
Signalrekonstruktion in der Radioastronomie

Signal Inference in Radio Astronomy

Maksim Greiner



München 2016

**Signalrekonstruktion in der
Radioastronomie**
Signal Inference in Radio Astronomy

Maksim Greiner

Dissertation
an der Fakultät für Physik
der Ludwig–Maximilians–Universität
München

vorgelegt von
Maksim Greiner
aus Berlin

München, den 31.5.2016

Erstgutachter: PD Dr. Torsten A. Enßlin
Zweitgutachter: Prof. Dr. Jochen Weller
Tag der mündlichen Prüfung: 27. Juli 2016

Contents

Zusammenfassung	xiii
Summary	xv
1 Introduction	1
1.1 The Milky Way	1
1.1.1 The interstellar medium	2
1.1.2 Observables related to the free electron density	5
1.2 Radio astronomy	9
1.2.1 Instrumentation	10
1.3 Signal inference and imaging	14
1.3.1 Frequentist and Bayesian methods	15
1.3.2 Estimators from quadratic cost functions	16
1.3.3 Other estimators	18
1.3.4 Evaluating posterior distributions	18
1.4 Outline of this thesis	20
2 Reconstructing the Galactic free electron density	21
2.1 Introduction	21
2.2 Reconstruction algorithm	23
2.2.1 Signal model	23
2.2.2 Necessary probability density functions	25
2.2.3 Galactic profile inference	31
2.2.4 Filter equations	32
2.3 Application to simulated data	33
2.3.1 Galaxy model	33
2.3.2 Simulated population and survey	35
2.3.3 Simulated dispersion measures and distances	35
2.3.4 Algorithm setup	38
2.4 Simulation evaluation	38
2.4.1 Density in the midplane	38
2.4.2 Vertical fall-off	39
2.5 Summary and conclusions	44

3	fastRESOLVE	47
3.1	Introduction	47
3.2	Algorithm	49
3.2.1	Signal model	49
3.2.2	Irregular sampling	50
3.2.3	The likelihood	52
3.2.4	Inference	53
3.3	Application	56
3.3.1	Abell 2199	56
3.4	Discussion & Conclusion	63
4	Log-transforming the matter power spectrum	65
4.1	Introduction	66
4.2	The conversion	67
4.2.1	Notation and definitions	67
4.2.2	The forward conversion	68
4.2.3	The backward conversion	69
4.3	The matter power spectrum	70
4.3.1	The mildly non-linear regime	72
4.3.2	The non-linear regime	72
4.3.3	Smearing of baryon acoustic oscillations	74
4.4	Discussion and conclusions	75
5	Conclusions and outlook	79
A	Appendix to reconstructing the Galactic free electron density	81
A.1	Parameters of the power spectrum prior	81
A.2	Functional derivatives of the Hamiltonian	82
A.3	Convergence	83
A.4	Goodness of fit	87
A.5	Reconstruction with real power spectrum and profile functions	88
A.6	Uncertainty map	88
A.7	Uncertainties of the vertical fall-off	89
B	Appendix to fastRESOLVE	91
B.1	Aliasing	91
B.2	Kullback-Leibler Divergence extremization	92
B.3	Approximate propagator	94
B.4	Likelihood variance estimation	95
B.5	Point source procedure	96
B.6	Simulation tests	97
B.6.1	Diffuse emission	97
B.6.2	point-like emission	99

B.7	Signal to noise	99
C	Appendix to log-transforming the matter power spectrum	101
C.1	Generating log-normal fields from a given power spectrum	101
C.2	The discretized conversion formula	101
C.3	The conversion for spherical harmonics	104
C.4	Supplements to Sect. 4.3	105
C.4.1	Log-distances	105
C.4.2	The effect of different grids	105
C.4.3	Strength of the smearing of BAO	108
C.4.4	Large-scale bias	111
C.4.5	Generating matter densities in spaces spanning in redshift	113
	Bibliography	115
	Danksagung	121

List of Figures

1.1	A cartoon of the Milky Way	2
1.2	All-sky free-free emission	6
1.3	The Galactic Faraday sky	7
1.4	Schematic illustration of a pulsar	8
1.5	The Arecibo radio telescope	10
1.6	The Very Large Array	12
1.7	Overview of the SKA configuration	14
1.8	The solution strategy employed in this thesis	15
2.1	Modeling diagram	25
2.2	Power spectrum of NE2001	34
2.3	Positions of simulated pulsars	35
2.4	Histogram of distance uncertainties	37
2.5	Reconstructions of medium strength fluctuation models	40
2.6	Recovered Galactic profile – horizontal	41
2.7	Reconstruction of models with varying fluctuation strengths	42
2.8	Reconstruction of contrast enhanced model	43
2.9	Recovered Galactic profile – vertical	43
2.10	Recovered z-profile	45
3.1	Total intensity contours of 3C 388 at 8415 MHz	57
3.2	Total intensity of 3C 388 at 8415 MHz	58
3.3	Superposition of fastRESOLVE and CLEAN	60
3.4	Relative uncertainty of the fastRESOLVE image	60
3.5	Total intensity of 3C 388 at 1665 MHz	62
4.1	Power enhancement w.r.t. linear spectrum – redshift 1 - 7	73
4.2	Power enhancement w.r.t. perturbation theory – redshift 1 - 7	73
4.3	Power enhancement w.r.t. linear spectrum – redshift 0 - 1	74
4.4	Power enhancement w.r.t. perturbation theory – redshift 0 - 1	75
4.5	Smearing of baryon acoustic oscillations	76
A.1	Power spectrum convergence	84
A.2	Propagated distance variance convergence	85

A.3	z-profile convergence	86
A.4	Radial profile convergence	86
A.5	Reconstruction with real power spectrum	89
A.6	Uncertainty map	90
B.1	Simulation results for purely diffuse emission	98
B.2	Simulation results for diffuse and point-like emission	100
C.1	Backwards conversion	102
C.2	Random log-normal field	102
C.3	Grid comparison forward	109
C.4	Grid comparison backward	110

List of Tables

2.1	Simulated data sets	36
2.2	Estimated sampling distances	37
3.1	Summary of the observations used in chapter 3	57
3.2	Flux of 3C 338 at 8415 MHz	59
A.1	Reduced χ^2 values 1	87
A.2	Reduced χ^2 values 2	88
C.1	Log-distances – different redshifts	106
C.2	Grids used for conversion	107
C.3	Log-distances – different grids	108
C.4	BAO best fit smoothing scales	109
C.5	Large-scale bias factors	112

Zusammenfassung

Diese Dissertation befasst sich mit dem Rekonstruieren von unvollständig gemessenen Signalen in der Radioastronomie. Es werden zwei bildgebende Algorithmen entwickelt, die im Formalismus der Informationsfeldtheorie hergeleitet werden. Beide basieren auf dem Prinzip der Bayesischen Analyse, die Informationen aus der unvollständigen Messung werden dabei durch a priori Informationen ergänzt. Hierfür werden beide Informationsquellen in Form von Wahrscheinlichkeitsdichten formuliert und zu einer a posteriori Wahrscheinlichkeitsdichte zusammengeführt. Die a priori Informationen werden dabei minimal gehalten und beschränken sich auf die Annahme, dass das ursprüngliche Signal bezüglich des Ortes nicht beliebig stark fluktuiert. Dies erlaubt eine statistische Abschätzung des ursprünglichen Signales auf allen Größenskalen.

Der erste bildgebende Algorithmus errechnet eine Abschätzung der dreidimensionalen freien Elektronendichte im interstellaren Medium der Milchstraße aus Dispersionsmessungen von Pulsaren. Die Dispersion der Radiostrahlung eines Pulsars ist proportional zu der Gesamtanzahl der freien Elektronen auf der Sichtlinie zwischen Pulsar und Beobachter. Jede gemessene Sichtlinie enthält somit Informationen über die Verteilung von freien Elektronen im Raum. Das Rekonstruktionsproblem ist damit ein Tomographieproblem ähnlich dem in der medizinischen Bildgebung. Anhand einer Simulation wird untersucht, wie detailliert die Elektronendichte mit Daten des noch im Bau befindlichen Square Kilometre Array (SKA) rekonstruiert werden kann. Die Ergebnisse zeigen, dass die großen Strukturen der freien Elektronendichte der Milchstraße mit den Daten des SKA rekonstruiert werden können.

Der zweite bildgebende Algorithmus trägt den Namen fastRESOLVE und rekonstruiert die Intensität von Radiostrahlung anhand von Messdaten eines Radiointerferometers. fastRESOLVE baut auf dem bestehenden Algorithmus RESOLVE auf. fastRESOLVE erweitert dessen Funktionalität um die separate Abschätzung von Punktquellen und rekonstruiert simultan auch die Messunsicherheit. Gleichzeitig ist fastRESOLVE etwa 100 mal schneller. Ein Vergleich des Algorithmus' mit CLEAN, dem Standardalgorithmus in der Radioastronomie, wird anhand von Beobachtungsdaten des Galaxienhaufens Abell 2199, aufgenommen mit dem Very Large Array, durchgeführt. fastRESOLVE kann feinere Details des Intensitätsverlaufs rekonstruieren als CLEAN. Gleichzeitig erzeugt fastRESOLVE weniger Artefakte wie negative Intensität. Außerdem liefert fastRESOLVE eine Abschätzung der Rekonstruktionsunsicherheit. Diese ist wichtig für die wissenschaftliche Weiterverarbeitung und kann mit CLEAN nicht errechnet werden.

Weiterhin wird ein Verfahren entwickelt, mit dem die Leistungsspektren von Gaußschen Feldern und die von log-normal Feldern ineinander umgewandelt werden können. Dieses ermöglicht die Verjüngung des Leistungsspektrums der großskaligen Dichtestruktur des Universums, was durch Vergleiche mit einer störungstheoretischen Methode und einem kosmischen Emulator validiert wird.

Summary

This dissertation addresses the problem of inferring a signal from an incomplete measurement in the field of radio astronomy. Two imaging algorithms are developed within the framework of information field theory. Both are based on Bayesian analysis; information from the incomplete measurement is complemented by a priori information. To that end both sources of information are formulated as probability distributions and merged to an a posteriori probability distribution. The a priori information is kept minimal. It reduces to the assumption that the real signal does not fluctuate arbitrarily strong with respect to position. This construction allows for a statistical estimation of the original signal on all scales.

The first imaging algorithm calculates a three-dimensional map of the Galactic free electron density using dispersion measure data from pulsars. The dispersion of electromagnetic waves in the radio spectrum that a pulsar emits is proportional to the total number of free electrons on the line of sight between pulsar and observer. Therefore, each measured line of sight contains information about the distribution of free electrons in space. The reconstruction problem is a tomography problem similar to the one in medical imaging. We investigate which level of detail of the free electron density can be reconstructed with data of the upcoming Square Kilometre Array (SKA) by setting up a simulation. The results show that the large-scale features free electron density of the Milky Way will be reconstructible with the SKA.

The second imaging algorithm is named fastRESOLVE. It reconstructs the radio intensity of the sky from interferometric data. fastRESOLVE is based on RESOLVE, but adds the capability to separate point sources and to estimate the measurement uncertainty. Most importantly, it is 100 times faster. A comparison of the algorithm with CLEAN, the standard imaging method for interferometric data in radio astronomy, is performed using observational data of the galaxy cluster Abell 2199 recorded with the Very Large Array. fastRESOLVE reconstructs finer details than CLEAN while introducing fewer artifacts such as negative intensity. Furthermore, fastRESOLVE provides an uncertainty map. This quantity is important for proper scientific use of the result, but is not available using CLEAN.

Furthermore, a formalism is developed, which allows the conversion of power spectra of Gaussian fields into the power spectra of log-normal fields and vice versa. This allows the rejuvenation of the power spectrum of the large-scale matter distribution of the Universe. We validate the approach by comparison with a perturbative method and a cosmic

emulator.

Chapter 1

Introduction

In this thesis we develop inference methods for various applications in radio astronomy. We develop a tomography method to study the free electron density in the Milky Way, an imaging algorithm for radio interferometers, and a statistical conversion method to rejuvenate the large-scale matter density. In this introductory chapter we briefly introduce the Milky Way and radio interferometers summarizing the most important physical aspects and addressing the relevance as a research topic. We furthermore discuss the current state of knowledge and open questions in these topics as well as the standard methods in inference to motivate the efforts made in the main part of this thesis.

1.1 The Milky Way

The Milky Way is a spiral galaxy, which means that most of its stars, gas, and dust are contained within a relatively flat disk with spiral-shaped density enhancements forming four spiral arms. The disk consists of a thin stellar disk and an even thinner disk made out of gas and dust. The central part of the disk is embedded in a stellar bulge, while the whole disk is embedded into a halo made out of gas, dust, and stars. The halo contains over 150 globular clusters, small spherical regions of high star density, which appear to be the oldest components of the Milky Way. Additionally, there are highly energetic cosmic rays made out of charged particles which are trapped in the magnetic field of the Milky Way. Finally, the whole arrangement of visible matter is contained in a large halo of dark matter, the dark halo. In Fig. 1.1 we show a cartoon of the Milky Way's structure.

The Milky Way is estimated to contain about 10^{11} stars (see e.g. Maoz 2016) and its total visible¹ mass is estimated to be 10^{11} solar masses (M_{\odot}) half of which is in stars (the average mass of a star in the Milky Way is half a Solar mass). The remaining visible mass is mostly gas and dust. The mass of the disk is roughly $10^{10} M_{\odot}$, 90% of it is in stars. The mass distribution of the disk follows approximately an exponential profile in the radial and

¹Visible in the sense of not being dark matter.

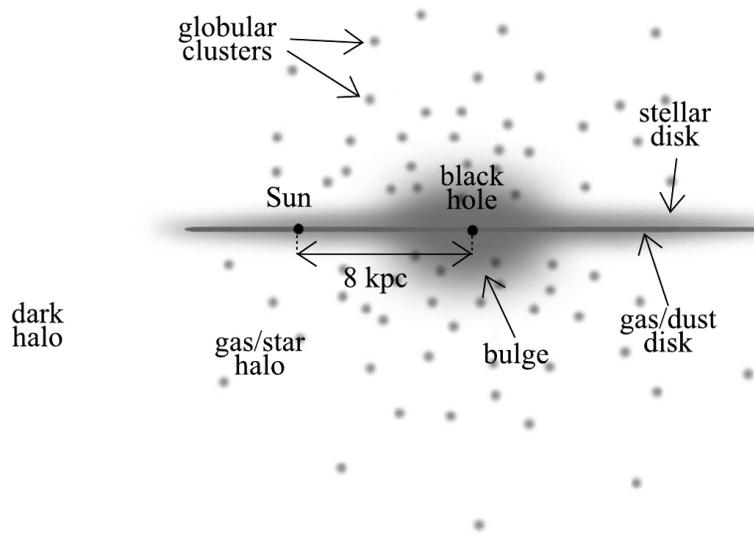


Figure 1.1: A cartoon of the Milky Way. Figure taken from Maoz (2016).

the vertical direction,

$$\rho(r, z) \approx \rho_0 \exp\left(-\frac{r}{H_r}\right) \exp\left(-\frac{|z|}{H_z}\right). \quad (1.1)$$

The radial scale length is $H_r = 2.5 \pm 0.5$ kpc. The vertical scale height H_z is about 400 pc. The Solar system is located 25 pc above the mid plane of the disk at a distance of 8.3 ± 0.3 kpc from the Galactic center. It is therefore located in the outer part of the Milky Way's disk.

The stellar bulge is of spheroidal shape with a size of about 1 kpc. Its density drops off with the cubic power of the distance to the center. The Galactic halo follows a similar profile but extends out to 50 kpc. The stars in the bulge and halo are among the oldest stars in the Milky Way with ages between 10 and 14 billion years.

The Galactic center contains a compact stationary object name Sagittarius A*. With a size of roughly 1 astronomical unit (AU), a mass of $4 \times 10^6 M_\odot$ but very little luminosity it is assumed to be a supermassive black hole. The dark halo of the Milky Way contains dark matter with a total mass of nearly $10^{12} M_\odot$ with a density profile proportional to $1/r^2$. This means about 90% of the Milky Ways total mass is in dark matter.

1.1.1 The interstellar medium

The gas, dust, magnetic field as well as the relativistic particles that populate the space between the stars form the interstellar medium. In the disk it makes up roughly 10% of the visible mass. The gas is mostly Hydrogen ($\sim 70\%$). Roughly half the mass of the gas is concentrated in relatively small clouds occupying only about 1 – 2% of the total

volume (see Ferrière 2001, and references therein). These clouds can be categorized into dark molecular clouds ($T \sim 10 - 20$ K), which are intransparent to visible light, diffuse atomic clouds ($T \sim 100$ K), which are transparent apart from absorption lines, and clouds which consist of an inhomogeneous mixture of the two. The remaining gas is extremely dilute with less than half a particle per cubic centimeter in the vicinity of the Sun. This gas is divided into the warm atomic phase ($T \sim 6000 - 10000$ K), the warm ionized phase ($T \sim 8000$ K), and the hot ionized phase ($T \sim 10^6$ K). Roughly 15% of all gas mass belong to the ionized phases ($\sim 1.6 \times 10^9 M_\odot$), mostly to the warm ionized phase.

The warm ionized gas is not homogeneously distributed. There are underdense and overdense regions and research indicates that there are spatial fluctuations on all scales (Armstrong et al. 1995). There is a great demand in the astronomical community for a three-dimensional map of the free electron density, since it plays a role in a number of effects which are interesting in the context of radio astronomy: dispersion, Faraday rotation, and emission processes.

Dispersion

The ionized gas disperses electromagnetic radiation since the unbound free electrons oscillate with the electric field of traversing light waves. This makes the effective speed of light v frequency dependent,

$$v = c \sqrt{1 - \frac{\nu_p^2}{\nu^2}}, \quad (1.2)$$

where c is the speed of light in vacuum, ν is the frequency of the radiation, and ν_p the plasma frequency. The plasma frequency squared is proportional to the density of free electrons, n_e . At the densities encountered in the warm ionized gas the plasma frequency stays below 100 kHz and for frequencies in the range of MHz to GHz Eq.(1.2) can be linearized in n_e . The time it takes for radiation from a source at distance d to arrive at Earth grows therefore linearly with respect to the line integral of the free electron density and can be written as

$$t = \frac{d}{c} + \frac{k_{DM}}{\nu^2} \int_0^d dr n_e, \quad (1.3)$$

where k_{DM} is the dispersion constant². The line integral in this equation is commonly referred to as the dispersion measure DM ,

$$DM = \int_0^d dr n_e. \quad (1.4)$$

²The dispersion constant is $k_{DM} = e^2/(2\pi m_e c)$, where e is the electron charge and m_e is the electron mass.

Faraday rotation

Another important way in which the warm ionized gas influences electromagnetic radiation is the Faraday effect. In the presence of free electrons and a magnetic field component parallel to the direction of radiation the angle of polarization gets rotated. This can be understood by regarding the linear polarized wave as a superposition of a right-circular and left-circular polarized wave. The electric field of the wave acts on the free electrons, creating a current that counteracts the electric field. However, due to the presence of the magnetic field and the resulting Lorentz force, one direction is preferred over the other. This results in different speeds of light for the right-circular and left-circular component and thus in a rotation of the polarization angle. Accumulating this effect over the whole line-of-sight between source and observer yields a difference in polarization angle of

$$\Delta_\phi = \frac{e^3 \lambda^2}{2\pi m_e^2 c^4} \int_0^d n_e \vec{B} \cdot d\vec{r}. \quad (1.5)$$

The wavelength independent part of this equation is called Faraday depth (see e.g. Schlickeiser 2002).

Emission processes

The warm ionized gas also emits radiation through recombination and scattering between the electrons and ions. The most visible emission from recombination is H α emission, which comes from the photons that get emitted when an electron in a Hydrogen atom transitions from the third to the second lowest energy level. The thereby emitted radiation has a wavelength of 656 nm and is easily identified as a spectral line. When an electron and an ion scatter without getting bound to each other they emit Bremsstrahlung. In astronomy this process is called free-free emission. Free-free emission is not concentrated to one emission line but emits a whole spectrum of radiation, where the emitted power per unit frequency is inversely proportional to the frequency squared. Both emission processes, H α and free-free, involve pairwise interactions between electrons and ions. The ratio in which these events happen is therefore proportional to the free electron density squared. The integral of the squared density over a line-of-sight is proportional to the received radiation (ignoring absorption effects). It is called the emission measure,

$$EM = \int_0^d dr n_e^2. \quad (1.6)$$

The demand for a map

As outlined in this section, the warm ionized gas in the interstellar medium alters electromagnetic radiation passing through it and also emits radiation itself. Whenever distant objects are observed the recorded radiation has been in some way affected by the free

electrons. Therefore, many studies have to take great care in treating these foreground effects in order not to arrive at false conclusions. Measurements of the cosmic microwave background (such as the WMAP and Planck missions, see e.g. Bennett et al. 2013; Planck Collaboration et al. 2015) have to be taken at several frequencies in order to (among other effects) map out the free-free contribution to the data. Measurements of extragalactic Faraday rotation probe the strength of the extragalactic magnetic field, a quantity of great interest to understand the evolution of the Universe, but the unknown Galactic contribution has to be subtracted. There are promising approaches to tackle this problem using information theory (e.g. Oppermann et al. 2015; Vacca et al. 2015), but it still remains challenging. Knowing the distribution of free electrons could significantly aid these efforts, as the Faraday effect depends on both, the magnetic field and the free electron density. Furthermore, understanding the distribution of free electrons could help understanding the evolution of the Milky Way since it is an essential part of its structure and is shaped by the complex interplay of ionizing radiation, recombination, and heating and cooling processes.

1.1.2 Observables related to the free electron density

Several observables contain information about the free electron density. They are in one way or the other related to the physical effects described in this section. Free-free emission can in principle be measured over the whole sky and several satellite based missions (such as the WMAP and Planck missions, see e.g. Bennett et al. 2013; Planck Collaboration et al. 2015) have measured it. In Fig. 1.2 we show the free-free all-sky map published by Planck Collaboration et al. (2015). However, they do not measure the free-free emission alone, but rather the sum of all sources of emission, and the free-free emission does not dominate at any frequency. The components are separated by combining measurements at different frequencies with the respective spectral formulae of the components. For the WMAP satellite the free-free antenna temperature at frequency ν is related to the emission measure EM through the spectral formula by Bennett et al. (2003) (for the Planck satellite there is a similar expression, see Table 4 in Planck Collaboration et al. 2015)

$$T_A^{\text{WMAP}}(\mu\text{K}) = 1.44EM_{\text{cm}^{-6}\text{pc}} \frac{1 + 0.22 \ln(T_e/8000 \text{ K}) - 0.14 \ln(\nu/41 \text{ GHz})}{(\nu/41 \text{ GHz})^2 (T_e/8000 \text{ K})^{1/2}}. \quad (1.7)$$

The temperature T_e of the electrons is not known exactly. It is usually set to the mean value of the warm ionized medium (8000 K). In reality the temperature presumably fluctuates with respect to position, also along a line-of-sight. Furthermore, the errors of inferred values are spatially correlated through the instruments response, but the employed fitting procedure yields only point-wise error estimates. This makes further processing of the resulting EM map to infer n_e challenging. H_α emission is more easily identified due to its discrete spectral line, but at a wavelength of 656 nm it is subject to absorption by dust particles, which has to be corrected for in order to infer EM , and the relation between H_α and EM depends on the temperature, too. The most important caveat to infer n_e with free-free and H_α emission is however the fact that all line-of-sight integrals of the inferred

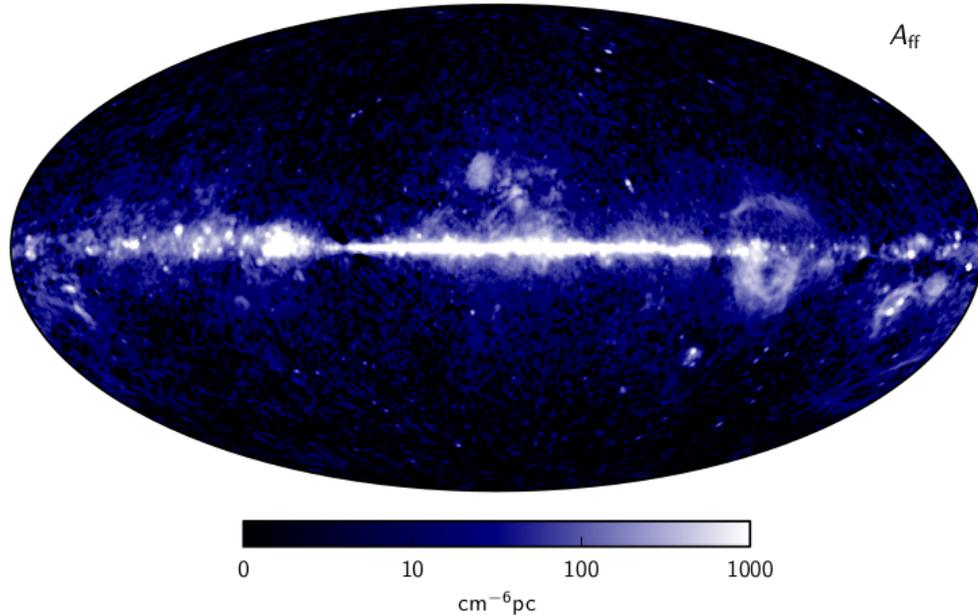


Figure 1.2: The all-sky free-free emission measure (EM) published by the Planck collaboration (Fig. 9 from Planck Collaboration et al. 2015).

