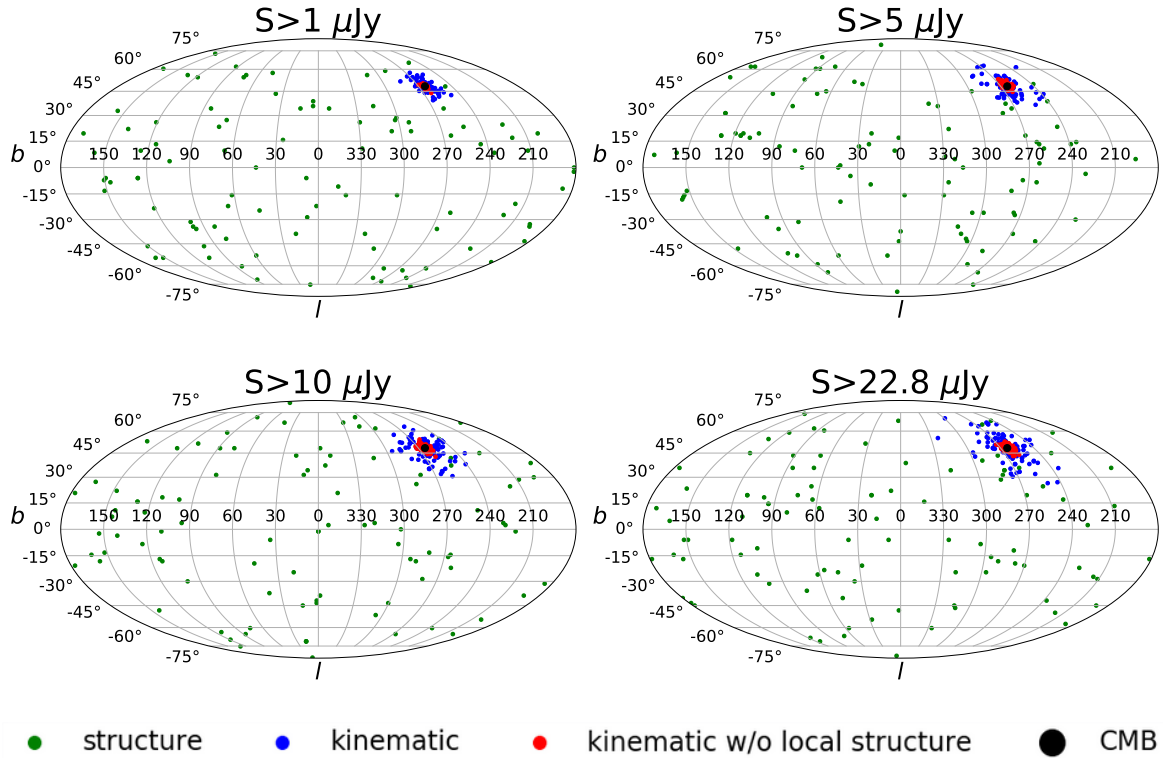


Figure A.1: Direction of the estimated dipole for 100 simulations, for full-sky in each of the four flux thresholds $S >$. It is a mollweide projection in galactic coordinates with l the longitude and b the latitude. The different dots are the CMB dipole (black), structure dipole (green), kinematic dipole (blue) and kinematic dipole without the local structure (red), without $z < 0.5$.

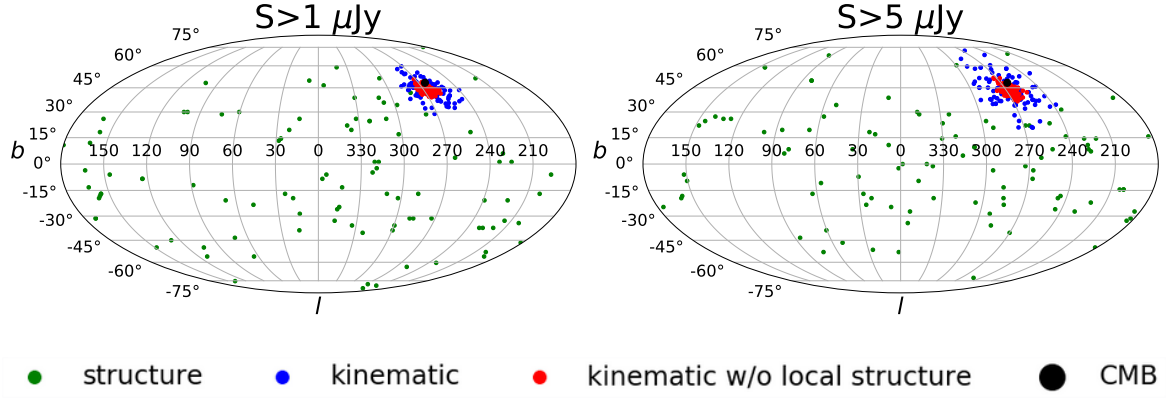


Figure A.2: Direction of the estimated dipole for 100 simulations, for the masked-sky and flux thresholds $S > 1 \mu\text{Jy}$ and $S > 5 \mu\text{Jy}$. It is a mollweide projection in galactic coordinates with l the longitude and b the latitude. The different dots are the CMB dipole (black), structure dipole (green), kinematic dipole (blue) and kinematic dipole without the local structure (red), without $z < 0.5$.