EM values go from the observer to infinity. There is no depth information in such EM values.

Faraday depth can be inferred by measuring the polarization of an object at several frequencies. There are some complications that make measured Faraday rotation values prone to errors (see Oppermann et al. 2012, 2015), but if sufficiently many wavelengths are used, the Faraday depth can be inferred reliably through a procedure called rotation measure synthesis (Brentjens & de Bruyn 2005), for both intragalactic and extragalactic sources. Oppermann et al. (2015) used the known extragalactic sources to infer an all-sky map of the Galactic Faraday depth, see Fig. 1.3. If their distances are known, intragalactic sources provide even three-dimensional information. However, as can be seen from Eq. (1.5), to infer n_e from Faraday depths involves the unknown Galactic magnetic field \vec{B} (or at least its parallel component). This reduces the information content of Faraday depths about the free electron density significantly.

The most direct way of measuring the free electron density on scales of parsecs and above is through dispersion measures of pulsars.

Pulsars

Atomic fusion is only exothermic for elements lighter than iron. Large stars that go through the whole fusion chain therefore build up more and more iron which (due to its density) forms an iron core. When the mass of the iron core reaches the Chandrasekhar limit ($1.4 M_\odot$), the core collapses under its own pressure and protons and electrons merge into neutrons through inverse beta decay (see e.g. Lieb & Yau 1987). The outer layers of the star

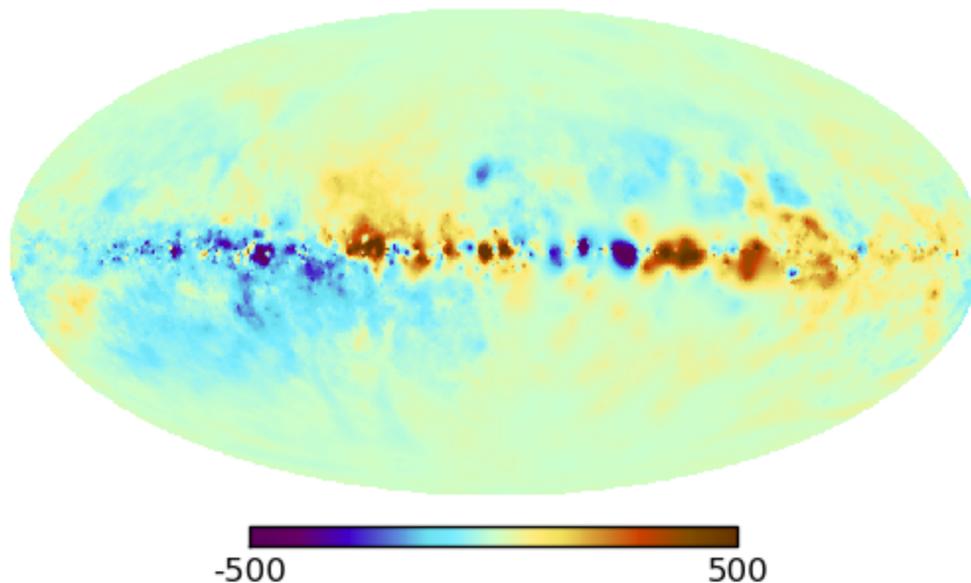


Figure 1.3: The Galactic Faraday depth in units of rad/m^2 (Fig. 15 from Oppermann et al. 2015).

are blown away and the whole star explodes in a core-collapse supernova leaving a highly compact remnant (for a more complete discussion see Heger et al. 2003, and references therein). The remnant becomes either a neutron star or black hole depending on its mass³ (see Chamel et al. 2013). The core collapse increases the density of the core by about 14 orders of magnitude. Since the angular momentum and the magnetic moment of the core are now concentrated on a much more compact object, the resulting neutron star can have magnetic fields of the order of $10^8 - 10^{15}$ Gauss and spin rapidly with periods as low as a few milliseconds. If the rotational axis and the magnetic dipole axis are not aligned, the magnetic field varies rapidly and large electric fields are induced which in turn accelerate ions and electrons in the vicinity of the neutron star (Gold 1968; Pacini 1968). The exact mechanism is not yet fully understood, but it results in a net flux of radiation in the direction of the magnetic dipole axis, see Fig. 1.4. Since this axis is subject to rotation a stationary observer sees the radiation as periodic pulses on the sky. Neutron stars which exhibit this kind of characteristic radiation are called pulsars.

Pulsars as probes of the free electron density

Pulsars emit radiation over a wide frequency spectrum. In the radio band this radiation passes through Galactic dust and other obstacles relatively unhindered. The radiation in one pulse is emitted at the same time for all frequencies, but the speed of light is frequency-dependent due to dispersion by the warm ionized interstellar medium. Therefore,

³The largest observed neutron star mass is about $2 M_{\odot}$ and the smallest observed black hole mass is about $8 M_{\odot}$.

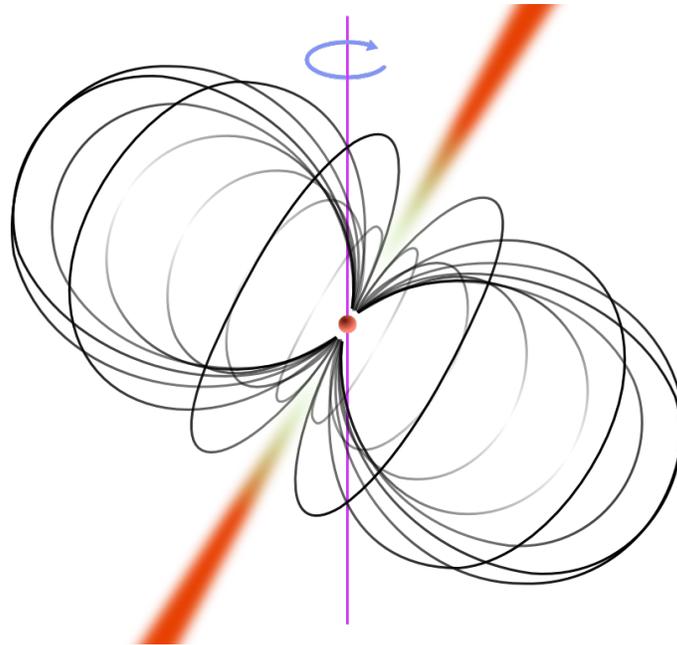


Figure 1.4: Schematic illustration of a pulsar. The sphere in the middle is the neutron star which rotates around the purple axis, the curved black lines illustrate the magnetic field and the red cones illustrate the emission (image taken from Wikimedia Commons and modified, published under the Creative Commons license).

measuring the arrival time of the pulse at a number of frequencies allows the observer to directly infer the line-of-sight integral of the free electron density between observer and pulsar (see Eq. (1.3)). Complemented with the distance of the pulsar this yields an excellent probe of the free electron density that is independent of any assumptions on other quantities. In consequence, most efforts on modeling or reconstructing the free electron density have focused on using pulsar dispersion measures.

The main challenge is finding enough pulsars since their emission is usually faint. Once a pulsar is discovered, measuring its dispersion measure is relatively easy. However, what is problematic is the distance information. The most reliable estimate of the distance of a pulsar is through parallax measurements, but these require instruments with a very high angular resolution. Pulsars can be observed best in the radio band or sometimes in X-rays (Bildsten et al. 1997); in the optical regime dust absorption and stellar light often prohibit their observation. Both frequency bands are not ideal for high angular resolutions, radio observations are limited by their large wavelengths (see section 1.2) and X-rays cannot be focused by conventional optics. Consequently, while there are about 2500 pulsars with known dispersion measures in the ATNF Pulsar Catalogue⁴ the distances are only known for fewer than 100 of them. This sparsity of data is the main challenge in estimating the free electron density from pulsar dispersion measurements. New types of radio observatories will have to be built in order to increase the abundance of data. The upcoming Square Kilometre Array (aimed to be operational in 2020) will provide unprecedented resolution and sensitivity and will likely find about 10000 pulsars and measure their dispersion measures and parallaxes (Smits et al. 2009, 2011). This will boost our knowledge on the three-dimensional distribution of free electrons in the Milky Way. We will discuss this further in section 1.2.1 and chapter 2.

1.2 Radio astronomy

Whichever observable one uses to probe the interstellar medium, one needs to observe deep into the dense part the Milky Way. Radiation in the optical regime is far from ideal to this purpose, since it gets absorbed by dust and drowned by stars. A spectrum which passes through dust and is not emitted in great intensity by stars is found in the radio regime, which brings us to the topic of radio astronomy.

Radio astronomy is concerned with observing objects that emit electromagnetic radiation in the radio spectrum. The astronomically relevant part of the radio spectrum ranges from 10^{-3} to 10^1 meters in wavelength. Radio waves in that range pass the Earth's atmosphere relatively unhindered making them very suitable for ground-based observations. At longer wavelengths it is the scattering and absorption by the ionosphere that limits the observability, at shorter wavelengths it is the interaction with water vapor in the troposphere. Therefore, observations in this spectrum have proven to be a very useful complement to observations in the optical regime and the field of radio astronomy has been an active field

⁴ The Australia Telescope National Facility Pulsar Catalogue (Manchester et al. 2005).
<http://www.atnf.csiro.au/people/pulsar/psrcat/>



Figure 1.5: The Arecibo radio telescope in Puerto Rico (image courtesy of the National Science Foundation).

of research since the 1930s. For a detailed introduction into the topic we refer the reader to Thompson et al. (1986) or any other book on radio astronomy.

1.2.1 Instrumentation

Radio waves can be received with either dipole antennae, which are directly illuminated, or dishes, which focus the the radiation on a receiver. Most radio telescopes are found in the form of dishes, as they can be used more efficiently over a broader wavelength range. Radio observations have several advantages compared to optical observations. Most notably, one can observe day and night and the observation is disturbed less by weather and atmosphere. Despite the advantages there are two disadvantages to radio waves compared to optical waves. First, radio waves from astronomical sources are very faint. Second, the smallest angle a telescope can resolve is proportional to the observed wavelength as can be seen from the Rayleigh criterion,

$$\theta = 1.22 \frac{\lambda}{D}, \quad (1.8)$$

where θ is the smallest resolvable angle, λ is the wavelength, and D is the diameter of the telescope. To increase both, the sensitivity and the resolution, radio dishes have to be very large in diameter. The largest dish-like radio telescope is the Arecibo radio telescope with a diameter of 305 meters (see Fig. 1.5).

Building dishes with diameters of several hundred meters involves substantial technical challenges and high costs. The large size of the Arecibo dish could only be reached by having a fixed dish, which can only observe a limited portion of the sky. Still, the resolution of Arecibo radio telescope is 10 times worse than that of a common 10 cm optical telescope due to the orders of magnitude difference of the observed wavelengths. To increase the resolution further radio astronomers therefore use arrays of radio dishes where the effective diameter in Eq. (1.8) is governed by the distance between the antennae instead of their individual sizes. This technique is called aperture synthesis.

Aperture synthesis

By combining several radio dishes to a radio interferometer astronomers can overcome the prohibitive efforts of building larger and larger single dishes to increase the resolution. The dishes are combined into one instrument by correlating the electromagnetic signals of the dished pairwise. This technique is called aperture synthesis.

Consider two antennae, labeled antenna 1 and 2 and a source so distant that the incoming wave front is planar. The electric signals received by the antennae around a frequency ν are

$$\begin{aligned} E_1(t) &= E_0(t, \hat{x}) \exp\left[i2\pi\left(t\nu + \vec{x} \cdot \vec{k}\right)\right] \\ \text{and} \quad E_2(t) &= E_0(t, \hat{x}) \exp\left[i2\pi\left(t\nu + (\vec{x} - \Delta\vec{x}) \cdot \vec{k}\right)\right]. \end{aligned} \quad (1.9)$$

Here, \vec{x} is the position of the source and $\Delta\vec{x}$ is the position of antenna 2 (both with respect to the location of antenna 1). The wave vector $\vec{k} \equiv \frac{-\vec{x}}{|\vec{x}|\lambda}$ is the same for both antennae due to the planar wave fronts. The time dependency in $E_0(t, \hat{x})$ describes variations in time around the plane wave oscillations. Since the whole sky (or at least the field of view of the antenna) contribute to the received signal, this equation has to be integrated over all directions $\hat{x} \equiv \frac{\vec{x}}{|\vec{x}|}$ leading to

$$\begin{aligned} E_1(t) &= \int_{S^{(2)}} d^2\hat{x} E_0(t, \hat{x}) \exp\left[i2\pi\left(t\nu + \vec{x} \cdot \frac{\hat{x}}{\lambda}\right)\right], \\ E_2(t) &= \int_{S^{(2)}} d^2\hat{x} E_0(t, \hat{x}) \exp\left[i2\pi\left(t\nu + (\vec{x} - \Delta\vec{x}) \cdot \frac{\hat{x}}{\lambda}\right)\right]. \end{aligned} \quad (1.10)$$

The signals from antenna 1 and 2 are correlated by averaging their product over time. This forms the visibility V_{12} ,

$$V_{12} = \frac{1}{T} \int dt \bar{E}_1(t) E_2(t), \quad (1.11)$$

where the bar denotes complex conjugation. Under the assumption that the radiation of different sources is uncorrelated,

$$\frac{1}{T} \int dt \bar{E}_0(t, \hat{x}) E_0(t, \hat{y}) \approx \delta(\hat{x} - \hat{y}) |E_0(\hat{x})|^2, \quad (1.12)$$



Figure 1.6: The Very Large Array in New Mexico (USA) (image courtesy of the National Radio Astronomy Observatory/AUI).

(where $\delta(\cdot)$ denotes the Dirac delta distribution) the correlated signal becomes

$$V(\vec{u}) = \int_{S^{(2)}} d^2\hat{x} I(\hat{x}) \exp[i2\pi \vec{u} \cdot \hat{x}], \quad (1.13)$$

where $\vec{u} \equiv \Delta\vec{x}/\lambda$ is the displacement vector between two antennae divided by the observed wavelength and $I(\hat{x}) \equiv |E_0(\hat{x})|^2$ is the intensity on the sky. As one can see, this equation is essentially a spherical Fourier transform. However, only the vectors \vec{u} that correspond to antenna displacements are probed by the measurements.

The most famous radio interferometer is the Very Large Array (VLA) built in the 70s (see Fig. 1.6). In 2012 it has been modernized and renamed to Karl G. Jansky Very Large Array. It can change its antenna configuration to adapt to different observation strategies. In configuration A, where the antennae are furthest apart, it can reach a resolution 100 times higher than the Arecibo radio telescope. This is the advantage of the aperture synthesis technique. It can be taken even further by distributing the antennae over several thousand kilometers. This method is called very-long-baseline interferometry.

There are disadvantages to aperture synthesis. As can be seen from Eq. (1.13) the intensity is only probed at “Fourier”-like vectors \vec{u} which correspond to antenna displacements, whereas a single dish probes in a continuous space. Since the dishes cannot be positioned arbitrarily close to each other there is usually a “gap” in the center of the measured Fourier-like space which means that largest scales and the total flux cannot be measured⁵. This makes the information recorded through aperture synthesis much more

⁵ In principle the total flux could be measured by correlating a dish’s signal with itself, but the signal of an individual dish in an interferometer is usually noise dominated. This noise can be averaged out when correlating the signals of two dishes, but not when correlating the signal with itself.

incomplete compared to a single dish of the same size. Furthermore, the total collecting area is typically smaller in an interferometer (all dish areas combined) than in a single dish, which lowers the sensitivity. These disadvantages make the imaging of interferometric data much more challenging. It will be the main topic of chapter 3.

Notable instruments

Apart from the already mentioned Arecibo telescope in the United States (US), which is a fixed full dish telescope, there are large steerable single dish telescopes such as the Green Bank Telescope in the US (100 m), the Effelsberg 100-m Radio Telescope in Germany, the Lovell Telescope in the United Kingdom (76 m), and the Sardinia Radio Telescope (64 m) in Italy. Notable radio interferometers include the aforementioned VLA and the Very Long Baseline Array in the US, the Westerbork Radio Telescope in the Netherlands, the Atacama Large Millimeter Array in Chile, the Australian Telescope Compact Array, and the Giant Metrewave Radio Telescope in India. Since about 2012 a new type of radio interferometer is operational in Europe, LOFAR, the Low-Frequency Array. Instead of using several dishes to form the interferometer, LOFAR uses vast arrays of omnidirectional dipole antennae. All processing that the geometry of a dish does automatically is emulated in software. In principle this means that the whole sky can be observed without reorienting the telescope. In practice the limitations come from the huge demands for computing hardware and software.

The Square Kilometre Array

The Square Kilometre Array (SKA) will be a giant radio interferometer of unprecedented sensitivity and dynamic range. Building is to start in 2018 and first light is to be seen in 2020. Its main stations will be placed in South Africa and Australia and there will be smaller remote stations in other African and Oceanian countries (see Fig. 1.7).

The SKA will consist of several thousand antennae and dishes with a total collecting area of 1 million square meters. This will make it fifty times more sensitive than the currently most powerful radio interferometer, the VLA, which consists of 27 antennae with a total collecting area of about 14000 m². Compared to the Arecibo radio telescope (~ 70000 m²) it will be 10-100 times more sensitive. It will continuously cover frequencies from 70 MHz ($\lambda = 4.3$ m) to 14 GHz ($\lambda = 2.1$ cm) thus covering most of the frequencies observable from the ground (see the SKA white paper, Aharonian et al. 2013, the exact numbers might change during the design process). The antennae and dishes will be distributed in a way to maximize the spatial dynamic range, i.e. to probe both, very large scales and very small scales. Many of the dishes will eventually be equipped with phased array feeds to enhance their field of view.

The SKA will be capable of Very Long Baseline interferometry while simultaneously surpassing the sensitivity of even the largest single dish telescopes. It will be built in two phases. Phase 1 will already enable fundamental advances in a number of science cases. In phase 1 the SKA should be able to detect neutral hydrogen at redshifts up to 20. This



Figure 1.7: Overview of the SKA configuration. Central stations consist of a core of 1 km and an inner region of 5 km in size. An intermediate regime extending out to 180 km consists of antennae distributed along 5 spiral arms and the outmost stations will be distributed along 3 spiral arms extending out to 3000 km. Image taken from Aharonian et al. (2013), image produced by SPDO/Swinburne Astronomy.

will enable the SKA to trace out the epoch of reionization, which is assumed to have taken place from around redshift 20 to redshift 6. A significant amount of pulsars is expected to be detected already in phase 1, but phase 2 promises to find about 30 000 normal and 3000 millisecond pulsars in the Milky Way. On the order of 10 000 parallaxes of pulsars will be measured as well (Smits et al. 2011). This will be a huge increase compared to the ~ 2500 pulsars (with ~ 100 parallaxes) known today and thus a major leap forward in the mapping of the free electron density of the Milky Way (see section 1.1.2).

1.3 Signal inference and imaging

The overall goal of any signal inference method is to find an estimate of a quantity s which is connected to a known data set d via some measurement process. The imaging algorithms derived and used in this thesis are based on information field theory, a formalism for Bayesian inference of continuous quantities. The general strategy is to formulate the measurement process (the forward model $s \rightarrow d$) as a probability distribution, the likelihood. This forward model can typically not be traced backwards, since the data rarely constrain all degrees of freedom of s and consequently there are (infinitely) many possible s that could have caused d . In Bayesian inference this problem is tackled by combining the likelihood with another probability distribution, the prior. It incorporates a priori knowledge (or assumptions) that serves to discriminate between the (from the point of view of the likelihood) equally likely signal configurations. A suitable prior breaks the degeneracy of the problem and the combination of prior and likelihood forms a well-defined posterior distribution for the signal. For a detailed introduction into the topic we recommend

$$\begin{array}{ccc}
 \mathcal{P}([p_1, \dots]) & & \mathcal{P}([q_1, \dots]) \\
 \swarrow & & \swarrow \\
 \mathcal{P}(s|[p_1, \dots]) & & \mathcal{P}(d|s, [q_1, \dots]) \\
 \searrow & & \swarrow \\
 \mathcal{P}(s, d, [p_1, \dots], [q_1, \dots]) & &
 \end{array}$$

$$\mathcal{P}(s, d, [q_1, \dots]) = \int \mathcal{D}[p_1, \dots] \mathcal{P}(s, d, [p_1, \dots], [q_1, \dots])$$

$$s, [q_1, \dots] = \arg \max_{(s, [q_1, \dots])} \mathcal{P}(s, d, [q_1, \dots])$$

Figure 1.8: The solution strategy employed in this thesis. The prior $\mathcal{P}(s)$ as well as the likelihood $\mathcal{P}(d|s)$ depend on a number of hyperparameters, $[p_1, \dots]$ and $[q_1, \dots]$ for which hyperpriors are formulated. All probability distributions are combined into one joint distribution. The hyperparameters are marginalized out analytically if possible. The resulting distribution is maximized with respect to the signal as well as the hyperparameters that could not be marginalized out analytically.

Jaynes & Bretthorst (2003). In this work, the posterior distribution is maximized to find the most likely signal configuration while all other (unknown) quantities are marginalized over or maximized for as well. We illustrate the strategy in Fig. 1.8. The involved solution strategies are called the maximum a posteriori approach and the empirical Bayes method. The exact methodology and prerequisites will be presented in the individual chapters of this thesis, where the imaging algorithms are derived.

This is by far not the only approach to signal inference. Within the context of Bayesian inference there are many other approaches to evaluate a posterior distribution. There are also numerous methods that do not formulate a prior, but operate solely on the likelihood and then there are methods that are not formulated in a probabilistic setting at all. The goal of this section is to give an overview of the signal inference methods currently in use.

1.3.1 Frequentist and Bayesian methods

Frequentist methods are approaches guided solely by the likelihood. The main argument is that the prior in Bayesian inference is subjective. It is formulated by the scientist who analyzes the data in a way that suits its purpose and there are infinitely many priors one can formulate which ultimately influence the result of the analysis. In the Bayesian way of reasoning this is not seen as a problem. All probabilities are subjective, but as long as they are formulated with reason, strong evidence through data will always prevail.

Frequentist methods often aim to construct an unbiased estimator. Given the same signal but different realizations of the data, the result of an unbiased estimator should be symmetrically distributed around the true signal. In a Bayesian method this spread

would be typically shifted or skewed into the direction in which the values of the signal are considered to be more likely a priori. From the Bayesian point of view the “bias” is the incorporation of a priori knowledge. A good prior improves the performance of the estimator on average (averaged over all possible signal and data realizations). An estimator that is unbiased for a fixed signal can introduce a systematic bias when applied to many (different) signals.

An example for this can be seen in the Lutz-Kelker bias for parallax measures. Imagine the parallax of an object to be measured as 1 arcsec with an uncertainty of 0.1 arcsec, for example. Since the volume of the spherical shell bounded by the parallaxes 0.9 arcsec and 1 arcsec is larger than of the shell bounded 1 arcsec and 1.1 arcsec it is likely that there are more objects in the outer shell⁶. This leads to more cases of the measured parallax of 1 arcsec originating from a true parallax which is smaller. An “unbiased” estimator would therefore yield parallaxes which are too large on average if applied to a large data set of parallaxes. A Bayesian estimator using a prior that incorporates the volume effect would correct that.

Besides the dispute over the advantages of biased and unbiased estimators frequentist methods often encounter the problem of regularization. If the likelihood does not constrain all degrees of freedom of the signal, the forward model cannot be inverted. The problem is ill-posed. In those cases frequentist methods have to be adapted to include some sort of regularization, which can often be mapped to a prior in the Bayesian sense.

The biggest difference between Bayesian and frequentist methods is however in their interpretation. Bayesian methods construct a posterior distribution, a probability distribution with respect to the signal. Therefore the derived answers can be interpreted as statements about the signal. Frequentist methods operate on the probability distribution of the data. In turn all derived answers have to be understood as statements about the data. For example the answer of a Bayesian method might be “ $s = 5 \pm 1$ ” = “the signal has an expected value of 5 with a credibility interval of 1”. This means that given all input knowledge and data the signal is believed to lie in this interval with a probability of 68%⁷. A frequentist method would yield an answer like “ 5 ± 1 ” = “the best fitting value for the signal is 5, the confidence interval is 1”. This means that if the measurement were to be repeated and the estimate calculated using the exact same procedure, the confidence interval would enclose the true value 68% of the time. The fundamental difference is that frequentist confidence intervals are always statements about repeated measurements whereas Bayesian credibility intervals are statements about the state of information.

1.3.2 Estimators from quadratic cost functions

All estimators aim to minimize some sort of cost function. The most common cost function in science is a quadratic cost function. Estimators based on such a cost function aim to minimize the L_2 -distance between the signal response $R(s)$ and the data d . From the

⁶Note that the distance is inversely proportional to the parallax.