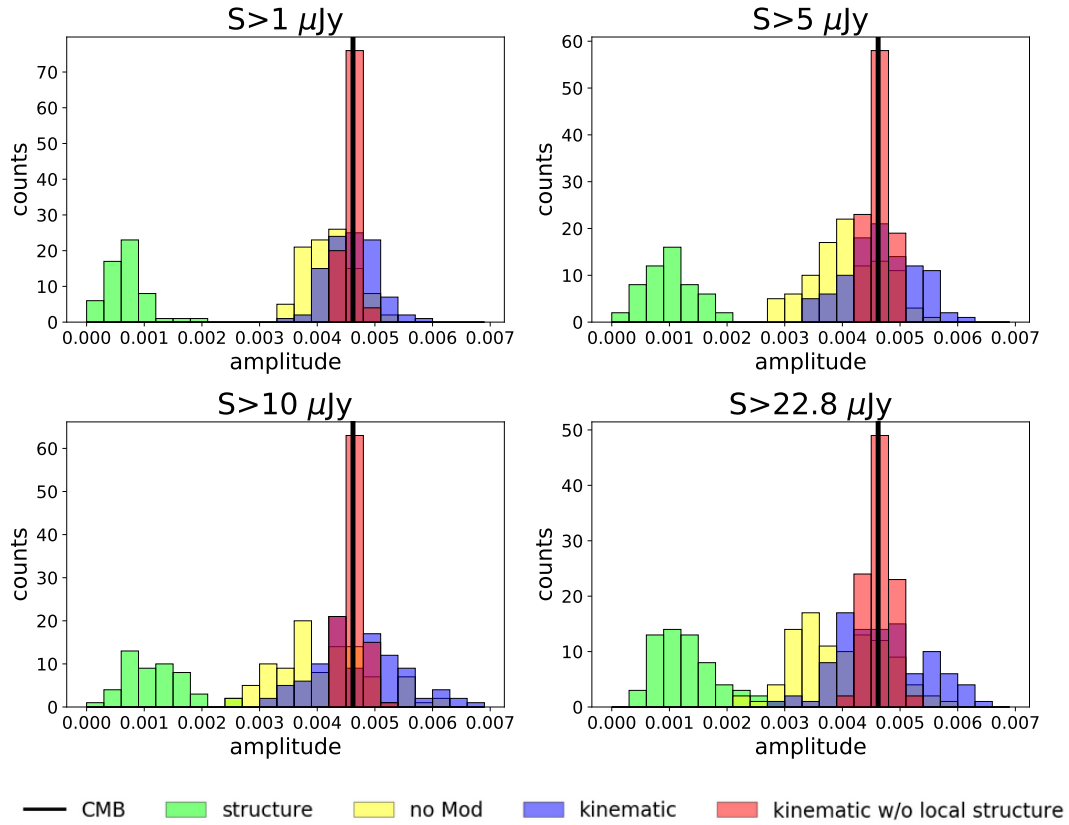


Figure A.3: Histograms of the dipole amplitude for four flux thresholds $S >$, each plot showing 100 simulations with the full-sky. The structure dipole (green), without kinematic dipole; the kinematic dipole (blue); kinematic dipole removing the local structure (red), without $z < 0.5$; kinematic dipole without modulating $z < 0.5$ (yellow) and the CMB dipole (black line) are presented.

In Figure A.4 figures from the estimations to study the systematic errors of the final optimization methods, similar to the one in Section 3.4, are shown. The change of direction from the full-sky maps to the masked-sky maps is seen in all plots, and corroborate the discussion made before.