⁷One can of course also specify a different interval, corresponding to a probability of 90%, for example.

probabilistic perspective this corresponds to assuming the data are subject to additive Gaussian⁸ noise n ,

$$d = R(s) + n, \quad n \leftarrow \mathcal{G}(n, N). \quad (1.14)$$

If the signal response $R(s)$ is linear in s , the resulting estimator is linear in d as well.

Least squares estimator

The least squares estimator is an estimator for overdetermined signals, i.e. the data contain more degrees of freedom than the signal. The estimator finds the signal configuration that minimizes the quadratic error between the modeled data and the measured data. Often it is not the case that the signal is overdetermined. To reduce the degrees of freedom the signal is modeled as a (typically linear) combination of basis functions or templates. The least squares estimator then finds the best weights w to these templates,

$$s = s[w], \quad w_{\text{best}} = \arg \min_w \sum_j (d_j - R(s[w]))_j^2 \quad (1.15)$$

The choice and amount of templates has of course great influence on the result. It can be seen as the incorporation of prior knowledge, but not with a probability distribution.

χ^2 estimator

The quadratic error with respect to the noise covariance is often called χ^2 ,

$$\chi^2 = (d - R(s))^\dagger N^{-1} (d - R(s)), \quad (1.16)$$

where \dagger denotes transposition and complex conjugation. For Gaussian noise n it follows the χ^2 distribution. The χ^2 estimator is a variant of the least squares estimator, which minimizes χ^2 instead of the quadratic error.

Regularized linear inversion

If the forward model can be expressed by a linear relationship between signal and data an estimate m of the signal can be found by matrix inversion.

$$d = Rs + n, \quad m = (R^\dagger R)^{-1} R^\dagger d, \quad (1.17)$$

However, the matrix $R^\dagger R$ is typically singular⁹ and cannot be inverted directly. Therefore one typically includes a regularization in the form of

$$m = (R^\dagger R + \epsilon \mathbb{1})^{-1} R^\dagger d, \quad \text{or} \quad m = (R^\dagger R + \Gamma^\dagger \Gamma)^{-1} R^\dagger d. \quad (1.18)$$

⁸The Gaussian distribution is defined as $G(a, A) := |2\pi A|^{-1/2} e^{-\frac{1}{2} a^\dagger A^{-1} a}$, where the \dagger denotes transposition and complex conjugation.

⁹It is not singular if the data constrain all degrees of freedom of the signal.

The latter case is called Tikhonov regularization. Γ is to express expected (or desired) features of the signal. The result corresponds to the a posteriori mean if the noise covariance is the identity and when a Gaussian prior with zero mean and covariance $S = (\Gamma^\dagger \Gamma)^{-1}$ is used,

$$P(s) = \mathcal{G}(s, S). \quad (1.19)$$

Thus linear inversion with Tikhonov regularization can be seen as a special case of the generalized Wiener filter (see e.g. Enßlin et al. 2009),

$$m = (R^\dagger N^{-1} R + S^{-1})^{-1} R^\dagger N^{-1} d. \quad (1.20)$$

1.3.3 Other estimators

Maximum likelihood estimators

Maximum likelihood estimators are the generalization of χ^2 estimator to arbitrary likelihoods. The objective function to minimize is the negative log-likelihood,

$$m = \arg \min_s [-\ln \mathcal{P}(d|s)], \quad (1.21)$$

which results in the χ^2 estimator if the likelihood is Gaussian. As for the quadratic estimator this may require additional regularization if not all of the degrees of freedom of the signal are constrained by the likelihood. An additive regularization in the form of

$$m = \arg \min_s [-\ln \mathcal{P}(d|s) + f(s)] \quad (1.22)$$

can be mapped to a prior $P(s) \propto \exp(-f(s))$. The maximum likelihood estimator can be seen as a maximum a posteriori estimator using a constant (hence uninformative) prior.

Least absolute deviations

The least absolute deviations is the analogon to the least squared estimator with the squares replaced by absolutes,

$$m_w = \arg \min_w \sum_j |d_j - R(s)_j|. \quad (1.23)$$

It can be seen as a maximum likelihood estimator where the likelihood is a Laplace distribution.

1.3.4 Evaluating posterior distributions

Maximum a posteriori

In this work we evaluate the posterior distribution using the maximum a posteriori Ansatz. In practice this means that the negative logarithmic joint probability is minimized with respect to the signal,

$$m = \arg \min_s [-\ln P(s, d)]. \quad (1.24)$$

This minimization can be done in several ways. There are gradient descent methods which follow the negative gradient to the minimum and there are Newton methods which use the second derivative in addition to the gradient to minimize the potential. In practice one often uses quasi-Newton methods which correct the negative gradient using an estimate of the curvature. The advantage of the maximum a posteriori Ansatz is that it is relatively fast. The disadvantages are that it can get stuck in a local minimum and that the most likely value is not necessarily the best estimate. Furthermore, it depends on the parametrization of the problem. Typically, the best¹⁰ estimate would be the posterior mean. However, unless the posterior mean coincides with the maximum a posteriori it typically has to be evaluated using relatively expensive sampling techniques.

Markov Chain Monte Carlo

In Monte Carlo sampling methods random realizations of the posterior distribution are generated to calculate moments such as the mean and the variance of the signal. If the posterior is of a form where samples can be generated directly, the procedure is very quick. However, this is rarely the case. Typically samples are generated by employing a random walk in the phase space of the signal. If the random walk fulfills detailed balance,

$$T(x^{(n+1)}|x^{(n)}) \mathcal{P}(x^{(n)}) = T(x^{(n)}|x^{(n+1)}) \mathcal{P}(x^{(n+1)}), \quad (1.25)$$

where $T(x^{(n+1)}|x^{(n)})$ is the transition probability from $x^{(n)}$ to $x^{(n+1)}$, the statistics of the chain follow the statistics of the distribution. Methods using such random walks are called Markov Chain Monte Carlo (MCMC) methods. The most famous example of a MCMC method is the Metropolis-Hastings algorithm (Metropolis et al. 1953; Hastings 1970). In the Metropolis-Hastings algorithm new samples are generated in a proposal step and are accepted or rejected randomly with probabilities that fulfill detailed balance. This involves careful tuning of the proposal mechanism in order to have the chain move fast enough to decorrelate in acceptable time without making the rejection rate too high. An Ansatz which increases the decorrelation speed without hurting the acceptance rate is the Hamiltonian Monte Carlo method. It introduces a pseudo-momentum and follows the Newtonian equations of movement to keep the acceptance rate close to 1. Another approach is to simulate a stochastic differential equation with a stationary solution that corresponds to the probability distribution one wants to evaluate. This is called Langevin sampling. For a detailed summary of the developments we refer the reader to Neal (2012) and the references therein.

Additional sampling methods which are applicable to evaluate posterior distributions are nested sampling (Skilling 2004), which has the advantage of calculating the normalization of the posterior while generating the samples, and Gibbs sampling (see George Casella 1992, and references therein), in which the probability distribution is broken down into conditional subdistributions from which samples can be generated directly.

¹⁰The posterior mean is the best estimate if one wants to minimize the squared error between real signal and estimate. Other cost functions are minimized by other estimates. The absolute error, for example, is minimized by the median.

1.4 Outline of this thesis

In chapter 2 we derive and test a tomographic algorithm that produces a three-dimensional map of the free electron density in the Milky Way from pulsar dispersion measures and distances. The main difference between our approach and previous work is the fact that we do not model the free electron density as a combination of templates or functions of which the parameters need to be found. Our modeling treats the density as a field retaining all of its degrees of freedom. This means that any feature or structure can arise in our map as long as the data warrant it sufficiently. Since the current abundance of appropriate data is too low, we investigate what level of detail will be reconstructible using simulated pulsar populations and observations. The generated data sets correspond to the predicted observations of the Square Kilometre Array, the next leap forward in radio interferometry.

Pulsar parallaxes are typically measured with radio interferometers, since they can provide the necessary resolution at the appropriate frequencies. The downside of the interferometric setup is that the emission on the sky is only sparsely measured in its harmonic space. This makes the imaging difficult, especially for diffuse emission. Consequently we develop and test fastRESOLVE, a Bayesian imaging algorithm for interferometric data, in chapter 3. The goal is to reduce the computing time by several orders of magnitude compared to the predecessor RESOLVE by avoiding expensive gridding routines. Furthermore, we include a Bayesian estimate of the likelihood variances to make the algorithm more robust against the bad surprises often found in real data. The algorithm is tested on real VLA observations of Abell 2199, a galaxy cluster which contains the powerful active galactic nucleus 3C 338. This nucleus is surrounded by extended jets and lobes visible in radio and provides an ideal test case for our algorithm against the standard method CLEAN.

Both of the derived algorithms reconstruct positive definite quantities, the electron density and the intensity of emission. In consequence the priors in both algorithms are so-called log-normal priors, which is to say that the logarithmic quantities follow a Gaussian distribution. Since the prior correlation structures are described by power spectra the need for a conversion formula between the power spectrum of a density and the power spectrum of the logarithmic density arises. In chapter 4 we derive such a formula from a statistical Ansatz. We show that the formula is also useful to rejuvenate the matter power spectrum. The matter power spectrum grows linearly in time at redshifts larger than ~ 10 until it picks up non-linear features at later times. Using our conversion formalism we investigate whether these nonlinearities arise naturally if the logarithmic density power spectrum is evolved instead. Previously this has been claimed on the basis of numerical simulations.

In chapter 5 we provide an outlook on how the work in this thesis could be continued and improved.

Chapter 2

Tomography of the Galactic free electron density with the Square Kilometer Array

Note: This chapter, as well as Appendix A, have been published in Astronomy & Astrophysics (Greiner et al. 2016a).

Abstract

We present a new algorithm for reconstructing the Galactic free electron density from pulsar dispersion measures. The algorithm performs a nonparametric tomography for a density field with an arbitrary amount of degrees of freedom. It is based on approximating the Galactic free electron density as the product of a profile function with a statistically isotropic and homogeneous log-normal field. Under this approximation the algorithm generates a map of the free electron density as well as an uncertainty estimate without the need of information about the power spectrum. The uncertainties of the pulsar distances are treated consistently by an iterative procedure. We tested the algorithm using the NE2001 model with modified fluctuations as a Galaxy model, pulsar populations generated from the Lorimer population model, and mock observations emulating the upcoming Square Kilometer Array (SKA). We show the quality of the reconstruction for mock data sets containing between 1000 and 10000 pulsars with distance uncertainties of up to 25%. Our results show that with the SKA nonparametric tomography of the Galactic free electron density becomes feasible, but the quality of the reconstruction is very sensitive to the distance uncertainties.

2.1 Introduction

Regions of the interstellar medium that are (partly) ionized play an important role in a number of effects such as pulse dispersion and scattering, and Faraday rotation. Addition-

ally, ionized parts of the interstellar medium emit radiation through free-free emission and H_α emission. The magnitude of these effects depends on the distribution of free electrons, the free electron density. It is therefore of great interest to model or reconstruct the free electron density as accurately as possible.

Reconstruction and modeling of the Milky Way has been an ongoing topic of research for many years. The free electron density has been modeled by Taylor & Cordes (1993), Cordes & Lazio (2002), and Gaensler et al. (2008), among others. For a comparison and discussion of various existing models see Schnitzeler (2012) and for a review of the mapping of HI regions see Kalberla & Kerp (2009). The interstellar magnetic field has been modeled by Sun et al. (2008), Sun & Reich (2010) and Jansson & Farrar (2012b,a). The dust distribution has been modeled for example by Berry et al. (2012), and even nonparametric tomography has been performed by Lallement et al. (2014) and Sale & Magorrian (2014).

We plan to use the dispersion measures (DM) of pulsar signals together with accurate pulsar distances to map the distribution of ionized gas in the Milky Way. The dispersion measure is defined as the line-of-sight integral over the free electron density between the observer and the pulsar,

$$DM = \int_{\text{pulsar}}^{\text{observer}} dr n_e, \quad (2.1)$$

where n_e is the three-dimensional free electron density. DM can be estimated by measuring the arrival time of a pulse at different frequencies, since the time delay is proportional to DM/ν^2 . While there is a vast number of known dispersion measures, very few of them are complemented by an independent distance estimate. The NE2001 model by Cordes & Lazio (2002) is currently the most popular model for the free electron density of the Milky Way. It uses 1143 DM measurements, 112 of which were complemented by distance estimates of varying quality. Additionally, it uses 269 pulsar scattering measurements, which only provide very indirect distance constraints.

We here perform nonparametric tomography of a simulation of the Galactic free electron density from pulsar dispersion measures complemented by independent distance estimates. By nonparametric tomography we mean a reconstruction with a virtually infinite¹ number of degrees of freedom using a close to minimal set of prior assumptions that only resolves structures that are supported by the data. Our assumptions are that the electron density is positive and spatially correlated and that the large-scale electron distribution only shows a variation with distance from the Galactic center and height above the Galactic plane. Both the correlation structure and the scaling behavior have to be inferred from the data. As a consequence, our reconstruction is focused on the large (kpc) scales of the Galactic free electron density. Small-scale structures such as HII regions and supernova remnants as well as spiral arms are only recovered if they are sufficiently deeply probed and constrained by the data.

Our tomography algorithm is derived from first principles in a Bayesian setting. This

¹ In numerical practice, the amount of degrees of freedom is the number of pixels used. However, the reconstruction will be resolution independent once the resolution is high enough.

has the advantage that all assumptions are clearly stated as priors. Additionally, it allows us to provide uncertainty maps of our reconstructions, which are important for any subsequent scientific analysis.

To obtain a meaningful map with minimal assumptions, a data set of high quality is needed, of course. Currently, there are around 100 pulsars known with reliable (independent) distance estimates. This only allows for a nonparametric reconstruction of the largest features in the Milky Way. New measurements with the Very Long Baseline Array will soon double the number of pulsars with accurate distances (see Deller et al. (2011)). However, with the planned Square Kilometer Array radio interferometer (SKA) the number of pulsars with parallax distance estimates might increase to around 10000 (see Smits et al. (2011)). In this paper we therefore investigate the feasibility of nonparametric tomography of the free electron density and demonstrate the performance of our algorithm by applying it to mock data sets similar to what the SKA might deliver. To this end, we create four Galaxy models from the NE2001 code by Cordes & Lazio (2002) with varying degrees of fluctuations and contrast as well as observational mock data sets for up to 10000 pulsars with distance estimates of varying quality and apply our algorithm to these data sets.

The remainder of this paper is structured as follows: First, we derive our tomography algorithm in Sect. 2.2, explaining our notation, our underlying assumptions, and all probability density functions involved. Second, we explain our Galaxy models and mock observations in detail in Sect. 2.3. In Sect. 2.4 we compare the electron density distributions reconstructed from mock observations with those from the Galaxy models used to produce them. We summarize our discussion in Sect. 2.5.

2.2 Reconstruction algorithm

The reconstruction algorithm applied in this work was derived within the framework of information field theory introduced by Enßlin et al. (2009). We also follow for most parts the notation used by them. To reconstruct the Galactic free electron density from pulsar dispersion measurements, we used a filter formalism very similar to the one presented by Junklewitz et al. (2016), which in turn is based on the critical filter formalism developed by Enßlin & Weig (2010), Enßlin & Frommert (2011), and refined by Oppermann et al. (2013).

2.2.1 Signal model

In the inference formalism we aim to reconstruct the free electron density field ρ , a three-dimensional scalar field. We assume it is related to the observed dispersion measure data DM by a linear measurement equation subject to additive and signal-independent measurement noise,

$$DM = R\rho + n, \tag{2.2}$$

where n is the measurement noise and $R\rho$ is the application of the linear response operator R on the field ρ ,

$$(R\rho)_i \equiv \int d^3x R(i, \vec{x}) \rho(\vec{x}). \quad (2.3)$$

The response operator R describes line-of-sight integrals through the density. It can be defined as

$$R(i, \vec{x}) = \int_0^{|\vec{d}_i|} dr \delta(\vec{x} - r\hat{\mathbf{d}}_i), \quad (2.4)$$

where \vec{d}_i is the position of pulsar i in a coordinate system centered on the Sun, and $\delta(\cdot)$ is the three-dimensional Dirac delta-distribution and $\hat{\mathbf{d}}_i := \vec{d}_i/|\vec{d}_i|$.

Formally, the free electron density is a continuous field. In practice, we reconstructed a discretized version of this field, for instance, a three-dimensional map with some pixel size. The discretized density field can be thought of as a vector of dimension N_{pix} with each component containing the field value in a specific pixel. The dispersion data DM and the noise n can be regarded as vectors of dimension N_{data} , where each component of DM contains a specific measurement result and the corresponding component of n the noise contribution to it. Thus, the response operator becomes a matrix with N_{pix} columns and N_{data} rows.

We parametrized the density as

$$\rho(\vec{x}) = \Delta(\vec{x})\tilde{\rho}(\vec{x}), \quad (2.5)$$

where Δ is the Galactic profile field that describes the the disk shape of the Milky Way. All deviations from the Galactic profile are described by $\tilde{\rho}$, for which we assumed no distinguished direction or position *a priori*. To ensure positivity of the density, these fields were in turn parametrized as

$$\begin{aligned} \Delta(x, y, z) &= \exp\left(\alpha\left(\sqrt{x^2 + y^2}\right) + \beta(|z|)\right), \\ \tilde{\rho}(x, y, z) &= \exp(s(x, y, z)). \end{aligned} \quad (2.6)$$

Thus, Δ can only represent the vertical and radial scaling behavior of the density and has the degrees of freedom of two one-dimensional functions. On the other hand, $\tilde{\rho}$ retains all degrees of freedom of a three-dimensional field and can represent arbitrary structures. Neither Δ nor $\tilde{\rho}$ are known *a priori* and will be inferred from the data.

We summarize our modeling in Fig. 2.1. The logarithmic density ρ is parametrized by three additive components, one three-dimensional field and two one-dimensional fields. As we outline in Sects. 2.2.2 and 2.2.3, all three fields are assumed to follow Gaussian statistics *a priori*. For the 1D fields a specific correlation structure was assumed, while the correlation structure of the three-dimensional field was unknown, but assumed to be homogeneous and isotropic. Therefore, our modeling prefers smooth structures, fluctuations that scale with the density, and exponential scaling in radial and vertical directions. Of course, this is

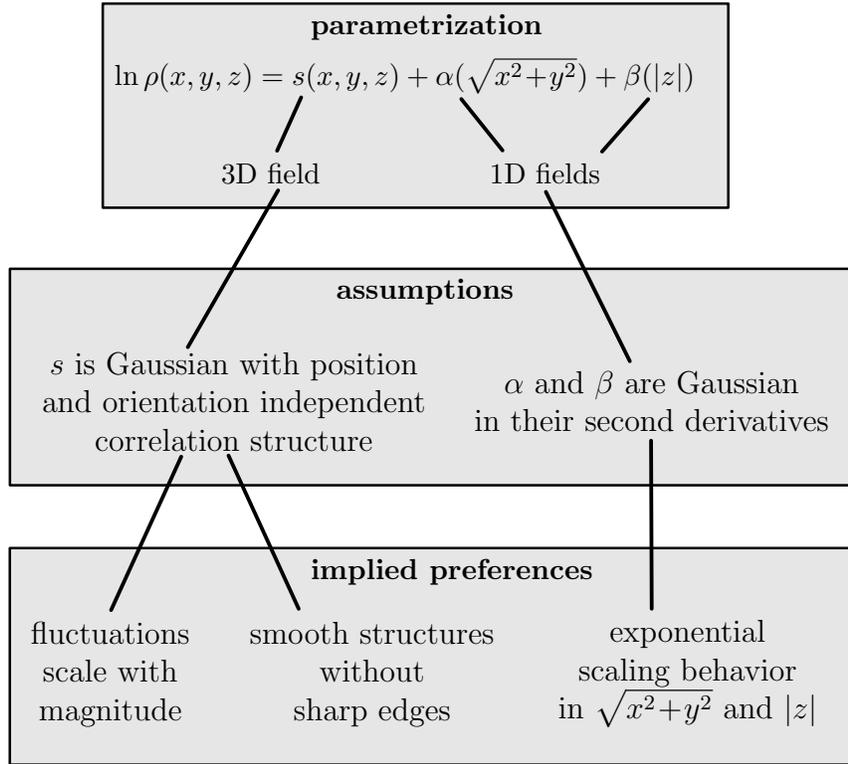


Figure 2.1: Diagram outlining the structure of our modeling.

a strong simplification of the Galaxy, where the behavior of the fluctuations can depend on the phase of the interstellar medium or the position within the Galaxy, for instance. However, all of these properties can be recovered if the data demand it, since all degrees of freedom are retained. They are just not part of the prior knowledge entering our inference.

2.2.2 Necessary probability density functions

Our goal is to derive an algorithm that yields an estimate of the logarithm of the Galactic free electron density. Hence, we constructed the posterior probability density function (PDF) $\mathcal{P}(s|\text{data})$, which is the PDF for the signal given the data set $\{DM, \vec{d}_{\text{obs(erved)}}\}$, using Bayes' theorem,

$$\mathcal{P}(s|\text{data}) = \frac{\mathcal{P}(s, DM|\vec{d}_{\text{obs}})}{\mathcal{P}(DM|\vec{d}_{\text{obs}})} = \frac{\mathcal{P}(s|\vec{d}_{\text{obs}})\mathcal{P}(DM|s, \vec{d}_{\text{obs}})}{\mathcal{P}(DM|\vec{d}_{\text{obs}})}. \quad (2.7)$$

On the right-hand side, we have three PDFs: the prior $\mathcal{P}(s|\vec{d}_{\text{obs}}) = \mathcal{P}(s)$, the likelihood $\mathcal{P}(DM|s, \vec{d}_{\text{obs}})$, and the evidence $\mathcal{P}(DM|\vec{d}_{\text{obs}})$. The evidence is independent of the signal and therefore automatically determined by the normalization of the posterior. The prior and the likelihood are addressed in the following sections. For notational convenience we

drop the dependence on the observed pulsar positions \vec{d}_{obs} throughout the rest of this paper.

Throughout this section we assume the Galactic profile field to be given. We address its inference in Sect. 2.2.3.

Likelihood

The likelihood $\mathcal{P}(DM|s)$ is the PDF that an observation yields dispersion measures DM assuming a specific realization of the underlying signal field s . If both the noise n and the pulsar distances $d_i \equiv |\vec{d}_i|$ were known, the relation between the dispersion measure data and signal would be deterministic,

$$\mathcal{P}(DM|s, n, d) = \delta(DM - R\rho - n), \quad (2.8)$$

with $\rho(\vec{x}) = \Delta(\vec{x})e^{s(\vec{x})}$. We do not know the realization of the noise, nor do we aim to reconstruct it. It was assumed to follow Gaussian statistics with zero mean and known covariance structure²,

$$\langle n_i n_j \rangle_{\mathcal{P}(n)} = N_{ij} = \delta_{ij} \sigma_i^2, \quad (2.9)$$

where σ_i is the root mean square error of the observation i , and we assumed independent measurements. Distance information is usually given in the form of parallaxes from which distance estimates can be derived. As all observables, these are subject to uncertainties, which is why the information about the distances of the pulsars is described by a PDF³, $\mathcal{P}(d) \equiv \mathcal{P}(d|\text{parallaxes})$, which can be non-Gaussian. Since we are doing inference on s , we need the noise and distance⁴ marginalized likelihood

$$\mathcal{P}(DM|s) = \int \mathcal{D}n \mathcal{D}d \mathcal{P}(DM|s, n, d) \mathcal{P}(n) \mathcal{P}(d), \quad (2.10)$$

where we assumed n and d to be independent of s and each other. The symbols $\mathcal{D}n$ and $\mathcal{D}d$ denote integration over the full configuration space of n and d , that is, the space of all possible configurations ($\mathcal{D}n \equiv \prod_i dn_i$).

Integration over n in Eq. (2.10) is trivial and yields

$$\mathcal{P}(DM|s) = \int \mathcal{D}d \mathcal{G}(DM - R\rho, N) \mathcal{P}(d), \quad (2.11)$$

where \mathcal{G} indicates a Gaussian PDF, $\mathcal{G}(x, X) := |2\pi X|^{-\frac{1}{2}} e^{-\frac{1}{2}x^\dagger X^{-1}x}$. Integration over d , however, cannot be done analytically, but the marginalized likelihood can be approximated by a Gaussian characterized by its first two moments in DM . The first moment is

$$\langle DM \rangle_{\mathcal{P}(DM|s)} = \tilde{R}\rho, \quad (2.12)$$

²We denote expectation values with respect to the underlying PDF as $\langle f(x) \rangle_{\mathcal{P}(x)} := \int \mathcal{D}x f(x) \mathcal{P}(x)$.

³We assumed here that the distance PDF is correctly derived from the parallax PDF taking Lutz-Kelker bias into account (see Verbiest et al. (2010)).