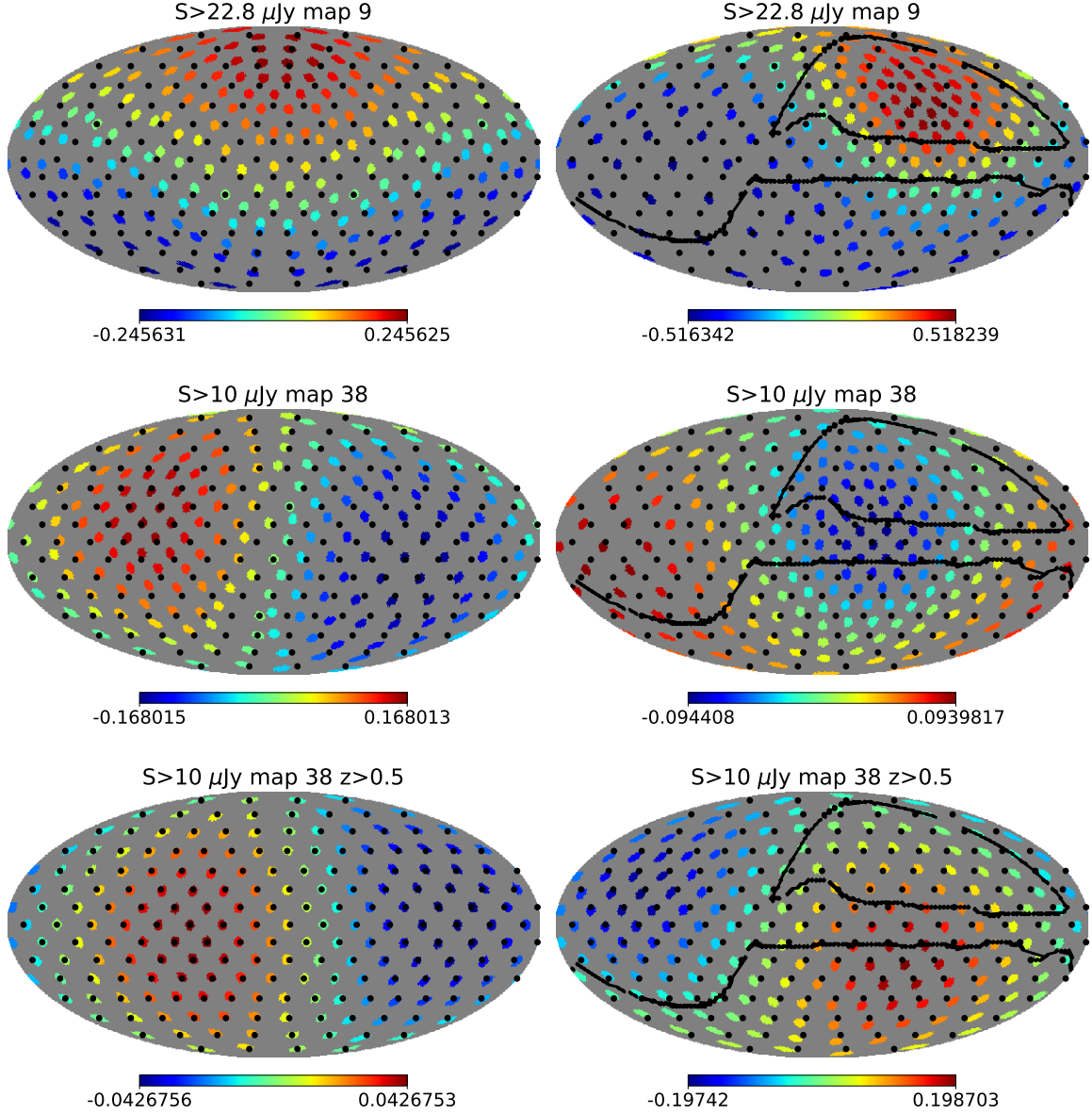


Figure A.4: Relative errors of the measured amplitude for the same map, giving them the directions of all pixel centres of an $N_{side} = 4$ map. The left plot is for the full-sky and the right for the masked-sky, the border of the mask can be seen. The coloured circles are in the direction estimated with the methods and the black points are the true directions. Here different maps are shown, two of them after removing the local structure ($z < 0.5$).

Following, all histograms of the measured longitude l (Figures A.5,A.5), the latitude b (Figures A.6,A.8) and the angle θ (Figure A.9) are shown. For the latter the ones of the masked-sky are shown in the Results and Analysis 4 chapter.

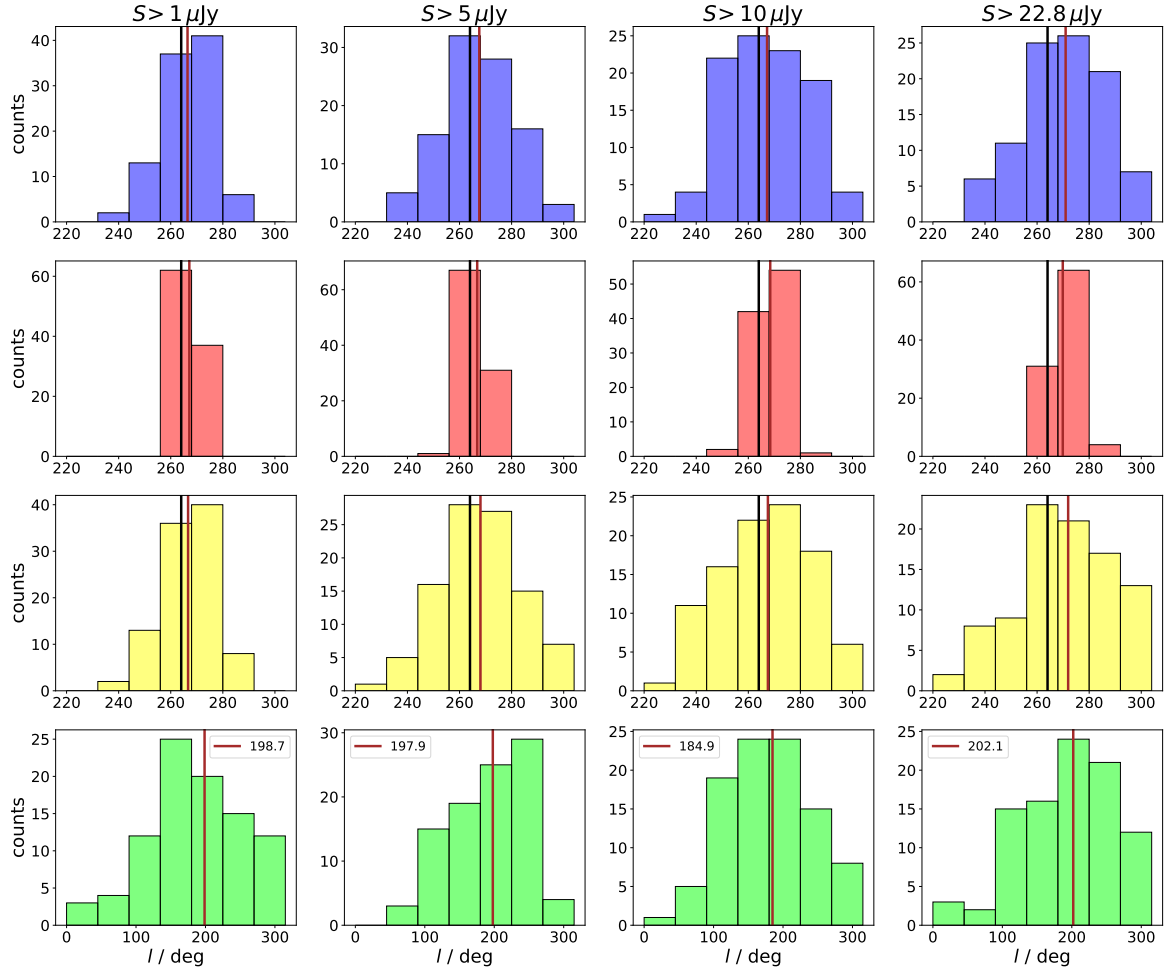


Figure A.5: Histograms showing the galactic longitude l of 100 masked-sky radio source simulations. Results for four flux thresholds are presented for the four samples; *full* (blue), $z > 0.5$ (red), *no Mod* (yellow) and *structure* (green). The value of the parameter in the CMB dipole (black line) and the mean of the distribution (brown) are also shown.