⁴Technically, we also need to marginalize over the position on the sky (i.e., the direction of the line of sight). But since the angular error of the pulsar position is small compared to the error in distance, we can neglect it and treat the direction as an exact value.

with

$$\tilde{R}_i(\vec{x}) = \langle R_i(\vec{x}) \rangle_{\mathcal{P}(d)} = \int_0^\infty dr \delta(\vec{x} - r\hat{\mathbf{d}}_i) P[d_i > r], \quad (2.13)$$

where $P[d_i > r]$ is the probability that the pulsar distance d_i is larger than r . The second moment is

$$\langle DM DM^\dagger \rangle_{\mathcal{P}(DM|s)} = N + \left\langle (R\rho)(R\rho)^\dagger \right\rangle_{\mathcal{P}(d)}. \quad (2.14)$$

For non-diagonal elements the second term on the right-hand side decouples,

$$\begin{aligned} \left\langle (R\rho)_i (R\rho)_j \right\rangle_{\mathcal{P}(d)} &= \left\langle (R\rho)_i \right\rangle_{\mathcal{P}(d)} \left\langle (R\rho)_j \right\rangle_{\mathcal{P}(d)} \\ &= \left(\tilde{R}\rho \right)_i \left(\tilde{R}\rho \right)_j \quad \text{for } i \neq j. \end{aligned} \quad (2.15)$$

Diagonal elements yield

$$\begin{aligned} \left\langle (R\rho)_i (R\rho)_i \right\rangle_{\mathcal{P}(d)} &= \int_{\mathbb{R}^3} d^3x \int_{\mathbb{R}^3} d^3y \rho(\vec{x})\rho(\vec{y}) \times \\ &\quad \left\langle R_i(\vec{x}) R_i(\vec{y}) \right\rangle_{\mathcal{P}(d)}, \end{aligned} \quad (2.16)$$

with

$$\begin{aligned} \left\langle R_i(\vec{x}) R_i(\vec{y}) \right\rangle_{\mathcal{P}(d)} &= \int_0^\infty dr \int_0^\infty dr' \delta(\vec{x} - r\hat{\mathbf{d}}_i) \delta(\vec{y} - r'\hat{\mathbf{d}}_i) \times \\ &\quad P[d_i > \max(r, r')]. \end{aligned} \quad (2.17)$$

Using these first two moments, we can approximate⁵ the likelihood $\mathcal{P}(DM|s)$ by a Gaussian $\mathcal{G}(DM - \tilde{R}\rho, \tilde{N})$ with

$$\tilde{N}_{ii} = N_{ii} + \rho^\dagger F^{(i)} \rho, \quad (2.18)$$

where⁶

$$\begin{aligned} F^{(i)}(\vec{x}, \vec{y}) &:= \left\langle R_i(\vec{x}) R_i(\vec{y}) \right\rangle_{\mathcal{P}(d_i)} - \tilde{R}_i(\vec{x}) \tilde{R}_i(\vec{y}) \\ &= \int_0^\infty dr \int_0^\infty dr' \delta(\vec{x} - r\hat{\mathbf{d}}_i) \delta(\vec{y} - r'\hat{\mathbf{d}}_i) \times \\ &\quad P[d_i > \max(r, r')] P[d_i < \min(r, r')]. \end{aligned} \quad (2.19)$$

⁵This corresponds to characterizing the likelihood by its cumulants and setting all but the first two cumulants to zero.

⁶We abbreviate $\xi^\dagger \zeta := \int d^3x \xi^*(\vec{x}) \zeta(\vec{x})$ and $\Xi \xi := \int d^3y \Xi(\vec{x}, \vec{y}) \xi(\vec{y})$ for continuous quantities.

The noise covariance matrix of this effective likelihood is signal dependent, which increases the complexity of the reconstruction problem. Therefore, we approximated the density in Eq. (2.18) by its posterior mean,

$$\rho^\dagger F^{(i)} \rho = \text{tr}(\rho \rho^\dagger F^{(i)}) \approx \text{tr}\left(\langle \rho \rangle_{\mathcal{P}(\rho|DM)} \langle \rho \rangle_{\mathcal{P}(\rho|DM)}^\dagger F^{(i)}\right). \quad (2.20)$$

Since $\langle \rho \rangle_{\mathcal{P}(\rho|DM)}$ depends on \tilde{N} , this yields a set of equations that need to be solved self-consistently (see Sect. 2.2.4).

Priors

The signal field s is unknown a priori, but we assumed that it has some correlation structure. We describe this correlation structure by moments up to second order in s . The principle of maximum entropy therefore requires that our prior probability distribution has a Gaussian form,

$$\mathcal{P}(s|S) = \mathcal{G}(s, S) := |2\pi S|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} s^\dagger S^{-1} s\right), \quad (2.21)$$

with some known correlation structure,

$$S(\vec{x}, \vec{y}) = \langle s(\vec{x}) s(\vec{y}) \rangle_{\mathcal{P}(s)}. \quad (2.22)$$

The first moment of s is set to zero because it can be absorbed into $\Delta(\vec{x})$. This means that the a priori mean of s is contained in $\Delta(\vec{x})$.

A priori, our algorithm has no preferred direction or position for s . This reduces the number of degrees of freedom of the correlation structure S . It is fully described by a power spectrum $p(k)$,

$$S(\vec{x}, \vec{y}) = \sum_k S^{(k)}(\vec{x}, \vec{y}) p(k), \quad (2.23)$$

where $S^{(k)}$ is the projection operator onto the spectral band k , with its Fourier transform defined as

$$S^{(k)}(\vec{q}, \vec{q}') = (2\pi)^3 \delta(\vec{q} - \vec{q}') \mathbb{1}_k(|\vec{q}|), \quad (2.24)$$

with

$$\mathbb{1}_k(|\vec{q}|) = \begin{cases} 1 & \text{for } |\vec{q}| = k \\ 0 & \text{otherwise} \end{cases}. \quad (2.25)$$

The power spectrum $p(k)$, however, is still unknown. The prior for the power spectrum is constructed out of two parts: First, an inverse gamma distribution $\mathcal{I}(p(k); \alpha_k, q_k)$ for each k -bin (see Appendix A.1), which is a conjugate prior for a Gaussian PDF, second a Gaussian cost-function that punishes deviations from power-law spectra (see Oppermann et al. (2013)),

$$\mathcal{P}(p) \propto \left\{ \prod_k \mathcal{I}(p(k); \alpha_k, q_k) \right\} \exp\left(-\frac{1}{2} (\log p)^\dagger T (\log p)\right). \quad (2.26)$$

T is an operator that fulfills

$$(\log p)^\dagger T(\log p) = \frac{1}{\sigma_p^2} \int d(\log k) \left(\frac{\partial^2 \log p(k)}{\partial (\log k)^2} \right)^2, \quad (2.27)$$

and σ_p is a parameter that dictates how smooth the power spectrum is expected to be. In our paper \log refers to the natural logarithm. We explain our choice of the parameters α_k , q_k , and σ_p in Appendix A.1.

Power spectrum posterior

With the signal and power spectrum priors, we can derive a posterior for the power spectrum,

$$\begin{aligned} \mathcal{P}(p|DM) &\propto \int \mathcal{D}s \mathcal{P}(DM|s, p) \mathcal{P}(s|p) \mathcal{P}(p) \\ &= \int \mathcal{D}s \mathcal{P}(DM|s) \mathcal{G}(s, S) \mathcal{P}(p). \end{aligned} \quad (2.28)$$

We calculated the integral using a saddle point approximation up to second order around the maximum for the s -dependent part,

$$\begin{aligned} \mathcal{P}(DM|s) \mathcal{G}(s, S) &\approx \mathcal{P}(DM|m) \mathcal{G}(m, S) e^{-\frac{1}{2}(s-m)^\dagger D^{-1}(s-m)} \\ &\propto \mathcal{G}(m, S) e^{-\frac{1}{2}(s-m)^\dagger D^{-1}(s-m)}, \end{aligned} \quad (2.29)$$

where m and D are defined as $m^{(s)}$ and $D^{(s)}$ in Sect. 2.2.2 and only s - and p -dependent factors are kept after the proportionality sign. With this approximation we arrive at

$$\mathcal{P}(p|DM) \propto |2\pi D|^{\frac{1}{2}} |2\pi S|^{-\frac{1}{2}} e^{-\frac{1}{2}m^\dagger S^{-1}m} \mathcal{P}(p). \quad (2.30)$$

Maximizing this PDF with respect to $\log(p)$ (see Oppermann et al. (2013)) leads to

$$p(k) = \frac{q_k + \frac{1}{2} \text{tr}(S^{(k)}(mm^\dagger + D))}{\alpha_k - 1 + \frac{1}{2} \varrho_k + (T \log p)_k}, \quad (2.31)$$

where $\varrho_k = \text{tr}(S^{(k)})$ is the number of degrees of freedom in the spectral band k . This formula for the power spectrum $p(k)$ should be solved self-consistently because m and D depend on $p(k)$ as well. Thus we arrive at an iterative scheme, where we look for a fixed point of Eq. (2.31).

Signal posterior

The signal posterior can be expressed as

$$\mathcal{P}(s|DM) = \int \mathcal{D}(\log p) \mathcal{P}(\log p|DM) \mathcal{P}(s|p, DM), \quad (2.32)$$

where $\mathcal{P}(s|p, DM)$ is the signal posterior with a given power spectrum. Instead of calculating the marginalization over $\log p$, we used Eq. (2.31) for the power spectrum; that is, we approximated $\mathcal{P}(\log p|DM)$ by a Dirac peak at its maximum. This procedure is known as the empirical Bayes method. The signal posterior with a given power spectrum is proportional to the product of the signal prior and the likelihood (see Eq. (2.7)),

$$\begin{aligned} \mathcal{P}(s, DM|p, \tilde{N}) \propto \exp\left(-\frac{1}{2}s^\dagger S^{-1}s\right) \times \\ \exp\left(-\frac{1}{2}(DM - \tilde{R}\rho)^\dagger \tilde{N}^{-1}(DM - \tilde{R}\rho)\right). \end{aligned} \quad (2.33)$$

However, as has been demonstrated in Sects. 2.2.2 and 2.2.2, S and \tilde{N} depend on the mean and the covariance of $\mathcal{P}(s|DM, p, \tilde{N})$ leading to a circular dependence that needs to be solved self-consistently.

We approximate the mean of the posterior by minimizing the joint Hamiltonian $\mathcal{H}(s, DM|p, \tilde{N}) := -\log \mathcal{P}(s, DM|p, \tilde{N})$ with respect to s ,

$$m^{(s)} \approx \arg \min_s \mathcal{H}(s, DM|p, \tilde{N}), \quad (2.34)$$

and its covariance by the inverse Hessian at that minimum,

$$D^{(s)} \approx \left(\frac{\delta^2}{\delta s \delta s^\dagger} \mathcal{H}(s, DM|p, \tilde{N}) \Big|_{s=m} \right)^{-1}. \quad (2.35)$$

These estimates are the maximum a posteriori (MAP) estimates of s . Consequently, $m^{(\rho)}$ and $D^{(\rho)}$ are estimated as

$$m^{(\rho)}(\vec{x}) \approx \Delta(\vec{x}) e^{m^{(s)}(\vec{x})} \quad (2.36)$$

and

$$D^{(\rho)}(\vec{x}, \vec{y}) \approx \Delta(\vec{x}) e^{m^{(s)}(\vec{x})} \left(e^{D^{(s)}(\vec{x}, \vec{y})} - 1 \right) e^{m^{(s)}(\vec{y})} \Delta(\vec{y}). \quad (2.37)$$

S is then constructed as

$$S(\vec{x}, \vec{y}) = \sum_k S^{(k)}(\vec{x}, \vec{y}) p(k), \quad (2.38)$$

with $p(k)$ given by Eq. (2.31). \tilde{N} is constructed using Eqs. (2.18) and (2.20) as

$$(\tilde{N})_{ij} = (N)_{ij} + \delta_{ij} \operatorname{tr} \left(m^{(\rho)} m^{(\rho)\dagger} F^{(i)} \right), \quad (2.39)$$

where δ_{ij} is the Kronecker delta.

2.2.3 Galactic profile inference

To infer the Galactic profile field Δ , we introduce $\tilde{s} \equiv s + \log(\Delta) \equiv \log(\rho)$. The Galactic profile is to capture the most prominent symmetries of a disk galaxy, namely its rotational symmetry and the scaling behavior with radial distance from the Galactic center and vertical distance from the Galactic plane. Using Eq. (2.6), $\mu \equiv \log(\Delta)$ becomes

$$\mu(x, y, z) = \alpha(r) + \beta(|z|), \quad \text{with } r \equiv \sqrt{x^2 + y^2}, \quad (2.40)$$

where α and β are one-dimensional functions describing the average behavior with respect to the radial distance and the vertical distance from the Galactic center. Including the shift by μ from s to \tilde{s} yields the signal prior

$$\mathcal{P}(\tilde{s}|\alpha, \beta) = \mathcal{G}(\tilde{s} - \mu, S). \quad (2.41)$$

We do not wish to assume specific functions α and β , but to infer them. To this end we chose a Gaussian prior,

$$\mathcal{P}(\alpha, \beta) \propto \exp\left(-\frac{1}{2\sigma_\alpha^2} \left(\frac{\partial^2 \alpha}{\partial r^2}\right)^2 - \frac{1}{2\sigma_\beta^2} \left(\frac{\partial^2 \beta}{\partial |z|^2}\right)^2\right), \quad (2.42)$$

with the second derivative of α (or β , respectively) as the argument. This prior prefers linear functions for α and β and thus Galactic profile fields with an exponential fall-off (or rise). To simplify the notation, we defined $\xi(r, |z|) = (\alpha(r), \beta(|z|))^T$ and introduced the linear operators Ξ and X , where

$$X\xi = \alpha + \beta \equiv \mu \quad (2.43)$$

and

$$\xi^\dagger \Xi \xi = \frac{1}{\sigma_\alpha^2} \left(\frac{\partial^2 \alpha}{\partial r^2}\right)^2 + \frac{1}{\sigma_\beta^2} \left(\frac{\partial^2 \beta}{\partial |z|^2}\right)^2. \quad (2.44)$$

Now we can write the Hamiltonian of ξ given a specific electron density as

$$\begin{aligned} \mathcal{H}(\xi|\tilde{s}) &= \frac{1}{2} (\tilde{s} - X\xi)^\dagger S^{-1} (\tilde{s} - X\xi) + \frac{1}{2} \xi^\dagger \Xi \xi + \text{const.} \\ &= \frac{1}{2} \xi^\dagger (X^\dagger S^{-1} X + \Xi) \xi - \tilde{s}^\dagger S^{-1} X \xi + \text{const.} \\ &\equiv \frac{1}{2} \xi^\dagger D_{(\xi)}^{-1} \xi - j_{(\xi)}^\dagger \xi + \text{const.} \end{aligned} \quad (2.45)$$

with $D_{(\xi)} = (X^\dagger S^{-1} X + \Xi)^{-1}$ and $j_{(\xi)} = X^\dagger S^{-1} \tilde{s}$. Since this Hamiltonian is a quadratic form in ξ , the mean of the corresponding Gaussian PDF is

$$\langle \xi \rangle_{(\xi|\tilde{s})} = D_{(\xi)} j_{(\xi)}. \quad (2.46)$$

2.2.4 Filter equations

Using the posterior estimates presented in the previous section, we arrive at the following iterative scheme for reconstructing the density ρ :

1. Make an initial guess for the power spectrum (e.g., some power law) and the additive term in the noise covariance (e.g., simple relative error propagation).

2. With the current estimates for p and \tilde{N} , the Hamiltonian, $\mathcal{H}(s, DM|p, \tilde{N}) + \text{const.} \equiv \log \mathcal{P}(s, DM|p, \tilde{N})$ is

$$\begin{aligned} \mathcal{H}(s, DM|p, \tilde{N}) &= \frac{1}{2} s^\dagger \left(\sum_k S^{(k)} p_k^{-1} \right) s \\ &+ \frac{1}{2} (e^s * \Delta)^\dagger \tilde{R}^\dagger \tilde{N}^{-1} \tilde{R} (e^s * \Delta) \\ &- DM^\dagger \tilde{N}^{-1} \tilde{R} (e^s * \Delta), \end{aligned} \quad (2.47)$$

where the asterisk denotes point-wise multiplication in position space.

3. The MAP estimate of this Hamiltonian is calculated as

$$m^{(s)} = \arg \min_s \mathcal{H}(s, DM|p, \tilde{N}), \quad (2.48)$$

with the covariance estimate (see Appendix A.2)

$$D^{(s)} = \left(\frac{\delta^2}{\delta s \delta s^\dagger} \mathcal{H}(s, DM|p, \tilde{N}) \Big|_{s=m} \right)^{-1}. \quad (2.49)$$

4. The updated power spectrum is the solution (with respect to $p(k)$) of the equation

$$p(k) = \frac{q_k + \frac{1}{2} \text{tr}(S^{(k)}(mm^\dagger + D))}{\alpha_k - 1 + \frac{1}{2} \varrho_k + (T \log p)_k}. \quad (2.50)$$

5. The updated effective noise covariance is calculated as

$$(\tilde{N})_{ii} = (N)_{ii} + \text{tr}(m^{(\rho)} m^{(\rho)\dagger} F^{(i)}), \quad (2.51)$$

with

$$m^{(\rho)}(\vec{x}) = \Delta(\vec{x}) \exp(m^{(s)}(\vec{x})). \quad (2.52)$$

6. The updated Galactic profile field is

$$\begin{aligned} \Delta &= \exp(Xm^{(\xi)}) \quad \text{with} \\ m^{(\xi)} &= (X^\dagger S^{-1} X + \Xi)^{-1} X^\dagger S^{-1} \log(m^{(\rho)}) \end{aligned} \quad (2.53)$$

7. Repeat from step 2 until convergence is reached.

When the solution of this set of equations is converged, the estimate of the density ρ is

$$\rho(\vec{x}) \approx m^{(\rho)}(\vec{x}) \pm \sigma^{(\rho)}(\vec{x}), \quad (2.54)$$

where the confidence interval $\sigma^{(\rho)}$ is defined as

$$\sigma^{(\rho)}(\vec{x}) := \sqrt{m^{(\rho)}(\vec{x}) (e^{D^{(s)}(\vec{x}, \vec{x})} - 1) m^{(\rho)}(\vec{x})}. \quad (2.55)$$

2.3 Application to simulated data

To test the reconstruction of the Galactic free electron density distribution with the SKA, we generated mock data sets of pulsars with various distance uncertainties. We simulated pulsar populations using the PSRPOPpy package by Bates et al. (2014), which is based on the pulsar population model by Lorimer et al. (2006). The generated populations take the observational thresholds of the SKA into account (mid-frequency). These data sets sample modified versions of the NE2001 model by Cordes & Lazio (2002) through dispersion measures.

2.3.1 Galaxy model

We deactivated⁷ all local ISM components as well as all clumps and voids in the NE2001 model. We kept the clump in the Galactic center, since it is the only one at a distinguished position. We evaluated⁸ the resulting free electron density model in a 512x512x64 pixel grid centered on the Galactic center with a pixel edge length of 75 pc. This means that our model extends out to 2400 pc from the Galactic plane. We assumed a density of zero outside of this regime when calculating the dispersion measures. The resulting density field is very smooth. We generated three Gaussian random fields that follow a power-law distribution with a spectral index⁹ of -4.66, but have different fluctuation amplitudes. We verified that

⁷ This was achieved by modifying the nelism.inp, neclumpN.NE2001.dat, and nevoidN.NE2001.dat files provided with the NE2001 code.

⁸ To obtain the three-dimensional free electron density from the compiled NE2001 code, we evaluated two positions in each pixel that have parallel line-of-sight vectors. The difference between their dispersion measures divided by the difference of their distance to the Sun was then taken as the free electron density in that pixel.

⁹ There is no physical reason for this choice, but a power law with this index seems to follow the spectrum of the log-density in the original model NE2001 rather well on large to medium scales.

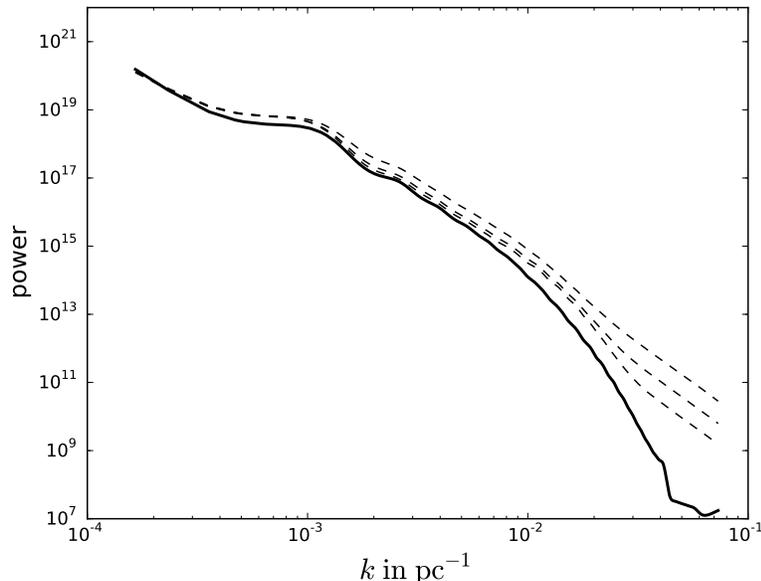


Figure 2.2: Power spectrum of the NE2001 field without local features, clumps, and voids compared to the three unenhanced Galaxy models. The thick solid line depicts the NE2001 spectrum, the thin dashed lines depict the spectra of the models with strong, medium, and weak fluctuations (from top to bottom). To calculate the spectra, the density peak in the Galactic center is masked.

the Sun is located in an underdensity in these random fields. Then we added these three random field maps to our smooth map of $\log(n_e)$ to create three different modified versions of NE2001. In Fig. 2.2 we depict the power spectra of the smooth NE2001 field (without local features, clumps, and voids) and the power spectra of its three modified versions.

Contrast-enhanced model

The three Galaxy models we generated from NE2001 have relatively little contrast in the sense that under- and overdense regions differ by relatively moderate factors. For example, the density in the region between the Perseus and Carina-Sagittarius arm where the Sun is located is only a factor of three lower than in the Perseus arm itself. Since the Perseus arm is a less than 1 kpc in width, any excess dispersion measure that is due to the arm can also be explained by an underestimated pulsar distance for many lines of sight. In consequence, we expect the reconstruction quality to improve if the input model has higher contrast. Therefore, we prepared one additional model with enhanced contrast. To this end, we took the input model with medium-strength fluctuations as described above. We divided out the scaling behavior in radial and vertical directions using the scale heights from NE2001. We squared the density and divided it by a constant to ensure that the mean density in

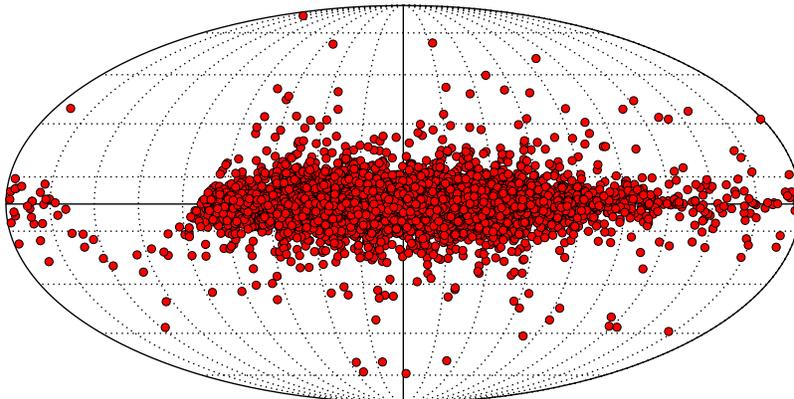


Figure 2.3: Positions of the simulated 10000 pulsars on the sky in Galactic coordinates.

the Galactic plane remains unchanged¹⁰. Finally we multiplied the resulting density with the scaling functions to restore the original scaling in radial and vertical directions.

This procedure yielded a Galaxy model sharing the same morphology and scaling behavior as the input model. Averaged over the lines of sight, the value of the dispersion measures is roughly unchanged. But the contrast is twice as strong, that is, the previously mentioned factor between the density in the Perseus arm and the inter-arm region is now squared from 3 to 9. We show a picture of the density in the Galactic plane of this model in Sect. 2.4.1, where we compare it with its reconstruction.

2.3.2 Simulated population and survey

We used the SKA template in the PSRPOPpy package, but reduced the maximum declination in equatorial coordinates to 50° (this is due to the SKAs position on the Southern Hemisphere, see e.g. Smits et al. (2009)). This yielded a detected population of roughly 14000 pulsars. Out of these, we took the first 1000, 5000, or 10000 pulsars as our test populations. The population was not ordered in any sense, which means that the first 1000 pulsars, for example, represent a random sample from the whole detected population. In reality, the Malmquist bias preferentially selects pulsars that lie close to the Sun. We chose a random selection, however, to see the effect of the population size on the quality of the reconstruction more clearly. In Fig. 2.3 we depict the population of 10000 pulsars projected onto the sky. The pulsars are concentrated toward the center of the Galaxy. The gap in the equatorial Northern Hemisphere is clearly evident in the left part of the plot.