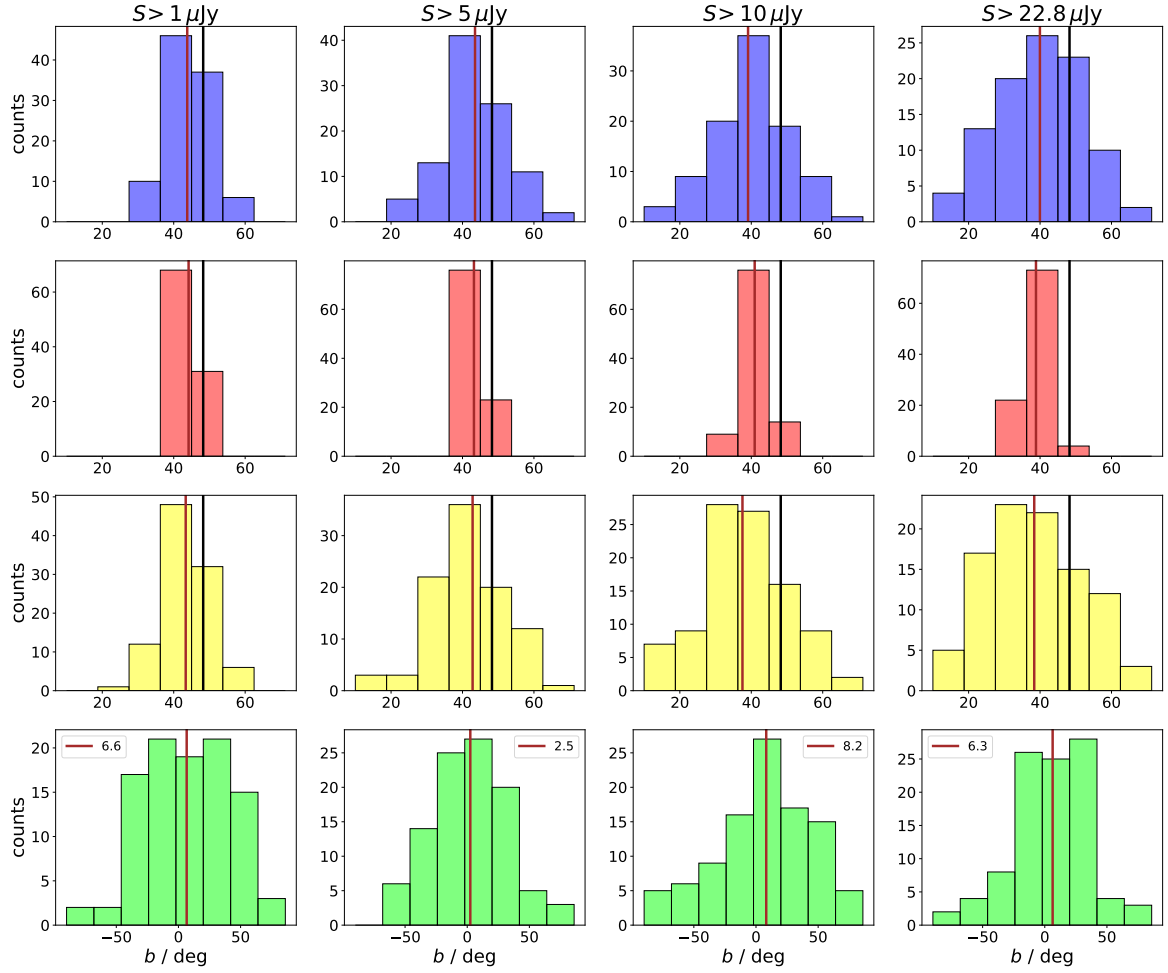


Figure A.6: Histograms showing the galactic latitude b of 100 masked-sky radio source simulations. Results for four flux thresholds are presented for the four samples; *full* (blue), $z > 0.5$ (red), *no Mod* (yellow) and *structure* (green). The value of the parameter in the CMB dipole (black line) and the mean of the distribution (brown) are also shown.

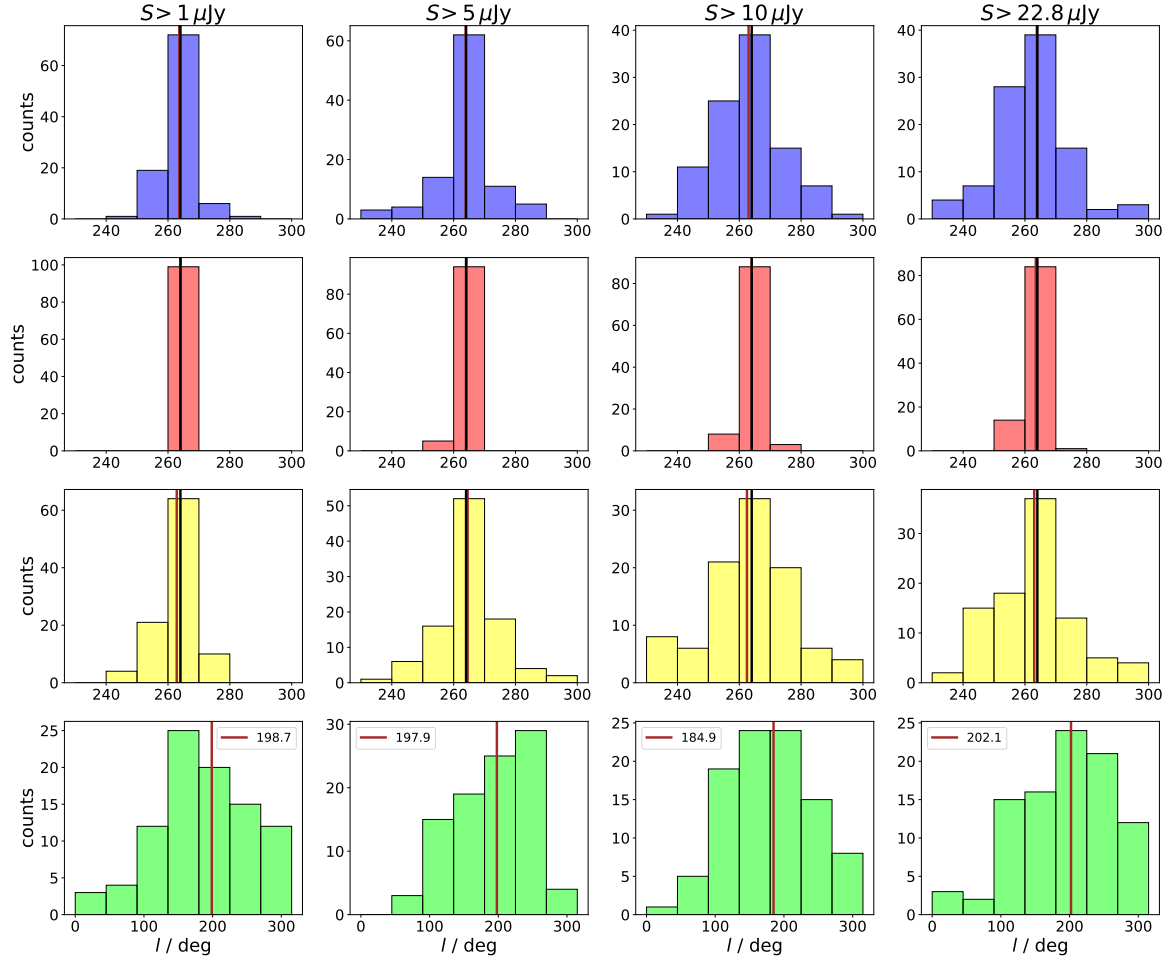


Figure A.7: Histograms showing the galactic longitude l of 100 full-sky radio source simulations. Results for four flux thresholds are presented for the four samples; *full* (blue), $z > 0.5$ (red), *no Mod* (yellow) and *structure* (green). The value of the parameter in the CMB dipole (black line) and the mean of the distribution (brown) are also shown.