2.3.3 Simulated dispersion measures and distances

We calculated the line integrals through the Galaxy models from the positions generated by the PSRPOPpy package to the Sun to generate simulated dispersion measures. We added

¹⁰ The bulge in the Galactic center is kept unchanged by the whole procedure.

Table 2.1: Types of data sets simulated for all Galaxy models. The columns indicate the number of pulsars, the rows the relative distance uncertainties.

	1000 pulsars	5000 pulsars	10000 pulsars
25% unc.	no	yes	yes
15% unc.	yes	yes	yes
5% unc.	yes	yes	no

Gaussian random variables to the pulsar distances to simulate measurement uncertainties of the distances. For each pulsar we generated one random number and scale this to 5%, 15%, or 25% of the distance of the pulsar. In reality the distance PDF would be non-Gaussian. The exact form depends on the combination of observables that are used to infer the distance. We used Gaussian PDFs for simplicity. As long as the real distance PDFs are unimodal, we do not expect this choice to have a significant effect on our study. We did not simulate additional measurement noise for the dispersion measures because it is expected to be small compared to the distance uncertainty. This leaves us with a number of data sets that are described in Table 2.1. In this table we omit the combinations of 1000 pulsars at 25% distance error (because we do not hope for a good reconstruction in that case) and 10000 pulsars at 5% distance error (because we deem it to be too unrealistic).

The aforementioned measurement scenarios were chosen to see the effect of the population size and the distance error on the reconstruction in isolation. A more realistic setting is of course a mix of distance uncertainties where more distant pulsars have larger distance errors on average. We therefore created one additional measurement scenario for 10000 pulsars, where we assigned the uncertainty magnitude of each pulsar randomly¹¹. The distance uncertainties are distributed as shown in Fig. 2.4. In this measurement set, 2969 pulsars have a 5% distance error, 3400 pulsars have a 15% distance error, and 3631 pulsars have a 25% distance error. Throughout the rest of this paper we refer to this data set as the mixed data set.

A very rough estimate of the scales that we can hope to resolve is given by the mean distance between neighboring pulsars and the average misplacement that is due to distance errors. The mean distance between neighboring pulsars is 490 pc for 1000, 290 pc for 5000, and 230 pc for 10000 pulsars. The average misplacement is 380 pc for 5%, 1100 pc for 15%, and 1900 pc for 25% distance errors and 1300 pc for the mixed data set. Interpreting these distances as independent uncertainties, we can combine them by adding the squares and taking the square root. This provides a rough estimate of sampling distances. In Table 2.2 we list these distances for each data set.

¹¹Each pulsar was assigned probabilities to belong to either the 5%, the 15%, or the 25% set. The probabilities depend on the pulsar distance, making more distant pulsars more likely to have higher uncertainties. The pulsar was then randomly assigned to an uncertainty set according to the probabilities.

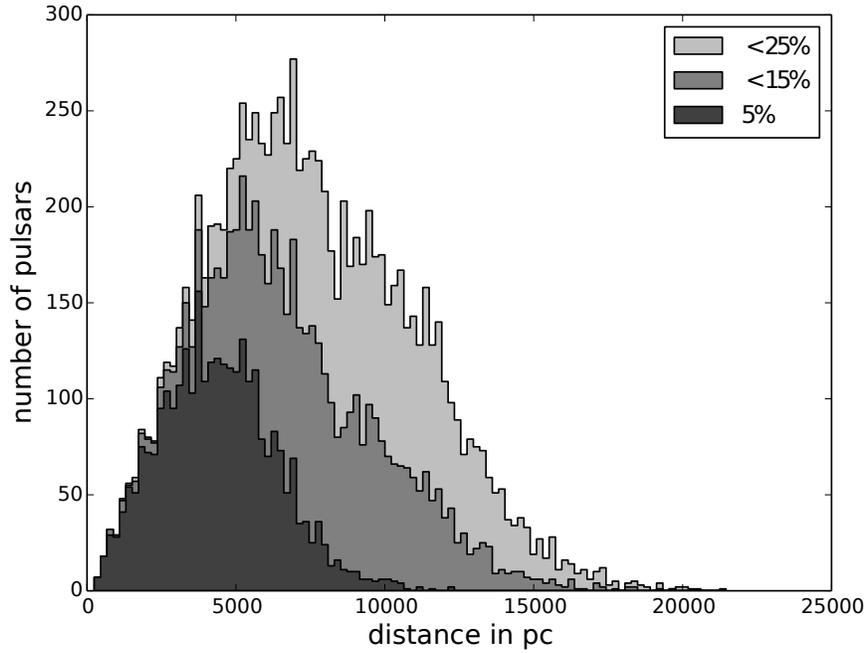


Figure 2.4: Histogram showing the distance uncertainty distribution with respect to the distance from the Sun in the mixed measurement set.

Table 2.2: Estimated sampling distances for each data set. The columns indicate the number of pulsars, the rows the relative distance uncertainties.

	1000 pulsars	5000 pulsars	10000 pulsars
25% unc.		1900 pc	1900 pc
15% unc.	1200 pc	1100 pc	1100 pc
5% unc.	600 pc	500 pc	

2.3.4 Algorithm setup

The algorithm was set up in a $128 \times 128 \times 48$ pixel grid centered on the Galactic center with pixel dimensions¹² of $281.25 \text{ pc} \times 281.25 \text{ pc} \times 250 \text{ pc}$. While the dispersion measures in our data sets were free from instrumental noise, the noise was assumed to be 2% in the algorithm. This provides a lower limit for the effective noise covariance (Eq. 2.18) and thus ensures stability of the inference without losing a significant amount of precision. The initial guess for the power spectrum is a broken power law with an exponent of -3.66 ¹³. For the propagated distance uncertainty it is

$$\sigma_i = \frac{\sqrt{\text{Var}[d_i]}}{d_i} DM_i, \quad (2.56)$$

where d_i is the distance of the pulsar (see Sect. 2.2.2). The initial guesses of the Galactic profile functions¹⁴ are

$$\alpha(r) = \frac{-r}{28000 \text{ pc}} \quad \text{and} \quad \beta(|z|) = \frac{-|z|}{1600 \text{ pc}}. \quad (2.57)$$

We discuss the convergence and final values of the power spectrum, effective errors, and profile functions in Appendix A.3.

2.4 Simulation evaluation

Our algorithm accounts for most of the variance in the data while regularizing the result to avoid overfitting. Most of the reconstructions shown in this section have corresponding reduced χ^2 values close to 1, indicating that they show all structures that are sufficiently well contained by the data. We discuss the reduced χ^2 values in detail in Appendix A.4.

2.4.1 Density in the midplane

The simulations show that with the amount of pulsars with reliable distance estimates that the SKA should deliver reconstruction of the free electron density in the vicinity of the Sun becomes feasible (see Fig. 2.5). However, small-scale features are difficult to identify in the reconstruction. Identifying spiral arms remains challenging as well, especially beyond the Galactic center. To resolve the spiral arms in the vicinity of the Sun, between 5000 and 10000 pulsars with distance accuracies between 5% and 15% are needed. As is

¹² We note that the pixels of our algorithm setup are significantly larger than those of the input models. This is on purpose, since in reality there will always be structure smaller than the chosen pixel size.

¹³ We might have used any power spectrum as an initial guess. It has negligible influence on the final result (see Appendix A.3). The index of -3.66 corresponds to Kolmogorov turbulence in three dimensions. Such a power spectrum produces spatially correlated structures while permitting fluctuations on all scales.

¹⁴ We note that while the priors for the profile functions prefer linear forms, all functional forms are allowed in principle.

evident from the figure, small distance uncertainties increase the quality of the reconstruction significantly. The reconstruction from 5000 pulsars with 5% distance uncertainty is better in quality than the one from 10000 pulsars with 15% distance uncertainty¹⁵. All reconstructions smooth out small-scale structure in the electron density, for example, at the Galactic center. If an overdensity appears at the wrong location, this indicates that the data do not constrain the overdensity well. For completeness we also show the recovered Galactic profile in the Galactic plane for 5000 pulsars with 5% distance uncertainty in Fig. 2.6. For other data sets the plot would look very similar. In Appendix A.5 we show a reconstruction where the Galactic profile and the correlation structure are known a priori and in Appendix A.6 we show and discuss the uncertainty estimate of the algorithm.

In Fig. 2.7 we compare the performance of the reconstruction algorithm for the three input model fluctuation strengths using 5000 pulsars with 5% and 15% distance uncertainty. The strength of the fluctuations clearly does not influence the quality of the reconstructions by much. The reconstructions of the models with stronger fluctuations exhibit stronger fluctuations as well, while all reconstructions omit or smear features to a similar degree. However, it clearly becomes more difficult to reconstruct the Perseus arm toward the Galactic anticenter if the fluctuations in the electron density are strong. This is to be expected because the spiral arm is also harder to recognize in the original model as the fluctuations become stronger.

In Fig. 2.8 we show the contrast-enhanced Galaxy model and its reconstruction using 5000 pulsars with distance uncertainties of 5%. As is clear from the figure, the algorithm is able to resolve much more detailed structure than the reconstruction of the unenhanced Galaxy model (bottom middle panel in Fig. 2.5). We stress that the pulsar population and their distance uncertainties are exactly the same for both cases. The increase in quality comes merely from the increased contrast and the resulting stronger imprint of under- and overdensities in the dispersion data. Therefore, we conclude that if the contrast of the real Galaxy is much stronger than in NE2001, our algorithm could resolve the Galaxy much better than the study on NE2001 indicates.

2.4.2 Vertical fall-off

A quantity of interest in any model of the Galactic free electron density is the drop-off of the average density with respect to distance from the Galactic plane. This behavior is shown in Fig. 2.9, which displays a vertical cut through the Galactic profile reconstructed using 5000 pulsars with 5% distance uncertainty.

In our parametrization the function β in Eq. (2.40) describes the average log-density at a certain distance from the Galactic plane. In Fig. 2.10 we show the estimates for β corresponding to the reconstructions shown in Fig. 2.5 along with their uncertainties (see Appendix A.7 for their calculation). The uncertainty regions reflect that a vertical fall-off can be explained by a global profile as well as by density fluctuations close to the Sun.

¹⁵ In principle, this behavior is expected because one measurement of a scalar quantity a with standard deviation σ contains the same amount of information as nine independent measurements with standard deviation 3σ (assuming Gaussian PDFs).

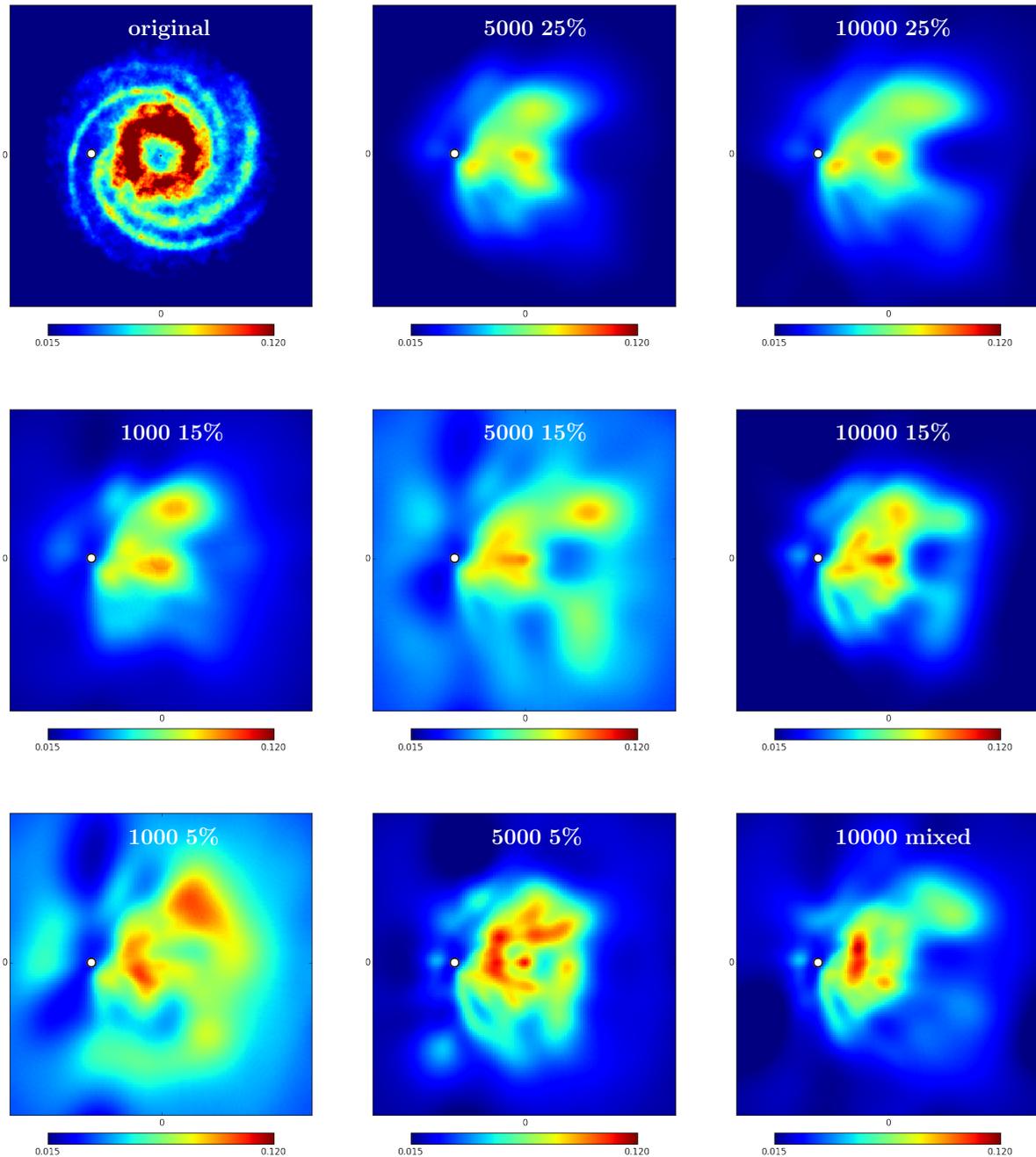


Figure 2.5: Several reconstructions of the Galaxy model with medium strength fluctuations. All panels show top-down views of the electron density in the Galactic plane using a linear color scale in units of cm^{-3} . The panels span 36000 pc in each dimension. The Sun is located at the white dot depicted in each panel. The rows show reconstructions with distance errors of 25%, 15%, and 5% (from top to bottom). The columns show reconstructions with 1000, 5000, and 10000 pulsars (from left to right). The layout follows Table 2.1. The top left panel shows the original input model (modified NE2001). The bottom right panel shows the reconstruction of the mixed measurement set.

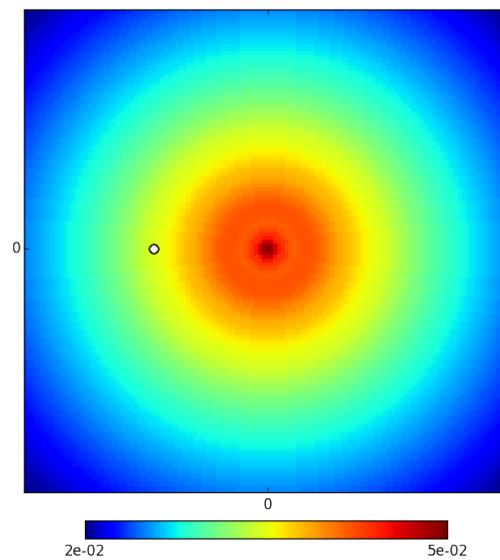


Figure 2.6: Recovered Galactic profile in the Galactic plane. Shown is a top-down view the Galaxy as in Fig. 2.5, but here in logarithmic color scale. The input model had medium strength fluctuations, it was recovered using 5000 pulsars with 5% distance uncertainty (corresponding to the bottom middle panel in Fig. 2.5). Other fluctuation strengths and data sets would yield a very similar image.

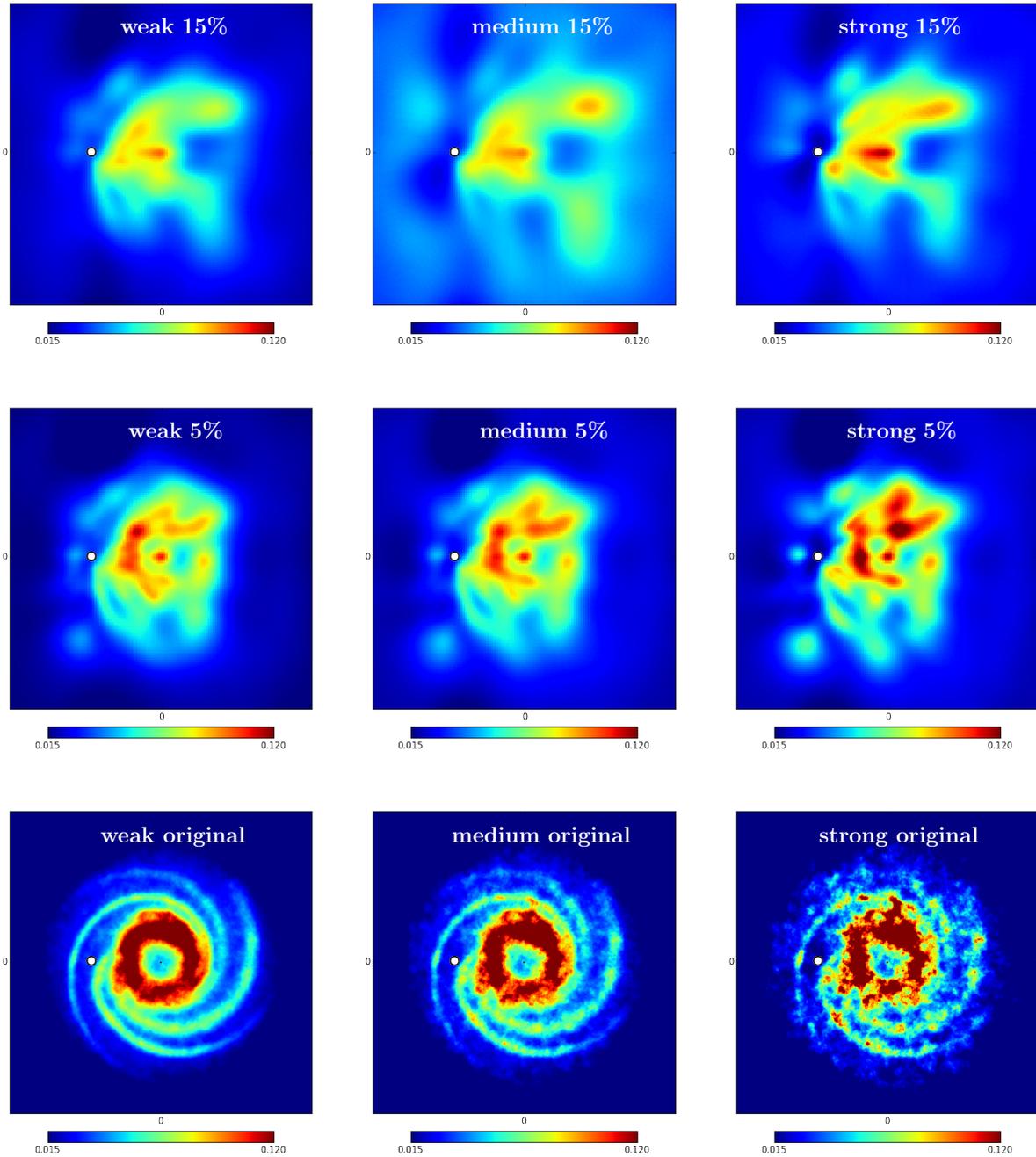


Figure 2.7: Reconstructions of the three Galaxy models with fluctuation strengths using 5000 pulsars. All panels show top-down views of the electron density in the Galactic plane using a linear color scale in units of cm^{-3} . The panels span 36000 pc in each dimension. The rows show reconstructions with distance errors of 15% and 5% (from top to bottom). The bottom row shows the original Galaxy models. The columns show reconstructions and original input models (modified NE2001) with weak, medium, and strong fluctuations (from left to right).

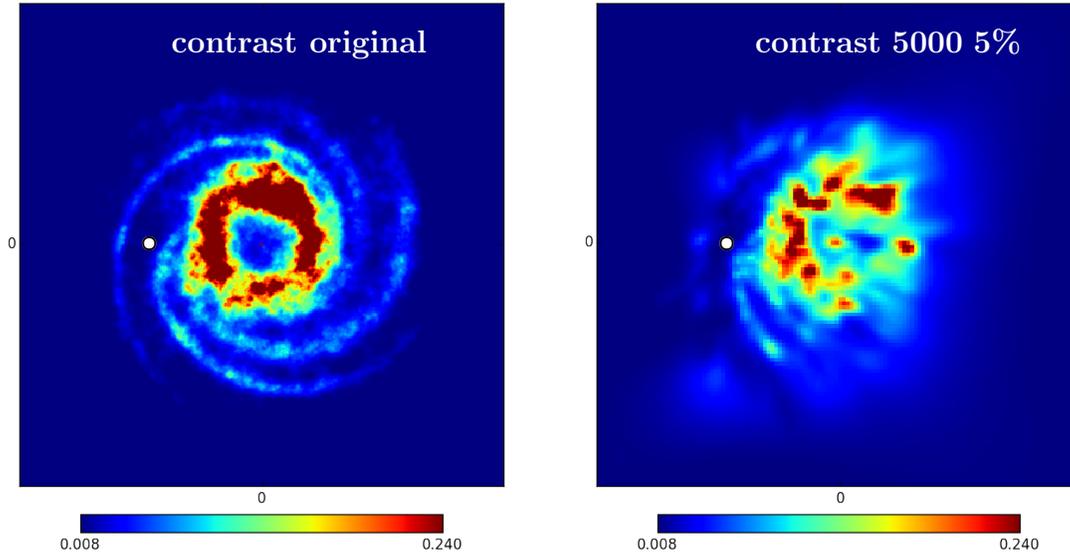


Figure 2.8: Input model (left) and reconstruction (right) of the contrast-enhanced Galaxy model using 5000 pulsars with distance uncertainties of 5%. Both panels show top-down views of the electron density in the Galactic plane using a linear color scale in units of cm^{-3} . The panels span 36000 pc in each dimension.

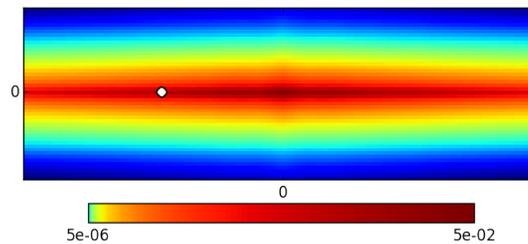


Figure 2.9: Vertical cut through the same Galactic profile as in Fig. 2.6. Shown is the slice containing the Sun (white dot) and the Galactic center (middle). The image spans $36 \text{ kpc} \times 12 \text{ kpc}$. The color scale is logarithmic.

This uncertainty is nearly independent of the quality of the data set, but depends on the strength of fluctuations on kpc scales. These are always present unless the data probe a simplistic disk. Therefore, there is a lower bound of precision to which our algorithm can determine the vertical fall-off behavior. We compared the reconstructed vertical scaling to a global and a local estimate generated from the original input model. The global estimate describes vertical fall-off throughout the whole model, whereas the local estimate describes the vertical fall-off close to the Sun¹⁶. For completeness we also provide best-fitting scale heights for exponential fall-offs in the figure, that is, we fit the vertical scaling to $n_e \propto e^{-|z|/H_z}$. The uncertainties of these estimates were calculated by performing the fit on multiple posterior samples of β . The local and global estimate both have significantly lower scale heights than the 950 pc from NE2001 (thick disk). This is probably due to the combination of the thick disk with the thin disk of NE2001 (which has a scale height of 140 pc). As is evident from the figure, the reconstructed z-profile is dominated by the local behavior of the density and agrees with it within the error bars throughout all data sets¹⁷ for the regime $|z| < 2400$ pc. However, the width of the uncertainty region prohibits a clear decision whether the vertical fall-off follows a single exponential function or a thick disk and a thin disk, as is the case for NE2001. In our input model we set n_e to zero for $|z| > 2400$ pc. In this regime our reconstruction is unreliable.

2.5 Summary and conclusions

We presented an algorithm that performs nonparametric tomography of the Galactic free electron density using pulsar dispersion measures and distances. The algorithm produces a three-dimensional map and a corresponding uncertainty map. It automatically estimates the correlation structure and the scales of the disk shape, requiring only approximate initial guesses for them. The uncertainties of pulsar distance estimates are consistently propagated.

Using our algorithm, we investigated the feasibility of nonparametric tomography with the upcoming Square Kilometer Array. To this end, we created three Galaxy models with various fluctuation strengths and one with enhanced contrast and simulated mock observations of these models using between 1000 and 10000 pulsars. Our results indicate that with the amount of pulsars that the SKA is expected to deliver, nonparametric tomography becomes feasible. However, detecting spiral arms in the free electron density from pulsar dispersion measures alone remains challenging if the input model has unenhanced contrast. We find that to distinguish the spiral arms in the vicinity of the Sun, between 5000 and 10000 pulsars with distance accuracies of between 5% and 15% are needed. The

¹⁶ The global estimate was calculated by averaging the logarithmic density at fixed vertical distances over the whole horizontal plane. The local estimate was calculated by averaging the logarithmic density at fixed vertical distances in a sub-area of the horizontal plane, which is centered on the Sun and has a size of $1500 \text{ pc} \times 1500 \text{ pc}$.