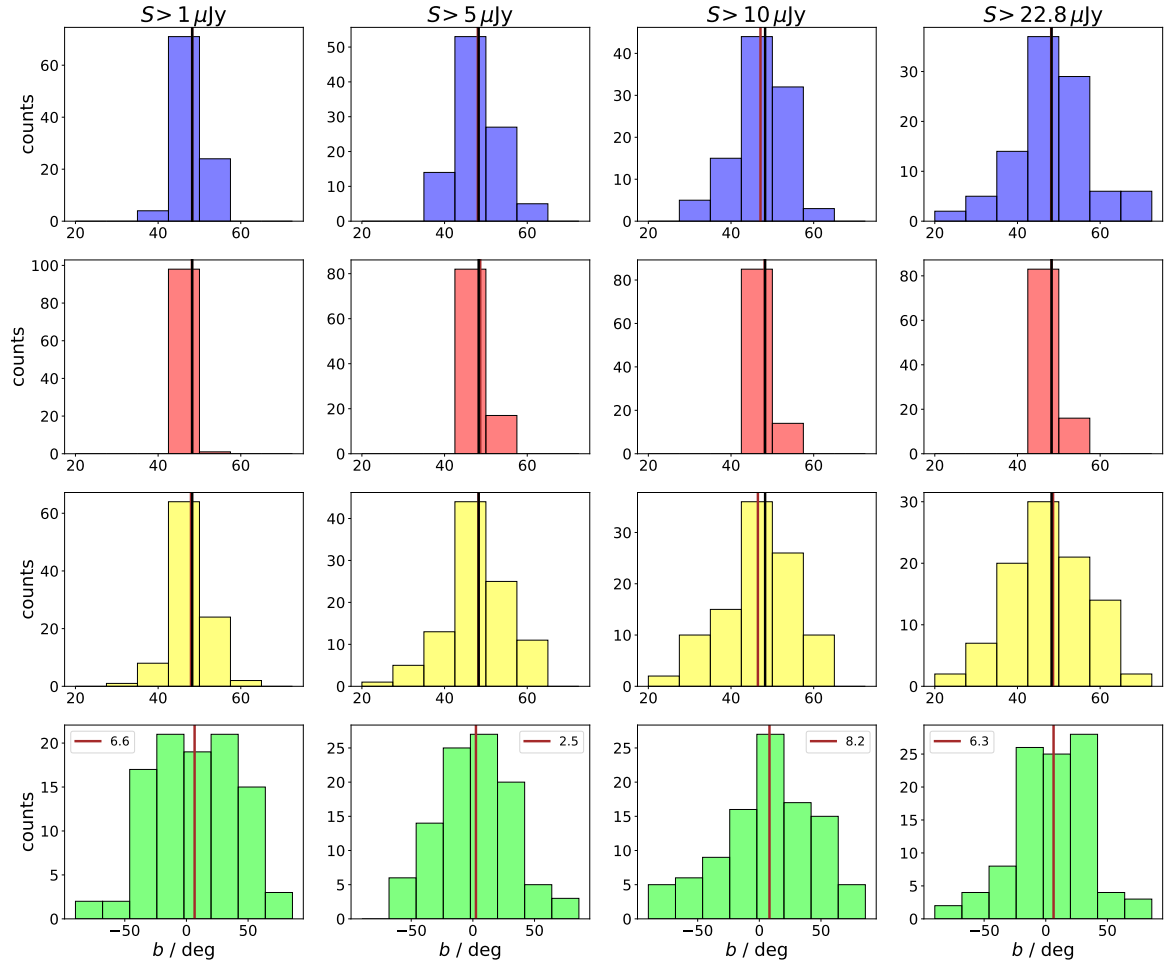


Figure A.8: Histograms showing the galactic latitude b of 100 full-sky radio source simulations. Results for four flux thresholds are presented for the four samples; *full* (blue), $z > 0.5$ (red), *no Mod* (yellow) and *structure* (green). The value of the parameter in the CMB dipole (black line) and the mean of the distribution (brown) are also shown.

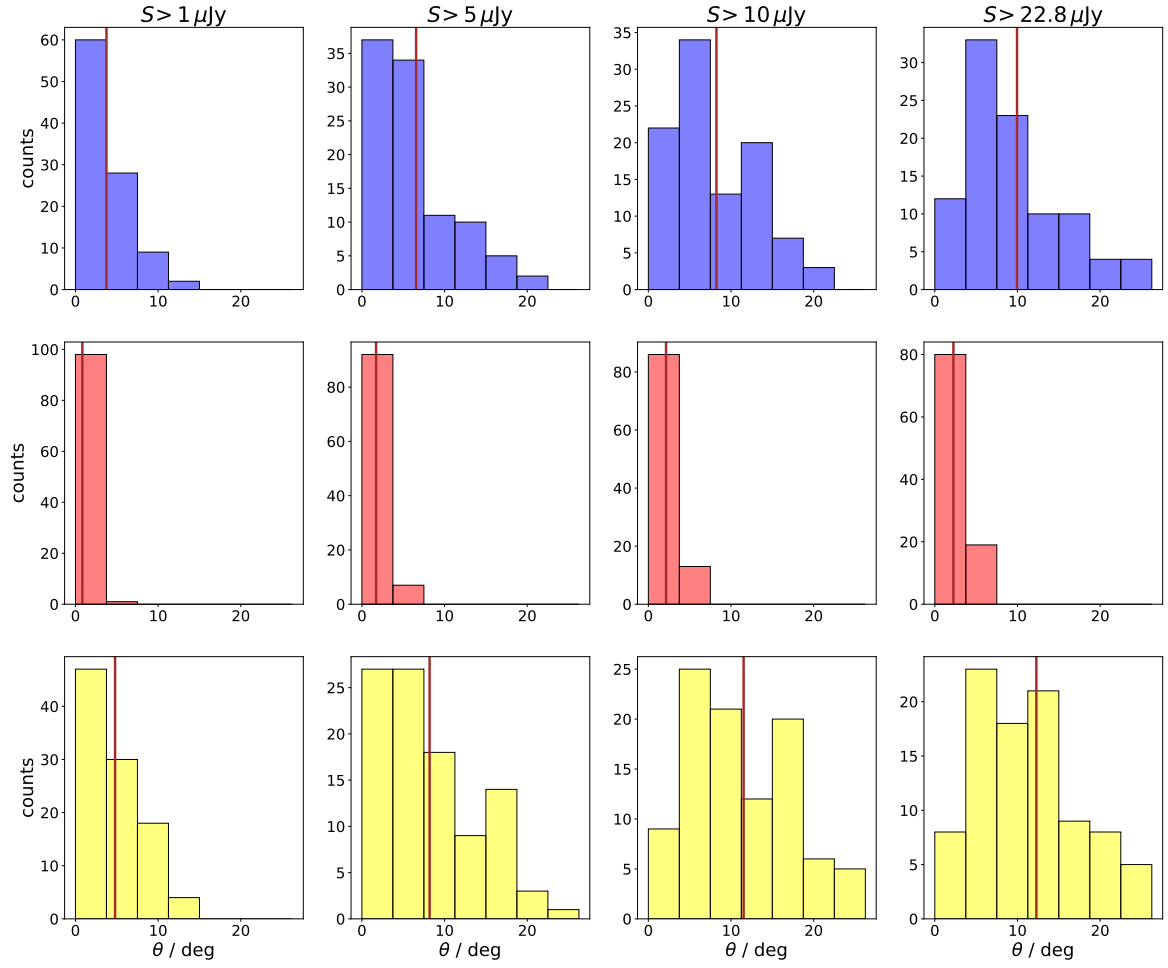


Figure A.9: Angle between the CMB dipole and the estimated dipole θ from 100 full-sky radio source simulations. Results for four flux thresholds are presented for three samples; *full* (blue), $z > 0.5$ (red) and *no Mod* (yellow). The mean of the distribution is shown as a brown line.