¹⁷ The reconstructed vertical fall-off is dominated by the near-Sun region since this is the part of the Galaxy where the density is reconstructed best.

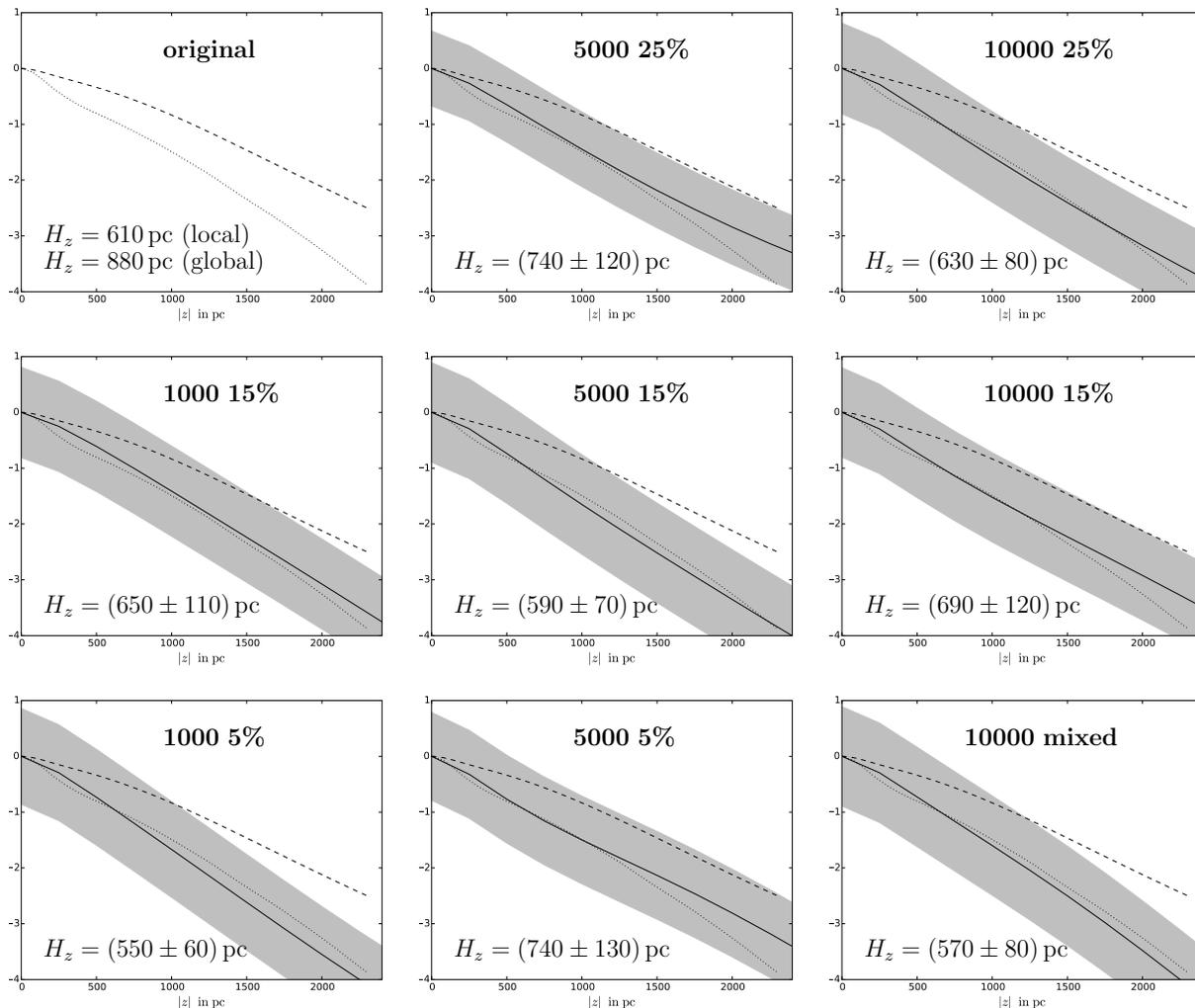


Figure 2.10: The recovered z -dependent fall-off in logarithmic units (function β in Eq. (2.40)), input model with medium strength fluctuations. The top left panel shows the global z -profile (dashed line) as well as the local z -profile (dotted line). In all other panels the solid line is the recovered z -profile while the global and local z -profile are replotted (in dashed and dotted respectively). The gray areas indicate the 1σ uncertainty around the recovered z -profile. In the bottom left corner of each panel we show the best fitting exponential scale height (and its 1σ uncertainty for the reconstructions).

vertical fall-off behavior of the free electron density was recovered for all mock data sets we investigated. However, a clear decision whether the vertical fall-off of free electron density is best described by a single exponential function or a thick disk and a thin disk could not be made by our algorithm.

One way to increase the sensitivity of the algorithm for Galactic features would be to include them into the prior description. By including higher order statistics (non-Gaussian priors), the inference could be made more sensitive for spiral arm structures in the electron density. In cosmology, for example, higher order statistics allowed for a better recovery of cosmological filaments (e.g., Jasche & Wandelt (2013)). Modeling HII regions and supernova remnants is also beyond the scope of Gaussian statistics. Another approach would be to include parametrized structures that are known from stellar observations, such as spatial templates for spiral arm locations. This would connect data-driven tomography (with infinite degrees of freedom) with classical model fitting.

The algorithm can also be used for other tomography problems with line-of-sight measurements, such as stellar absorption coefficients. It could also be extended to infer vector fields, enabling inference of the Galactic magnetic fields from pulsar rotation measures. Furthermore, a joint reconstruction of the Galactic free electron density and the magnetic field using pulsar dispersion, measures, pulsar rotation measures, and extragalactic Faraday sources needs to be investigated.

Acknowledgements

We wish to thank Niels Oppermann and Marco Selig for fruitful collaboration and advice. We also wish to thank Henrik Junklewitz, Jimi Green, Jens Jasche and Sebastian Dorn for useful discussions. The calculations were realized using the NIFTY¹⁸ package by Selig et al. (2013). Some of the minimizations described in Sect. 2.2.4 were performed using the L-BFGS-B algorithm (Byrd et al. (1995)). We acknowledge the support by the DFG Cluster of Excellence “Origin and Structure of the Universe”. The simulations have been carried out on the computing facilities of the Computational Center for Particle and Astrophysics (C2PAP). We are grateful for the support by Frederik Beaujean through C2PAP. This research has been partly supported by the DFG Research Unit 1254 and has made use of NASA’s Astrophysics Data System.

¹⁸<http://www.mpa-garching.mpg.de/ift/nifty/>

Chapter 3

fastRESOLVE: fast Bayesian imaging for aperture synthesis in radio astronomy

Note: This chapter, as well as Appendix C, have been published on arXiv e-prints (Greiner et al. 2016b) and have been submitted to Astronomy & Astrophysics for peer review and publication.

Abstract

The standard imaging algorithm for interferometric radio data, CLEAN, is optimal for point source observations, but suboptimal for diffuse emission. Recently, RESOLVE, a new Bayesian algorithm has been developed, which is ideal for extended source imaging. Unfortunately, RESOLVE is computationally very expensive. In this paper we present fastRESOLVE, a modification of RESOLVE based on an approximation of the interferometric likelihood that allows us to avoid expensive gridding routines and consequently gain a factor of roughly 100 in computation time. Furthermore, we include a Bayesian estimation of the measurement uncertainty of the visibilities into the imaging, a procedure not applied in aperture synthesis before. The algorithm requires little to no user input compared to the standard method CLEAN while being superior for extended and faint emission. We apply the algorithm to VLA data of Abell 2199 and show that it resolves more detailed structures.

3.1 Introduction

A fundamental observation method in modern radio astronomy is the aperture synthesis technique (see e.g. Ryle & Hewish 1960; Thompson et al. 1986; Finley & Goss 2000). Antennas of large interferometers are correlated to achieve resolutions comparable to a single dish instrument of the size of the whole array. The downside of the technique is an increase in the complexity of data processing, since the interferometer only measures

irregularly spaced sample points of the Fourier transform of the sky brightness leaving large unsampled regions in the Fourier plane. An inverse Fourier transform would suffer from strong aliasing effects severely distorting the image and misplacing regions of high brightness etc. Therefore, the brightness distribution on the sky has to be estimated in a more elaborate way.

The most widely used imaging algorithm in radio astronomy is CLEAN, developed by Högbom (1974). The underlying assumption of CLEAN is that the image is composed of uncorrelated point sources. Consequently, CLEAN is very effective in imaging observations of point source dominated fields (see e.g. Thompson et al. 1986; Taylor et al. 1999; Sault & Oosterloo 2007). There are multiple variants and extensions to CLEAN reducing the computational effort (Clark 1980), improving the performance for multi-frequency observations (Sault & Wieringa 1994), and implementing corrections (Schwab 1984). Reconstruction of extended objects has been addressed by implementing differently scaled kernel functions instead of sharp point sources as a basis, resulting in the Multi-Scale (MS) CLEAN and Adaptive Scale Pixel (ASP) algorithm (Bhatnagar & Cornwell 2004; Cornwell 2008; Rau & Cornwell 2011). However, how to choose the scales for MS CLEAN remains a non-trivial task left to the user and an implementation of the ASP algorithm is yet to be published.

The underlying assumption of the image being composed of point sources (or kernel functions) remains a fundamental ingredient in CLEAN, which hinders its performance in reconstructing extended diffuse sources. Furthermore, it is not known how to propagate measurement uncertainty through CLEAN and consequently, no uncertainty map for the reconstruction is provided. To overcome this issues a new algorithm called RESOLVE (**R**adio **E**xtended **S**ources **L**ognormal decon**V**olution **E**stimator) has been developed by Junklewitz et al. (2016). RESOLVE was developed within the framework of *information field theory* of Enßlin et al. (2009) and fulfills two main objectives¹:

1. It should be optimal for extended and diffuse radio sources.
2. It should include reliable uncertainty propagation and provide an error estimate together with an image reconstruction.

RESOLVE is a Bayesian algorithm. As such it needs to transform between image and data space many times in order to calculate the estimate of the brightness distribution on the sky. These transformations are costly, as they involve a Fourier transform to an irregular grid. In consequence, RESOLVE is computationally much slower than CLEAN. However, the Bayesian approach has the advantage that it requires fewer input parameters by the user since regularization of the result happens automatically and the additional time needed to determine some of these input parameters in CLEAN partly compensates for the additional computation time. Nevertheless, significant performance gains have to be achieved in order to make RESOLVE a widely used imaging algorithm for interferometric data. Furthermore, a Bayesian algorithm finds an estimate by weighing the prior against the likelihood. This weighting relies on an accurate description of the uncertainty of the recorded visibilities

¹The objectives are a direct quote from Junklewitz et al. (2016).

which is not always given for interferometric data. Algorithms like CLEAN, which operate solely on the likelihood are insensitive to a global factor in the measurement uncertainty, but they require a manually chosen cut-off criterion to avoid overfitting. Choosing this criterion properly is again related to estimating the measurement uncertainty.

In this paper we introduce fastRESOLVE, an algorithm that reduces the computational cost of RESOLVE by avoiding the costly irregular Fourier transforms. Instead, the data are gridded to a regular Fourier grid in an information theoretically optimal way. Additionally, we include a measurement variance estimator into the procedure allowing us to overcome the dependency of an accurate description of the measurement uncertainty. Furthermore, we introduce a hybrid approach to simultaneously reconstruct point-sources and diffuse emission using a preprocessing similar to CLEAN. Thus, we tackle the major disadvantages of RESOLVE compared to CLEAN, bringing it a step closer to a widely applicable imaging algorithm for interferometric data.

The remainder of this paper is structured as follows. In section 3.2 we derive fastRESOLVE focusing on the approximations and add-ons that distinguish fastRESOLVE from RESOLVE. In section 3.3 fastRESOLVE is tested and compared to CLEAN using archival data of Abell 2199. Section 3.4 provides a summary of the findings as well as an outlook to further developments.

3.2 Algorithm

3.2.1 Signal model

The signal of interest is the intensity of electromagnetic radiation at some frequency across a patch of the sky, $I(x, y)$. The position on the sky is here described by the Cartesian coordinates x and y . The coordinates span a plane tangential to the celestial sphere with the point of tangency being the center of the observation. This plane is called the uv -plane. The relationship between the intensity I on the sky and the visibilities V an interferometer records is described by the interferometer equation, which – for a small patch of the sky (see e.g. Thompson et al. 1986) – is essentially a Fourier transformation:

$$V(u, v) = \int dx dy e^{-2\pi i (ux + vy)} B(x, y) I(x, y). \quad (3.1)$$

The term $B(x, y)$ describes the primary beam of the instrument. The coordinates u and v describe the displacement (projected onto the uv -plane) between a pair of antennas in units of the observed wavelength. Since there is a finite set of antenna combinations forming a finite set of uv -locations only discrete visibilities are observed. We label these with an index i ,

$$V_i \equiv V(u_i, v_i) \quad (3.2)$$

For notational convenience we introduce the response operator R describing the linear relationship between the data d and the intensity I ,

$$R_i(x, y) = \exp(-2\pi i (u_i x + v_i y)) B(x, y). \quad (3.3)$$

To further improve readability we use a compact matrix notation for objects living in the continuous sky. The application of R to I is denoted as,

$$V_i = (RI)_i := \int dx dy R_i(x, y) I(x, y). \quad (3.4)$$

Similarly we define scalar products of continuous quantities,

$$I^\dagger j := \int dx dy I^*(x, y) j(x, y), \quad (3.5)$$

where $*$ denotes complex conjugation. The \dagger symbol therefore denotes transpositions and complex conjugation. For objects living in the discrete data space, there are analogous operations with sums instead of integrals, e.g.,

$$(R^\dagger V)(x, y) := \sum_i R_i^*(x, y) V_i. \quad (3.6)$$

The incompleteness of the uv -plane is the main challenge in interferometric imaging. Additionally, the recorded data are subject to instrumental noise. If the full uv -plane was measured and there was no instrumental noise, the intensity I could be simply recovered by inverse Fourier transformation. The noise is mostly of thermal origin and here assumed to be signal independent², additive Gaussian noise,

$$\begin{aligned} d &= V + n, \\ \mathcal{P}(n) &= \mathcal{G}(n, N), \end{aligned} \quad (3.7)$$

where \mathcal{G} describes a multivariate Gaussian probability distribution,

$$\mathcal{G}(n, N) := |2\pi N|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} n^\dagger N^{-1} n\right), \quad (3.8)$$

with the covariance matrix $N := \langle nn^\dagger \rangle_{\mathcal{P}(n)}$. The covariance matrix is assumed to be diagonal in this paper, meaning that the noise contamination of different visibilities is not correlated,

$$N_{ij} = \delta_{ij} \sigma_i^2. \quad (3.9)$$

Here, σ_i is the 1σ uncertainty interval of the data point i .

3.2.2 Irregular sampling

Any algorithm that aims to find an estimate of I given d will have to apply the response R and thus evaluate the interferometer equation Eq. (3.1) multiple times in order to ensure compatibility between the data and the estimated intensity. Therefore, the speed of this operation is of crucial importance for the performance of the algorithm and decisive in

²The independence of the noise is of course an approximation.

whether the algorithm is feasible to apply or not. The interferometer equation Eq. (3.1) involves a Fourier transform. In numerical applications this is typically carried out using a Fast Fourier transform (FFT). The FFT scales as $\mathcal{O}(N \log N)$ which is much faster than the $\mathcal{O}(N^2)$ of a direct Fourier transform. However, the FFT is only applicable if the sampling points are arranged in a regular grid. For a two-dimensional space such as a patch of the sky this means that x and y have to be distributed according to

$$x = j\Delta x, \quad \text{for } -N_x/2 \leq j < N_x/2, \quad (3.10)$$

where j is an integer and N_x is the number of pixels in the x -dimension and Δx is the pixel edge length³. With a corresponding equation for y and the assumption of periodic boundary conditions the sampling theorem states that a conjugate basis of k and q allows for lossless back and forth Fourier transformations if k and q are distributed according to

$$\begin{aligned} k &= j\Delta k, \quad \text{for } -N_x/2 \leq j < N_x/2, \\ \Delta k &= \frac{1}{N_x\Delta x} \end{aligned} \quad (3.11)$$

with a corresponding expression for q and y . Only with such a correspondence between the sample points of x , y and k , q is the FFT applicable.

The sample points of x and y can be freely chosen in the interferometer equation. However, the sample points on the left-hand-side are the uv -points sampled by the interferometer and cannot be freely chosen. There is basically no antenna setup in the real world where the uv -points can be brought into a form like Eq. (3.11) with an acceptable number for N_x . Simply mapping the uv -points to their closest (regular) kq -point yields strong aliasing effects contaminating the results. Calculating the aliasing explicitly yields a gridding operation between the visibility in a regular Fourier grid and the visibilities at irregularly spaced uv -points:

$$V(k, q) = \sum_j G_j(k, q) V_j, \quad (3.12)$$

where G is the gridding operator (see Appendix B.1),

$$G_j(k, q) \approx \frac{1}{\Delta k \Delta q} \text{sinc}\left(\frac{k - u_j}{\Delta k}\right) \text{sinc}\left(\frac{q - v_j}{\Delta q}\right). \quad (3.13)$$

In this paper we use the normalized sinc function, $\text{sinc}(x) := \sin(\pi x)/(\pi x)$ (with $\text{sinc}(0) = 1$ of course). To make the notation more compact, we will use the vector notation $\vec{k} \equiv (k, q)$ and $\vec{x} \equiv (x, y)$ throughout the rest of this paper.

³The origin of the coordinate system can of course be shifted arbitrarily.

3.2.3 The likelihood

The likelihood is a probability density function describing the possible outcomes of a measurement assuming a known signal configuration. Since we are assuming signal independent and additive Gaussian noise, the likelihood is a Gaussian centered on the noise-free visibilities,

$$\begin{aligned} \mathcal{P}(d|I) &= \mathcal{G}(d - RI, N) \\ &= |2\pi N|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} (RI - d)^\dagger N^{-1} (RI - d)\right). \end{aligned} \quad (3.14)$$

The goal of our inference is to find an estimate of the intensity I . Therefore, the only terms of the likelihood that are of relevance to us are the ones dependent on I ,

$$\mathcal{P}(d|I) \propto \exp\left(-\frac{1}{2} I^\dagger M I + j^\dagger I\right), \quad (3.15)$$

with the measurement precision operator M and the information source j ,

$$\begin{aligned} M &= R^\dagger N^{-1} R, \\ j &= R^\dagger N^{-1} d. \end{aligned} \quad (3.16)$$

The measurement precision M can be represented as

$$M = \hat{B} F \hat{B}, \quad (3.17)$$

where \hat{B} is a diagonal operator in position space with the primary beam on the diagonal and

$$\begin{aligned} F(\vec{k}, \vec{k}') &= \sum_{j, j'} G_j^*(\vec{k}) (N^{-1})_{jj'} G_{j'}(\vec{k}') \\ &= \sum_j \frac{1}{\sigma_j^2} G_j^*(\vec{k}) G_j(\vec{k}'). \end{aligned} \quad (3.18)$$

If the sampled uv -points were matching the k and q points of the regular Fourier grid perfectly, the operator F would be a simple (inverse noise weighted) mask in Fourier space, a diagonal operator. The points do not match, of course, so F is not diagonal. The full matrix F cannot be stored in the memory of a computer in realistic scenarios, so to apply F the gridding operator G has to be applied twice, which is numerically expensive, even though it can be done quicker than by invoking a direct Fourier transform (see e.g. Briggs et al. 1999).

We want to approximate F by \tilde{F} , a diagonal operator in Fourier space, while preserving as much information as possible. This would enable us to apply M without the expensive gridding operation, thus increasing the numerical speed significantly. As a criterion of

information loss we use the Kullback-Leibler divergence, sometimes called relative entropy,

$$D_{\text{KL}}[\tilde{\mathcal{P}}(x)|\mathcal{P}(x)] := \int \mathcal{D}x \tilde{\mathcal{P}}(x) \ln \frac{\tilde{\mathcal{P}}(x)}{\mathcal{P}(x)}, \quad (3.19)$$

where $\mathcal{D}x \equiv dx_1 dx_2 dx_3 \dots$ denotes integration over the full phase space (all degrees of freedom of x). By minimizing the Kullback-Leibler divergence with respect to \tilde{F} we find that most information is preserved if \tilde{F} is simply the diagonal of F ,

$$\tilde{F}(\vec{k}, \vec{k}') = \delta(\vec{k} - \vec{k}') F(\vec{k}, \vec{k}'). \quad (3.20)$$

For the detailed calculation we refer the reader to Appendix B.2.

3.2.4 Inference

In this section we outline the inference algorithm. The algorithm is the original RESOLVE algorithm which is derived in detail by Junklewitz et al. (2016) save for two additions (Sec. 3.2.4 and Sec. 3.2.4) and an additional approximation (Appendix B.3). The inference combines the interferometric likelihood with a log-normal prior with a homogeneous and isotropic covariance,

$$I = \exp(s), \quad \mathcal{P}(s) = \mathcal{G}(s, S), \quad S = \sum_i p_i S^{(i)}, \quad (3.21)$$

where p_i are unknown power spectrum parameters and $S^{(i)}$ are disjoint projection operators onto the different spectral bands⁴. The reasoning behind this choice for a prior is that the intensity is a positive definite quantity. Furthermore, the algorithm should not assume any preferred direction or position prior to considering the data. This leaves us with an unknown power spectrum for which we choose a hyper-prior which prefers (but does not enforce) the power spectrum to be a power law (Oppermann et al. 2013; Enßlin & Weig 2010; Enßlin & Frommert 2011). Using a maximum a posteriori Ansatz for s and the empirical Bayes method⁵ (see e.g. Robbins 1956) for p this yields the following set of equations for the logarithmic map m , its covariance estimate D , and the power spectrum p :

$$m = \arg \min_s (-\ln \mathcal{P}(s, d|p)), \quad (3.22)$$

$$D = \left(-\frac{\delta^2}{\delta s \delta s^\dagger} \ln \mathcal{P}(s, d|p) \Big|_{s=m} \right)^{-1}, \quad (3.23)$$

$$p_i = \frac{q_i + \frac{1}{2} \text{tr}[(mm^\dagger + D) S^{(i)}]}{\alpha_i - 1 + \varrho_i/2 + (T \ln p)_i}. \quad (3.24)$$

⁴For a more detailed introduction of the concept we refer the reader to the references provided in this section.

⁵This can also be seen in the context of Variational Bayes, see e.g. Selig & Enßlin (2015).

Here, $q_i = \text{tr } S^{(i)}$ is the amount of degrees of freedom in the i th band whereas q_i , α_i , and T come from the hyper-prior of the power spectrum parameters (see Junklewitz et al. (2016) & Oppermann et al. (2013) for details). In Eq. (3.24) D is used in an approximated form which is numerically cheaper to calculate. This approximation is described by Eq. B.24 in Appendix B.3. The solutions for m and D depend on p , which in turn depends on m and D . Therefore, these equations have to be iterated until a fixed point is reached. The resulting map m is the estimate of the logarithmic intensity I while D is an estimate of the posterior covariance of m .

Measurement variance estimation

The reconstruction of the diffuse emission is sensitive to the variances in the likelihood Eq. (3.14). However, these variances are often not available in a way that is consistent with Eq. (3.14). Therefore, we need to estimate these variances along with the emission itself. The estimate for σ^2 is derived in Appendix B.4 as

$$\sigma_i^2 = |(Re^m - d)_i|^2. \quad (3.25)$$

This equation is prone to overfitting, since individual data points can yield a very small σ leading to an even stronger fit in the next iteration. We therefore introduce a regularization,

$$\sigma_i^2 \leftarrow t \sigma_i^2 + \frac{1-t}{N_{\text{data}}} \sum_j \sigma_j^2, \quad (3.26)$$

where $0 \leq t \leq 1$. If the data are properly calibrated (e.g. outliers have been flagged), we recommend to set t to 0, i.e., setting all σ_i equal. If the data contain a significant amount of outliers t has to be increased to allow the σ_i to weigh them down. For such cases we found $t = 0.9$ to be a good value. If different data sets are combined we recommend to perform the regularization in Eq. (3.26) separately for each data set.