The remaining simulations like the ones in Figure 4.5 are shown here, those in the referenced figure are included for symmetry.

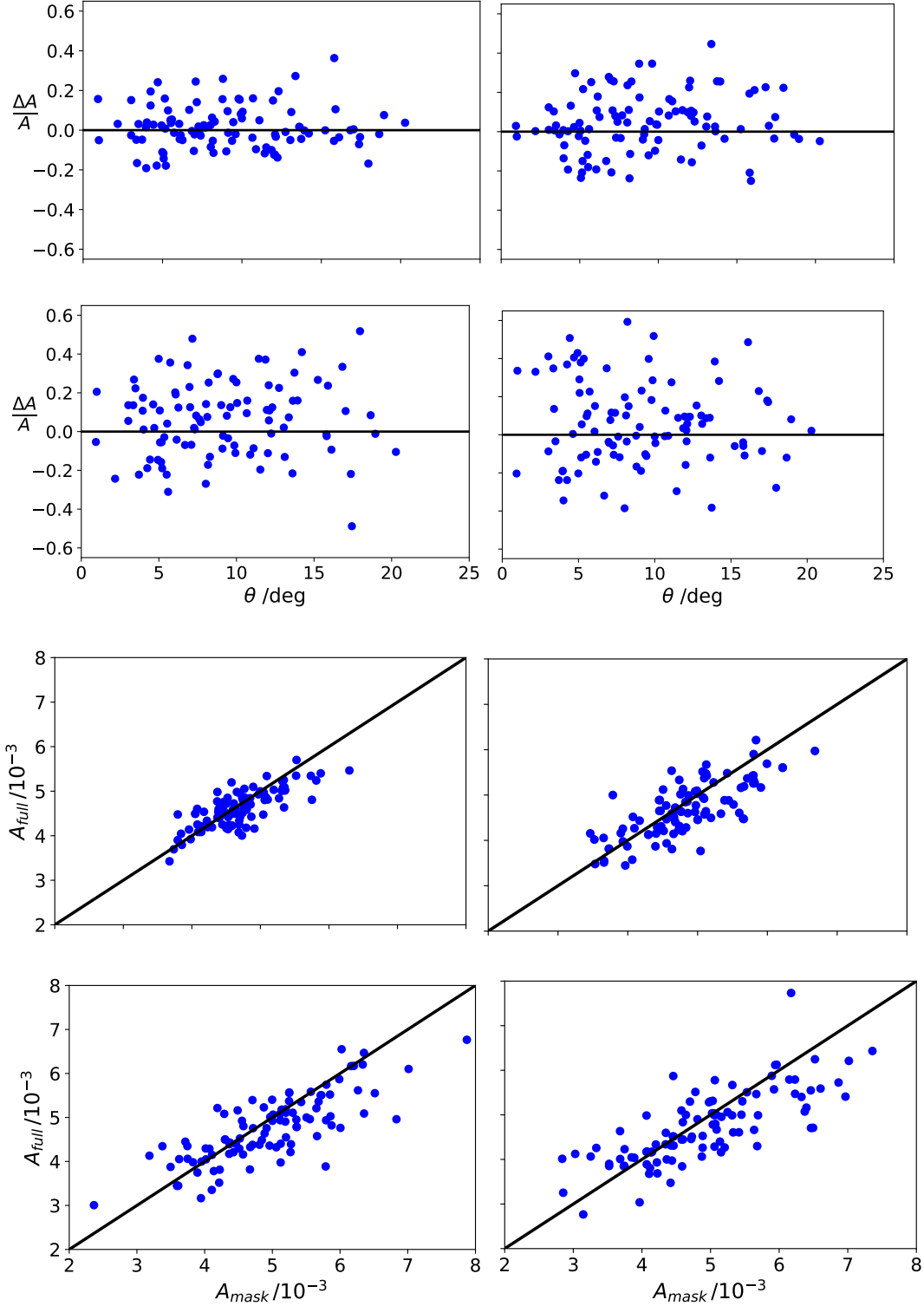


Figure A.10: The top four show the relative error of the measured amplitude $\Delta A/A$ against the angle θ between the estimated direction and the CMB direction with masked-sky. The flux threshold increases from left to right and from top to bottom, with the fluxes 1 μJy , 5 μJy , 10 μJy and 22.8 μJy . The bottom four are the the amplitude measured for full-sky against the measured for masked-sky.