Point source removal

The prior assumptions outlined in this section so far implicitly assume the intensity to be spatially correlated due to the prior assumption of a statistically homogeneous and isotropic log-normal field. This assumption is obviously not very well suited for point sources, which are present in most radio observations. Point sources are often orders of magnitude brighter than the extended sources of radio emission and can thus complicate the reconstruction of the latter, especially if the measurement operator in the likelihood has been approximated. To overcome this we subtract point sources from the data by an iterative procedure somewhat similar to CLEAN. To that end the intensity is divided into the extended emission intensity I and the point source emission intensity I_{point} . The goal is to determine I_{point} first and apply the algorithm for extended emission on the point source removed data. The iterative procedure is stopped when the amount of pixels containing a point source reaches a cut-off criterion. One could of course use CLEAN itself for this,

but we implemented our own procedure to understand all the subtleties. We describe it in Appendix B.5. A more rigorous procedure which combines RESOLVE with the D³PO algorithm (Selig & Enßlin 2015) is in preparation (Junklewitz et al. in prep., working title pointRESOLVE).

After I_{point} is determined and subtracted the remaining data should contain mostly extended emission and can be further processed using Eqs. (3.22) to (3.24) to reconstruct the extended emission.

Filter procedure

The filter equations and procedures described in this section allow us to determine the point-like emission, diffuse emission, the power spectrum of the diffuse emission, and the noise level (measurement variance). Due to the non-convex nature of the system of equations the starting point of the solving strategy is of importance. The following initial guesses for the power spectrum and noise variances proved to be useful.

The initial power spectrum is set to $p_i \propto \ell_i^{-3}$, where ℓ is the average length of the k -vectors in band i . The prefactor is chosen so that the variance of a corresponding random field is ~ 1 . The monopole p_0 is chosen to be practically infinity (but still a floating point number). This means that the relative fluctuations of I are a priori assumed to be of order e while its absolute scale is free.

The initial noise variances σ_i^2 are set to half of the average data power. This corresponds to an assumed signal-to-noise ratio of 1.

With these starting values we solve the equations in the following way:

1. separate point sources according to Appendix B.5
2. update the power spectrum by iterating Eqs. (3.22) to (3.24) several times
3. repeat from step 1 until no more point sources are found
4. update the variances according to Sec. 3.2.4
5. update the map according to Eq. (3.22) and repeat from step 4 until the variances do not change anymore.

The resulting estimates of $I = e^m$, I_{point} , p , and σ^2 are our final results.

Posterior uncertainty map

After the iterative procedure described in 3.2.4 has converged, the posterior variance of s can be approximated (within the maximum a posteriori approach) by

$$\left\langle \left[s(\vec{x}) - \langle s(\vec{x}) \rangle_{\mathcal{P}(s|d)} \right]^2 \right\rangle_{\mathcal{P}(s|d)} \approx D(\vec{x}, \vec{x}), \quad (3.27)$$

where we use the approximative form of D described Eq. (B.23) in Appendix B.3. Using this result we estimate posterior variance of $I = e^s$ as

$$\left\langle \left[I(\vec{x}) - \langle I(\vec{x}) \rangle_{\mathcal{P}(s|d)} \right]^2 \right\rangle_{\mathcal{P}(s|d)} \approx e^{m(\vec{x})} \left(e^{D(\vec{x}, \vec{x})} - 1 \right) e^{m(\vec{x})}. \quad (3.28)$$

It is important to note that this uncertainty estimate is derived within a saddle-point approximation to a non-convex problem. Consequently, it cannot detect artifacts in the maximum a posteriori solution m . The uncertainty estimate is rather to be interpreted as the uncertainty assuming m was a good estimate. It does not consider that structures could be misplaced due to aliasing for example. However, it is still useful to have even a crude uncertainty estimate compared to having no estimate at all, which is the case for all established imaging methods for aperture synthesis.

Furthermore the uncertainty estimate only considers the statistical error of diffuse component of the image. It does not include an estimate of the systematic uncertainty. The point-like component is regarded as fixed. fastRESOLVE does not provide an uncertainty estimate for it. This will be fixed by the soon to be published pointRESOLVE addition mentioned in Sec. 3.2.4.

3.3 Application

In this section we demonstrate the performance of fastRESOLVE. In the original publication (Junklewitz et al. 2016) of RESOLVE its performance on simulated data is tested and compared thoroughly to CLEAN and the Maximum Entropy Method. Instead of reproducing these comparisons for fastRESOLVE, we validate the behavior of fastRESOLVE against standard RESOLVE to check that the approximations and add-ons in fastRESOLVE do not compromise the results. These mock tests can be found in Appendix B.6. The tests show that fastRESOLVE is consistent with RESOLVE, but the latter recovers the input data to a higher accuracy. This is to be expected given the approximations involved.

Here we want to focus on the performance using real data. To this end we use observations of Abell 2199 and compare their fastRESOLVE images to their respective CLEAN images. This is the first time that a real data reconstruction using any variant of RESOLVE is published.

3.3.1 Abell 2199

We applied fastRESOLVE to radio observations of the cool core galaxy cluster A2199. The brightest galaxy in this cluster is the powerful active galactic nucleus (AGN) 3C 338, a restarting Fanaroff-Riley I radio galaxy with two symmetric jets on parsec-scales, displaced from the large-scale emission of the lobes. 3C 338 is associated with the multiple nuclei optical cD galaxy NGC 6166, while X-ray observations of the cluster indicate the presence of cavities associated with the lobes. This radio source has been studied by several authors (e.g. Burns et al. 1983; Fanti et al. 1986; Feretti et al. 1993; Giovannini et al.

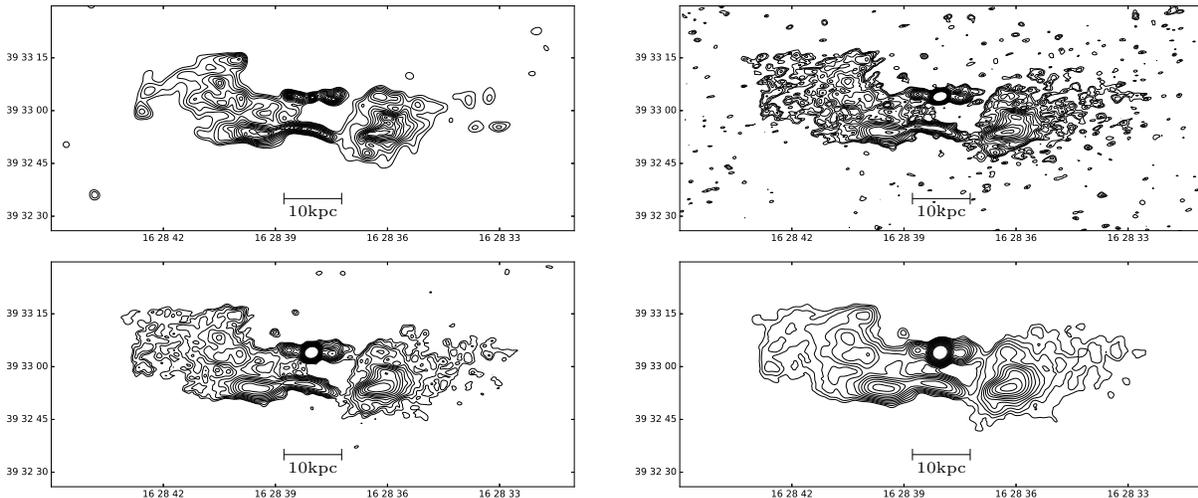


Figure 3.1: Total intensity contours of 3C 388 at 8415 MHz. The top-left panel shows the fastRESOLVE image, the other panels show the CLEAN images at uniform (top right), intermediate (bottom left) and natural (bottom right) weighting. Contour lines start at $5.2 \times 10^5 \text{ Jy/rad}^2$ and increase by factors of $\sqrt{2}$. The coordinates are J2000 declination on the y-axis and J2000 right ascension on the x-axis.

1998; Markevitch & Vikhlinin 2007). Recently, Vacca et al. (2012) derived the intracluster magnetic field power spectrum by using archival VLA observations over the frequency band 1 – 8 GHz. In this work we use the VLA data at 8145 GHz in C and BC configurations presented in their paper. The details of the observations are given in Table 3.1. We refer to their paper for the description of the data calibration. The combined dataset at 8145 GHz was used to produce images of the source with the CLEAN algorithm implemented in AIPS with natural, uniform, and intermediate weighting. These images are compared with the image obtained with fastRESOLVE.

Table 3.1: Summary of the observations used in this paper. We refer to Vacca et al. (2012) for further details.

Frequency (MHz)	Bandwidth (MHz)	Project	Date	Time (h)	VLA configuration	RA (J2000)	Dec (J2000)
8415	50	BG0012	94-Nov.-17	3.0	C	16h28m38.251s	+39d33'04.22''
8415	50	GG0038	00-Feb.-26	3.8	BC	16h28m38.243s	+39d33'04.19''
1665	50	GG0005	91-Jun.-18/19	8.5	A	16h28m38.232s	+39d33'04.14''

In Figs. 3.1 and 3.2 we compare the fastRESOLVE reconstruction with the CLEAN reconstructions using different weightings. Additionally we show a superposition of the fastRESOLVE image with the uniform weighting CLEAN image in Fig. 3.3. The reconstructions agree in their large-scale features. Small-scale features, which are comparable to the size of the beam, show the advantage of fastRESOLVE. The brightest structures are sharper and more detailed than in the CLEAN image with uniform weighting, while the

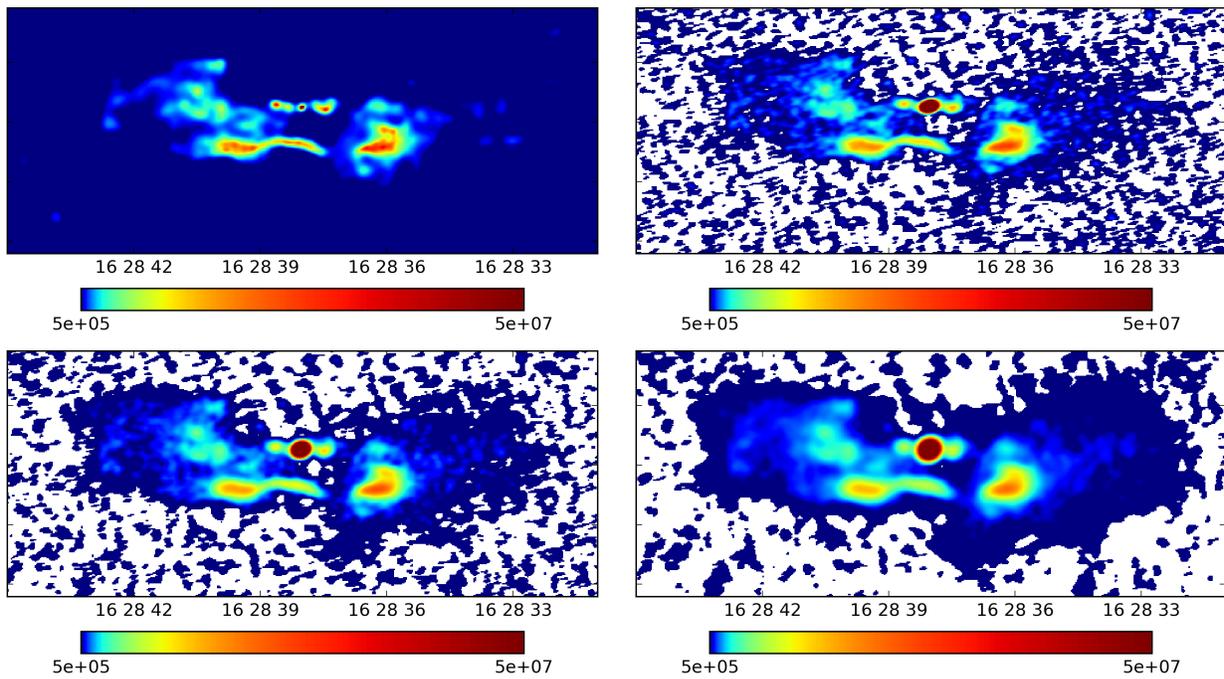


Figure 3.2: Total intensity of 3C388 at 8415 MHz. The top-left panel shows the fastRESOLVE image, the other panels show the CLEAN images at uniform (top right), intermediate (bottom left) and natural (bottom right) weighting. The units are Jy/rad^2 . The white regions in the CLEAN image are regions with negative intensity. The panels show the same field of view as in Fig. 3.1

Table 3.2: Flux of 3C 338 derived with the different approaches.

weighting/mask	CLEAN flux	fastRESOLVE flux
uniform	160 mJy	157^{+18}_{-14} mJy
intermediate	166 mJy	164^{+19}_{-16} mJy
natural	182 mJy	169^{+22}_{-17} mJy
no mask	–	187^{+35}_{-24} mJy

faint structures are almost as smooth as in the natural weighting image. In fastRESOLVE the effective resolution is chosen automatically through the interplay of prior and likelihood, i.e. structures which are noise dominated are washed out. This allows fastRESOLVE to adapt the resolution locally, depending on the strength of the data for that specific location (see the detailed discussion in Junklewitz et al. 2016). Therefore, the faint regions are smoother in fastRESOLVE whereas the bright regions are sharper. One can find many features which are hinted at in the CLEAN image to be much more detailed in the fastRESOLVE image. This effect of superior resolution compared to CLEAN will be thoroughly discussed in Junklewitz et al. (in prep.) using VLA data of the radio galaxy Cygnus A. Furthermore, the CLEAN image exhibits negative regions, whereas the fastRESOLVE image is strictly positive.

In Table 3.2 we list the total flux estimates obtained with fastRESOLVE and CLEAN. The estimates are derived by masking out all the regions with a flux below 0.06mJy/beam (three times the noise of the CLEAN images) and integrating⁶ the remaining flux for each CLEAN image. This yields three different masks for which we list the corresponding fluxes of the fastRESOLVE image, as well. Since the fastRESOLVE image has no negative regions there is no imperative need for a mask. We therefore list the total flux of the fastRESOLVE image, too. All derived CLEAN fluxes agree with the corresponding fastRESOLVE fluxes within the credibility intervals. Since fastRESOLVE separates point-like and diffuse flux, we can state an estimate for the total point-like flux. It is 84mJy. An uncertainty estimate for the point-like flux is not provided by fastRESOLVE.

Fig. 3.4 shows the estimate of the relative uncertainty (1σ credibility interval) of the fastRESOLVE image, a quantity that is not accessible for a CLEAN reconstruction. We see that the relative uncertainty is larger for faint regions and smaller for bright regions. This is because bright regions have a larger impact on the data and are therefore easier to separate from the noise. It is important to note that this is only an estimate of the credibility interval using a saddle point approximation not the result of a full sampling of the posterior distribution. Thus it is likely that the true credibility interval is larger.

⁶All of the integration in this paragraph are performed over a square with an edge length of $2'33.6''$ centered on the pointing and including all the source emission. This is to avoid artifacts towards the edge of the primary beam which can be present in both CLEAN and fastRESOLVE.

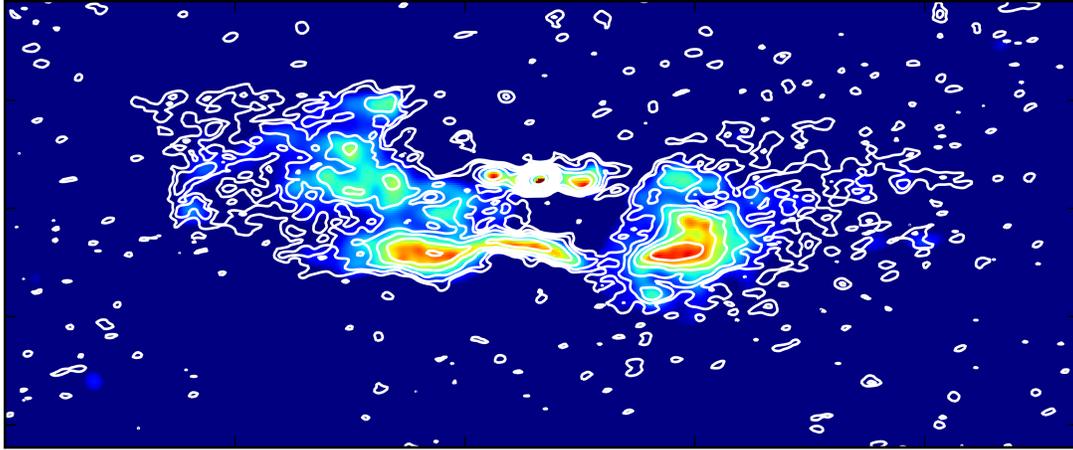


Figure 3.3: Superposition of the fastRESOLVE and CLEAN image of 3C 388 at 8415 MHz. The color plot shows the fastRESOLVE image, the contour lines show the CLEAN image (uniform weighting). The base contour, the color code, and the field of view are the same as in Figs. 3.1 & 3.2. The contour lines increase by factors of 2.

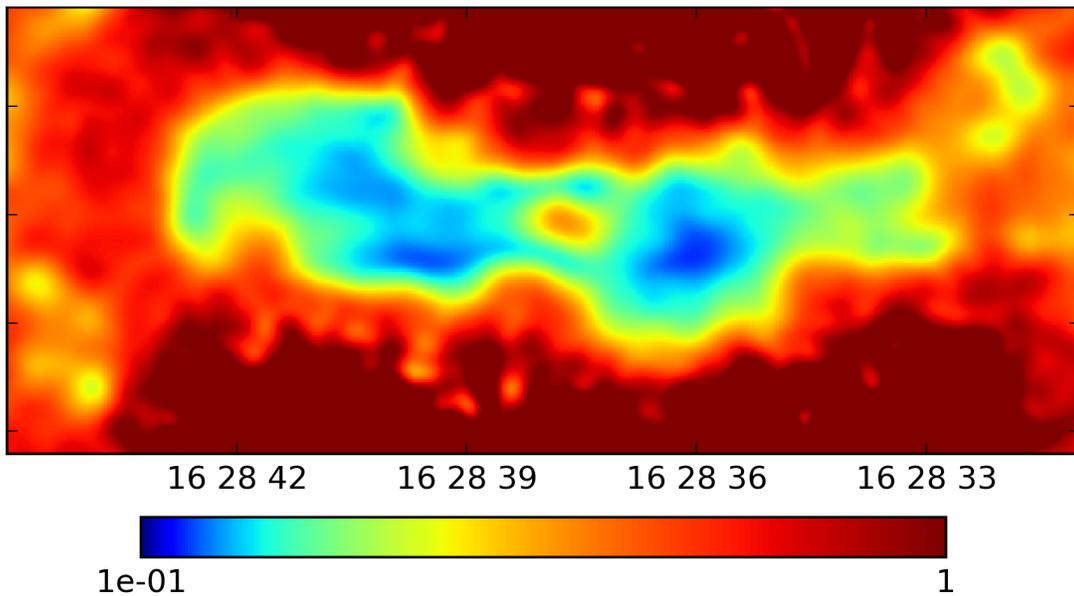


Figure 3.4: Relative uncertainty of the fastRESOLVE image of 3C 388 at 8415 MHz. The same field of view as in Fig. 3.1 is shown.

fastRESOLVE provides an estimate of the likelihood variances as well. The variance⁷ of the visibilities is uniformly ($t = 0$ in Eq. (3.26)) estimated as $\sigma^2 = (13.6 \text{ mJy})^2$. Thus the average signal-to-noise ratio is $SNR \approx 37$ for the point-like emission, $SNR \approx 1$ for the diffuse emission and $SNR \approx 40$ for the combined emission. The point source is much stronger imprinted in the data than the diffuse flux, since a sharp peak excites all Fourier modes (see Appendix B.7 for the SNR definition used in this work).

Computing time

Comparing the computing time of fastRESOLVE to CLEAN and RESOLVE is not straight forward. RESOLVE and fastRESOLVE do not require user input during the runtime. CLEAN typically needs repeated user input (interactive CLEANing) to select the appropriate regions in which flux is expected. Therefore, its performance cannot be characterized in computation time alone. In fastRESOLVE and RESOLVE the convergence criteria of the minimization in Eq. (3.22) play a decisive role for the run-time and since they use different minimization schemes it is not entirely clear how to compare these criteria. RESOLVE uses a gradient descent with Wolfe conditions while fastRESOLVE uses the L-BFGS-B algorithm (Byrd et al. 1995). However, in Appendix B.6.1 we find the computation time of fastRESOLVE to be roughly 100 times shorter compared to RESOLVE in our mock test. What we compare here is the time it takes to evaluate the likelihood and its gradient. Any efficient minimization algorithm will have to evaluate both repeatedly. After removing flagged data points the data set used in the previous section consists of 750489 uv -points. Evaluating the likelihood in RESOLVE took 11.9 seconds using a single core of an Intel Xeon E5-2650 v3 @ 2.30GHz. Using the same setup the likelihood in fastRESOLVE took 52 milliseconds to evaluate. The gradients took 27.1 seconds in RESOLVE and 77 milliseconds in fastRESOLVE. Combined with the evaluation of the prior, which is the same in RESOLVE and fastRESOLVE, this yields speed up by a factor of over 100. The fastRESOLVE likelihood scales as $\mathcal{O}(N_{\text{pix}} \log N_{\text{pix}})$ and the RESOLVE likelihood scales as $\mathcal{O}(N_{\text{pix}} \log N_{\text{pix}} + WN_{\text{uv}})$, where W is a precision parameter and N_{uv} is the amount of uv -points. The prior scales as $\mathcal{O}(N \log N)$ for both algorithms. Therefore, the speedup factor should increase with the amount of uv -points and decrease with the amount of image pixels.

In the fastRESOLVE run presented in the previous chapter it took 3 hours and 22 minutes on one CPU core to calculate the map m and 4 hours and 8 minutes on 2 cores to calculate the uncertainty map. Applying CLEAN to the data took approximately 10 minutes of interactive work using one core of an AMD FX-8350 @ 4GHz. This is still a big time difference – it is roughly a factor of 20 for the image alone – but it makes it feasible to run fastRESOLVE on a conventional notebook. The calculation of the fastRESOLVE image was performed using rather strict convergence criteria to be on the safe side. By relaxing the convergence criteria we managed to decrease the computation time of the map to 50 minutes without significantly changing the result. Another factor is the FFT implementation. Currently, fastRESOLVE uses the numpy FFT implementation. By using

⁷Note that σ^2 is an estimate of the statistical uncertainty of the measured visibilities, not of the reconstructed flux.

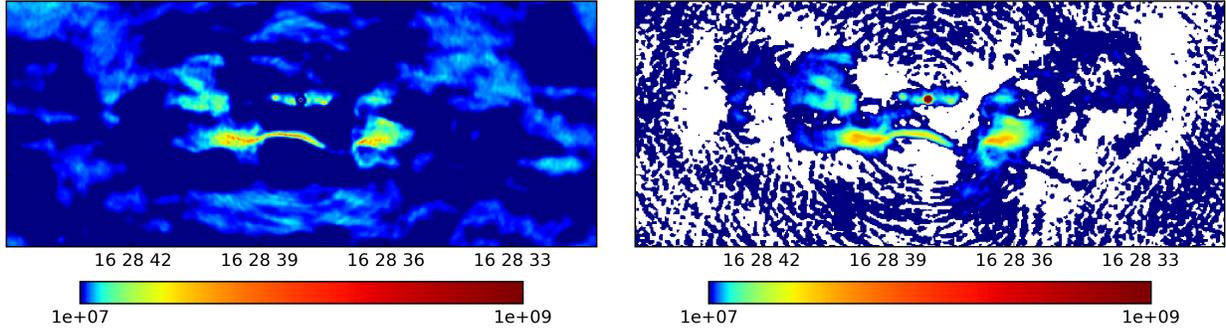


Figure 3.5: Total intensity of 3C 388 at 1665 MHz. The left panel shows the fastRESOLVE image, the right panel shows the CLEAN image at uniform weighting. The units are Jy/rad^2 . The white regions in the CLEAN image are regions with negative intensity. The panels show the same field of view as in Fig. 3.1

the FFTW (Frigo & Johnson 2005), there could be another speedup by a factor of 2 and even more if one uses more than one core (which is not possible in the numpy FFT).

A2199 – 1665MHz

We tested fastRESOLVE on several other data sets of 3C 388 at frequencies between 1.6 GHz and 8.4 GHz and compared the results to the corresponding CLEAN images. Most comparisons lead us to the same conclusions as the 8415 MHz data in the previous section. In this section we show the results of the data set for which fastRESOLVE performed worst in our opinion. The data were recorded at 1665 MHz, but the antenna configuration was different than in the previous section so that similar scales are probed by both, the 1665 MHz and the 8415 MHz data. The details of the data sets can be found in Tab. 3.1.

In Fig. 3.5 we show the fastRESOLVE image and the CLEAN image at uniform weighting. One can see that the fastRESOLVE image shows significant flux in regions where there should be no emission and that there is some artificial clumpiness on small scales. Furthermore, there is a ring of low emission around the nucleus, which is probably an artifact. Such a ring might be the result of overfitting the point-like nucleus, but we could not eliminate this artifact even by manually adjusting the size and strength of the point-like contribution to the image. The ring seems to be supported by the data which leads us to the suspicion that it is due to an imperfect calibration. Such a miscalibration could also explain the other artifacts in the fastRESOLVE image. Generally, it is fair to say that fastRESOLVE is more sensitive to calibration problems than CLEAN. However, even with the addressed problems the fastRESOLVE image still compares nicely to the CLEAN image in our opinion. They both resolve structures to a similar degree and they both exhibit a number of artifacts (eg. negative flux regions in the CLEAN image).

3.4 Discussion & Conclusion

We presented fastRESOLVE, a Bayesian imaging algorithm for aperture synthesis data recorded with a radio interferometer. It is much cheaper than RESOLVE in computation time without losing the key advantages RESOLVE has compared to CLEAN (optimality for diffuse emission and uncertainty propagation). Furthermore, fastRESOLVE is capable of reconstructing point sources and estimating the measurement variance of the visibilities. The estimation of the measurement variances is completely a new approach in aperture synthesis. It permits the application of Bayesian imaging schemes which rely on an accurate description of the measurement variances, which is often not provided for radio interferometric data.

We tested fastRESOLVE on archival VLA data of Abell 2199. We showed reconstructions at 1665 MHz and 8415 MHz. For the 8415 MHz data fastRESOLVE could resolve more detailed structures than CLEAN in uniform weighting while introducing fewer artifacts than CLEAN in natural weighting (especially no regions negative brightness). This behavior could be observed at most of the other frequencies we used for testing. We showed the 1665 MHz data set as a negative example. For this data set fastRESOLVE slightly over-fitted small-scale structures and introduced artifacts that put additional flux into empty regions. However, the overall quality of the image was still comparable to CLEAN at uniform weighting. Therefore, fastRESOLVE could be regarded as a replacement or at least supplement to CLEAN.

The speed-up of fastRESOLVE is achieved by introducing an approximation scheme in which the data are gridded onto a regular Fourier grid to allow for a quicker evaluation of the likelihood and its derivative. This approximation scheme could also serve as a general compression scheme in aperture synthesis, if the amount of uv points exceeds the amount of grid points of the desired image. This will be easily the case for measurements by the upcoming Square Kilometre Array (SKA). In such cases, one could save j and the diagonal description of M (see Sec. 3.2.3). Both need the same amount of space as the final image. Calibration, flagging, as well as the pixelization cannot be easily changed afterwards (if at all), but one preserves a description of the likelihood and can still apply different visibility dependent imaging and modeling schemes.

fastRESOLVE will be included into the main RESOLVE package as an optional feature. A hybrid version in which the result of fastRESOLVE is used to speed up a final RESOLVE run will also be implemented. Within this inclusion the multi-frequency capability of RESOLVE will also be enabled for fastResolve. The open source Python code is publicly available⁸. As has been mentioned earlier, there is another add-on under development which will enable a more consistent treatment of point-sources. All of these efforts will be joined to make the RESOLVE algorithm optimal for both, diffuse and point sources while keeping the computation time minimal.

⁸ ASCL: <http://ascl.net/1505.028>
github: <http://github.com/henrikju/resolve>

Acknowledgements

We wish to thank Martin Reinecke for his help with the C implementation of the gridding function. The calculations were realized using the NIFTY⁹ package by Selig et al. (2013). Some of the minimizations in this work were performed using the L-BFGS-B algorithm (Byrd et al. 1995).

⁹<http://www.mpa-garching.mpg.de/ift/nifty/>

Chapter 4

Log-transforming the matter power spectrum

Note: This chapter, as well as Appendix C, have been published in Astronomy & Astrophysics (Greiner & Enßlin 2015).

Abstract

We investigate whether non-linear effects on the large-scale power spectrum of dark matter, namely the increase in small-scale power and the smearing of baryon acoustic oscillations, can be decreased by a log-transformation or emulated by an exponential transformation of the linear spectrum. To that end we present a formalism to convert the power spectrum of a log-normal field to the power spectrum of the logarithmic Gaussian field and vice versa. All ingredients of our derivation can already be found in various publications in cosmology and other fields. We follow a more pedagogical approach providing a detailed derivation, application examples, and a discussion of implementation subtleties in one text. We use the formalism to show that the non-linear increase in small-scale power in the matter power spectrum is significantly smaller for the log-transformed spectrum which fits the linear spectrum (with less than 20% error) for redshifts down to 1 and $k \leq 1.0 h \text{ Mpc}$. For lower redshifts the fit to the linear spectrum is not as good, but the reduction of non-linear effects is still significant. Similarly, we show that applying the linear growth factor to the logarithmic density leads to an automatic increase in small-scale power for low redshifts fitting to third-order perturbation spectra and Cosmic Emulator spectra with an error of less than 20%. Smearing of baryon acoustic oscillations is at least three times weaker, but still present.

4.1 Introduction

In cosmology and astrophysics densities are often described as a combination of the mean density and fluctuations around it,

$$\rho = \rho_0 (1 + \delta), \quad (4.1)$$

where the density contrast field δ is often small and accurately described by a Gaussian random field. However, as soon as large overdensities occur ($\delta > 1$) a description of δ by Gaussian statistics does not suffice, since δ cannot go below minus one. A prominent example of such a density is the large-scale structure, where the overdensity is accurately described by Gaussian statistics at early epochs, but becomes highly non-linear at later epochs and small scales. In these non-linear regimes the density is much more accurately described by log-normal statistics, i.e., $\log(1 + \delta)$ following Gaussian statistics. This has already been noted by Hubble (1934) and Hamilton (1985) and theoretically investigated by Coles & Jones (1991). N-body simulations calculated by Kayo et al. (2001) are also in agreement with a log-normal distribution. However, as Carron (2011) showed, correlation functions and spectra contain very limited information in the case of highly non-linear log-normal statistics as there are other distribution functions that would produce the same moments. For a perturbative treatment of the power spectrum of $\log(1 + \delta)$ see Wang et al. (2011). Another example is the density in turbulent and isothermal clouds, which is very accurately described by a log-normal distribution (see e.g., Passot & Vázquez-Semadeni (1998) or Nordlund & Padoan (1999)).

In many cases, however, only the linear power spectrum (the power spectrum of δ) is available (e.g., from observations or simulations). It is therefore useful to convert a linear power spectrum into a logarithmic power spectrum (and vice versa). In this work we derive such a conversion formula under the assumption that the power spectrum is the power spectrum of the underlying statistical process. This condition is not necessarily met if the power spectrum is calculated from observational data or simulations, but in most cases the formula works nevertheless. Formulas relating the correlation functions of Gaussian and log-normal fields are widely known throughout several fields. In cosmology we refer the reader to Coles & Jones (1991), Politzer & Wise (1984), and Percival et al. (2004). The first goal of this paper is to provide a pedagogical introduction along with application examples enabling the reader to directly apply the derived formulas in a numerical setting while avoiding normalization and prefactor mistakes which can drastically influence the outcome due to the non-linearity of the exponential and logarithmic functions. The second goal is to apply the presented formalism to investigate how much a log-transformation of the large-scale density reduces the non-linear enhancement of small-scale power for decreasing redshifts.

The remainder of this paper is structured as follows. First, we go through the derivation of the conversion formalism in Sect. 4.2 in which we state all definitions explicitly before the calculation. Second, we apply the conversion to the large-scale matter spectrum in Sect. 4.3 to test the validity of the log-normal approximation of the cosmic matter density. Finally, we conclude this paper with a summary and outlook in Sect. 4.4.

In the Appendix, we describe how to generate log-normal random field following a given spectrum (Appendix C.1). Furthermore, we present a consistent discretization of the conversion formulas (Appendix C.2) their spherical harmonics equivalent (Appendix C.3). We also discuss in greater detail a number of aspects from Sect. 4.3 in Appendix. C.4, e.g., how the choice of gridding affects the results.

4.2 The conversion

4.2.1 Notation and definitions

We denote the log-normal field at position \vec{x} by $\rho(\vec{x})$ and the underlying Gaussian field by $s(\vec{x})$,

$$\rho(\vec{x}) = \rho_0 e^{s(\vec{x})} = e^{s(\vec{x})+m}, \quad (4.2)$$

where s is dimensionless and the units of ρ as well as a proportionality constant are in ρ_0 and $m \equiv \log \rho_0$. In this work, \log denotes the natural logarithm.

The field s is assumed to follow a Gaussian distribution with zero mean¹,

$$\mathcal{P}(s) = \frac{1}{Z} \exp\left(-\frac{1}{2} \int d^u x d^u y s(\vec{x}) (S^{-1})(\vec{x}, \vec{y}) s(\vec{y})\right), \quad (4.3)$$

with u being the dimensionality of space and $Z = \det(2\pi S)^{\frac{1}{2}}$. $S(\vec{x}, \vec{y})$ is the auto-correlation function of s . It is defined as

$$S(\vec{x}, \vec{y}) := \langle s(\vec{x})s(\vec{y}) \rangle_{\mathcal{P}(s)} \equiv \int \mathcal{D}s \mathcal{P}(s) s(\vec{x})s(\vec{y}). \quad (4.4)$$

Here, $\mathcal{D}s$ denotes integration over the full phase space of s . The inverse of S is defined by the relation

$$\int d^u y S(\vec{x}, \vec{y}) (S^{-1})(\vec{y}, \vec{z}) = \delta_{\mathbb{D}}(\vec{x} - \vec{z}), \quad (4.5)$$

where $\delta_{\mathbb{D}}(\vec{x} - \vec{z})$ is the u -dimensional Dirac delta distribution.

In Fourier space, we denote position vectors by \vec{k} or \vec{q} . Fourier transformed fields are denoted by their argument (e.g., $s(\vec{k})$). We define the Fourier transformations over continuous position space as

$$s(\vec{k}) = \int d^u x e^{i\vec{k}\cdot\vec{x}} s(\vec{x}), \quad s(\vec{x}) = \int \frac{d^u k}{(2\pi)^u} e^{-i\vec{k}\cdot\vec{x}} s(\vec{k}). \quad (4.6)$$

If a field obeys statistical homogeneity its auto-correlation function depends only on the separation, $S(\vec{x}, \vec{y}) = S(\vec{x} - \vec{y})$, and is therefore diagonal in Fourier space,

$$\begin{aligned} S(\vec{k}, \vec{q}) &= \int d^u x d^u y e^{i\vec{k}\cdot\vec{x} - i\vec{q}\cdot\vec{y}} \langle s(\vec{x})s(\vec{y}) \rangle_{\mathcal{P}(s)} \\ &= (2\pi)^u \delta_{\mathbb{D}}(\vec{k} - \vec{q}) P_s(\vec{k}). \end{aligned} \quad (4.7)$$

¹A non-zero mean can be absorbed into m .

The quantity $P_s(\vec{k})$ is the statistical power spectrum of s .

The auto-correlation function and statistical power spectrum of ρ are defined analogously,

$$\begin{aligned} C_\rho(\vec{x}, \vec{y}) &= \langle \rho(\vec{x})\rho(\vec{y}) \rangle_{\mathcal{P}(\rho)} = \rho_0^2 \langle e^{s(\vec{x})} e^{s(\vec{y})} \rangle_{\mathcal{P}(s)}, \\ C_\rho(\vec{k}, \vec{q}) &= (2\pi)^u \delta_{\text{D}}(\vec{k} - \vec{q}) P_\rho(\vec{k}). \end{aligned} \quad (4.8)$$

In the second line we used that ρ obeys statistical homogeneity, too.

4.2.2 Converting the logarithmic power spectrum to the linear power spectrum – the forward conversion

Suppose the statistical power spectrum of s , $P_s(\vec{k})$, and the mean m are known. We are looking for the statistical power spectrum of $\rho(\vec{x}) = e^{s(\vec{x})+m}$. The auto-correlation functions are related by

$$\begin{aligned} C_\rho(\vec{x}, \vec{y}) &= \langle e^{s(\vec{x})+m} e^{s(\vec{y})+m} \rangle_{\mathcal{P}(s)} \\ &= e^{2m + \frac{1}{2}S(\vec{x}, \vec{x}) + \frac{1}{2}S(\vec{y}, \vec{y}) + S(\vec{x}, \vec{y})}, \end{aligned} \quad (4.9)$$

since $\mathcal{P}(s)$ is Gaussian. Using the statistical homogeneity of C_ρ and S , we write

$$C_\rho(\vec{0}, \vec{x}) = e^{2m + S(\vec{0}, \vec{0}) + S(\vec{0}, \vec{x})}. \quad (4.10)$$

This relation between the correlation function of a Gaussian and log-normal field is well known in cosmology (see, e.g., Coles & Jones (1991)). Combined with Eqs. (4.7) and (4.8) this yields

$$\begin{aligned} &\int \frac{d^u k}{(2\pi)^u} e^{-i\vec{k}\cdot\vec{x}} P_\rho(\vec{k}) \\ &= \exp\left(2m + \int \frac{d^u k}{(2\pi)^u} \left(e^{-i\vec{k}\cdot\vec{x}} + 1\right) P_s(\vec{k})\right), \end{aligned} \quad (4.11)$$

and therefore

$$\begin{aligned} P_\rho(\vec{k}) &= \int d^u x e^{i\vec{k}\cdot\vec{x}} e^{2m} \\ &\quad \times \exp\left(\int \frac{d^u q}{(2\pi)^u} \left(e^{-i\vec{q}\cdot\vec{x}} + 1\right) P_s(\vec{q})\right). \end{aligned} \quad (4.12)$$

Eq. (4.12) is our *forward conversion* formula.

If $P_s(\vec{k})$ is isotropic, the equation can be simplified in spherical coordinates, since in that case the argument of the exponential function is isotropic in \vec{x} , too. In the three-

dimensional isotropic case, integration of the angular part yields

$$P_\rho(k) = 4\pi \int_0^\infty dr r^2 \frac{\sin(kr)}{kr} e^{2m} \times \exp\left(\int_0^\infty \frac{dq}{2\pi^2} q^2 \left(\frac{\sin(qr)}{qr} + 1\right) P_s(q)\right), \quad (4.13)$$

where we denoted $P(k) \equiv P(\vec{q})|_{|\vec{q}=k}$. While this equation only has one-dimensional integrals, its numerical evaluation involves some subtleties, which is why we perform the numerical calculations in this work using full-dimensional fast Fourier transforms, even if the power spectrum is isotropic. However, integration of the angular part can be performed for all of the following equations in a very similar manner.

4.2.3 Converting the linear power spectrum to the logarithmic power spectrum – the backward conversion

Suppose the statistical power spectrum of ρ , $P_\rho(k)$, is known and we want to know the power spectrum of s . We therefore need to solve Eq. (4.11) for $P_s(k)$,

$$\begin{aligned} & \log\left(\int \frac{d^u k}{(2\pi)^u} e^{-i\vec{k}\cdot\vec{x}} P_\rho(\vec{k})\right) - 2m \\ &= \int \frac{d^u k}{(2\pi)^u} \left(e^{-i\vec{k}\cdot\vec{x}} + 1\right) P_s(\vec{k}). \end{aligned} \quad (4.14)$$

We invert the operation in front of $P_s(k)$ using

$$\begin{aligned} & \int d^u x \left(e^{i\vec{q}\cdot\vec{x}} - \frac{1}{2}\delta_D(\vec{x}) (2\pi)^u \delta_D(\vec{q})\right) \left(e^{-i\vec{k}\cdot\vec{x}} + 1\right) \\ &= (2\pi)^u \delta_D(\vec{k} - \vec{q}), \end{aligned} \quad (4.15)$$

to arrive at

$$\begin{aligned} P_s(\vec{k}) &= \int d^u x e^{i\vec{k}\cdot\vec{x}} \log\left(\int \frac{d^u q}{(2\pi)^u} e^{-i\vec{q}\cdot\vec{x}} P_\rho(\vec{q})\right) \\ &\quad - m (2\pi)^u \delta_D(k) \\ &\quad - \frac{1}{2} (2\pi)^u \delta_D(k) \log\left(\int \frac{d^u q}{(2\pi)^u} P_\rho(\vec{q})\right). \end{aligned} \quad (4.16)$$

The two correction terms on the right-hand-side do not affect modes with $\vec{k} \neq \vec{0}$,

$$P_s(\vec{k} \neq \vec{0}) = \int d^u x e^{i\vec{k}\cdot\vec{x}} \log\left(\int \frac{d^u q}{(2\pi)^u} e^{-i\vec{q}\cdot\vec{x}} P_\rho(\vec{q})\right), \quad (4.17)$$

and there is a degeneracy between the monopole $P_s(\vec{k}=\vec{0})$ and the mean,

$$P_s(\vec{k}=\vec{0}) + Vm = \int d^u x \log \left(\int \frac{d^u q}{(2\pi)^u} e^{-i\vec{q}\cdot\vec{x}} P_\rho(\vec{q}) \right) - \frac{V}{2} \log \left(\int \frac{d^u q}{(2\pi)^u} P_\rho(\vec{q}) \right), \quad (4.18)$$

where we have identified $V = (2\pi)^u \delta_D(\vec{0})$ as the volume of the system. We recommend to set the monopole to zero to fix the mean. Eqs. (4.17) & (4.18) are our *backward conversion* formulas.

We note that knowledge about the monopole of ρ is crucial for this conversion to work. If the monopole is not supplied, one can estimate it from the mean of ρ as

$$P_\rho(\vec{k}=\vec{0}) = V \langle \rho \rangle^2. \quad (4.19)$$

Using the assumption of statistical homogeneity the backward conversion formula can also be applied in real space to get $S(\vec{x}, \vec{y})$ from $C_\rho(\vec{x}, \vec{y})$,

$$S(\vec{x}, \vec{0}) = \log \left(C_\rho(\vec{x}, \vec{0}) \right) - \frac{1}{2} \log \left(C_\rho(\vec{0}, \vec{0}) \right) - m. \quad (4.20)$$

This relation can (in a different form) also be found in Coles & Jones (1991) (Eq. (30) therein).

4.3 The matter power spectrum

In this section, we use Eqs. (4.12), (4.17), and (4.18) to test the range of validity of a log-normal approximation to the power spectrum of dark matter. In order to apply the formulas on a computer one needs to discretize them correctly, since global prefactors are important in a non-linear transformation such as the exponential and logarithmic functions. We present a consistent discretization in Appendix C.2.

The cosmic matter density ρ is typically parametrized as the mean density ρ_0 and mass density contrast δ according to Eq. (4.1). The three-dimensional isotropic power spectrum of these fluctuations, $P_\delta(k)$, is usually referred to as the matter power spectrum of the large-scale structure. At high redshifts the fluctuations are small, $\delta \ll 1$. With decreasing redshifts, the magnitude of the fluctuations increases. In linear theory, each Fourier mode of the fluctuations is enhanced by the same linear growth factor $D(z)$, yielding a very simple relation between the matter power spectrum at different redshifts,

$$D^2(z') P_\delta(k, z) = D^2(z) P_\delta(k, z'), \quad (4.21)$$

where $P_\delta(k, z)$ denotes the power spectrum of δ at redshift z . The functional form of the linear growth factor is slightly different for different cosmologies. We use the cosmological parameters determined by Planck Collaboration et al. (2014). The prefactor of the power

spectrum is determined by the σ_8 normalization² which is the variance of δ convolved with a spherical top hat function with an $8 \text{ Mpc}/h$ radius. Its value at redshift zero is determined by Planck Collaboration et al. (2014) to be $\sigma_8|_{z=0} = 0.83$. Together with the growth factor this determines the linear power spectrum $P_\delta(k)$ at all redshifts, since in linear theory it retains its shape. However, the linear description of the redshift dependence of the matter power spectrum fails for low redshifts, where the fluctuations can be on the order of 1. This has been successfully treated up to redshift 1 by third order perturbation theory (henceforth 3PT) around the linearly evolved spectrum by Jeong & Komatsu (2006). For redshifts below 1, the non-linearities in the matter power spectrum can be modeled using the Cosmic emulator (henceforth CosmicEmu) based on Heitmann et al. (2009), Heitmann et al. (2010) and Lawrence et al. (2010).

Using the formalism presented in this paper we investigate whether some of the non-linearities resolved by third order perturbation theory and the Cosmic emulator arise naturally if linear growth is applied to the logarithmic density contrast instead of the density contrast itself. A similar question has been investigated by Neyrinck et al. (2009). Using data from the Millennium Simulation by Springel et al. (2005) Neyrinck et al. (2009) compared the power spectrum of $\log(1 + \delta)$ at different redshifts. They found a remarkable reduction of non-linearities in the power spectra up to $k = 1.0 h \text{ Mpc}^{-1}$ at all redshifts. They also found a bias factor for large scales, which is not apparent in our figures. We discuss the origin of this factor in Appendix C.4.4.

Around redshift 7 the 3PT corrections by Jeong & Komatsu (2006) start to become significant. Jeong & Komatsu (2006) find a good agreement between the 3PT spectrum and N-body simulations for wavevectors up to $k = 1.4 h \text{ Mpc}^{-1}$ and redshifts higher or equal to 1. For redshifts 1 to 0 we use matter spectra calculated using CosmicEmu. According to Heitmann et al. (2009), Heitmann et al. (2010) and Lawrence et al. (2010) these are accurate to 1% up to $k = 1.0 h \text{ Mpc}^{-1}$. We therefore assume, that they are also reasonably accurate up to $k = 1.4 h \text{ Mpc}^{-1}$.

We denote the logarithmic density contrast as

$$s = \log(1 + \delta). \quad (4.22)$$

At high redshifts, where $\delta \ll 1$, we have $s \approx \delta$. Therefore, at high redshifts applying the linear growth factor to s has the same effect as applying it to δ . We keep applying the growth factor to s instead of δ throughout the whole redshift spectrum in order to test whether some of the non-linearities appearing at lower redshifts arise naturally this way. Therefore we apply the forward conversion formula (Eq. (4.12)) to the linearly evolved spectrum at redshifts 0 to 7 and the backward conversion formulas (Eqs. (4.17) and (4.18)) to the 3PT spectra and the CosmicEmu spectra respectively.

To that end, we calculate the matter power spectrum at several redshifts between 1 and 7 by applying the 3PT code³ by Jeong & Komatsu (2006) to the linear power spectrum

²This is the normalization of the linear matter power spectrum. If for example the $\sigma_8|_{z=1}$ value is 0.51 for the linear spectrum, it would be 0.55 for the 3PT spectrum.

³<http://www.mpa-garching.mpg.de/~komatsu/CRL/powerspectrum/density3pt/>

calculated using CAMB⁴ (see, e.g., Lewis et al. (2000)) and between 0 and 1 using the CosmicEmu code⁵.

The lowest spectral length covered is $k = 0.004 h \text{Mpc}^{-1}$ for the 3PT code and $k = 0.0075 h \text{Mpc}^{-1}$ for the CosmicEmu code. Therefore, we let our numerical setup cover the region $0.0075 h \text{Mpc}^{-1} \leq k \leq 1.4 h \text{Mpc}^{-1}$. There are several ways to resolve this spectral range and because of the non-linear nature of the exponential function the result of the conversion formulas is not independent of the choice. The smoothing onto a grid of a logarithmic function is thoroughly discussed by Wang et al. (2011). We perform the calculation on four different grids to demonstrate this difference. We present the results of one grid here and discuss the differences between the grids in Appendix C.4.2.

The mean of $(1+\delta)$ is 1, the mean of e^s , however, is higher than 1, since the fluctuations are not symmetric. In order to properly compare the resulting spectra we therefore absorb a factor of $\langle e^s \rangle$ into ρ_0 , i.e., in the following discussion we compare the power spectra of δ and $(e^s/\langle e^s \rangle - 1)$.

4.3.1 The mildly non-linear regime

For redshifts higher or equal to 1, the non-linear corrections (calculated using 3PT) are rather mild. In this regime the model of a linearly evolved log density contrast works rather well, as we depict in Fig. 4.1. For a more quantitative comparison we depict the forward converted spectra divided by the 3PT spectra and the backward converted spectra divided by the linear spectra in Fig. 4.2. For redshifts 1 to 7 and $k \leq 1.0 h \text{Mpc}^{-1}$ the maximal log-distance between the converted spectra and the 3PT spectra stays below 0.17. The distances are slightly lower for the backward conversion. As one can see in Fig. 4.2, the distance is strongest around $k \approx 0.2 h \text{Mpc}^{-1}$. For a full list of the log-distances see Appendix C.4.1. In the region where $1.0 h \text{Mpc}^{-1} < k < 1.4 h \text{Mpc}^{-1}$ the results are susceptible to the choice of gridding. We discuss this in detail in Appendix C.4.2.

4.3.2 The non-linear regime

For redshifts smaller than 1, the non-linear corrections add small-scale power which is comparable to the total linear power. On the grid we chose, the total power of the CosmicEmu spectrum is twice as high as the power of the linear spectrum at redshift 0.8. In this regime the agreement between the converted spectra and the CosmicEmu (or linear spectra, respectively) is much weaker. We depict the non-linearities generated by the forward conversion and the reduction of non-linearities in the backward conversion in Fig. 4.3. A more error focused comparison with the forward converted spectra divided by the CosmicEmu spectra and with the backward converted spectra divided by the linear spectra can be found in Fig. 4.4. The maximal log-distance for $k \leq 1.0 h \text{Mpc}^{-1}$ between the backward converted spectra and the CosmicEmu spectra ranges between 0.17 for $z = 1$ and 0.8 for

⁴<http://camb.info/>

⁵<http://www.lanl.gov/projects/cosmology/CosmicEmu/emu.html>

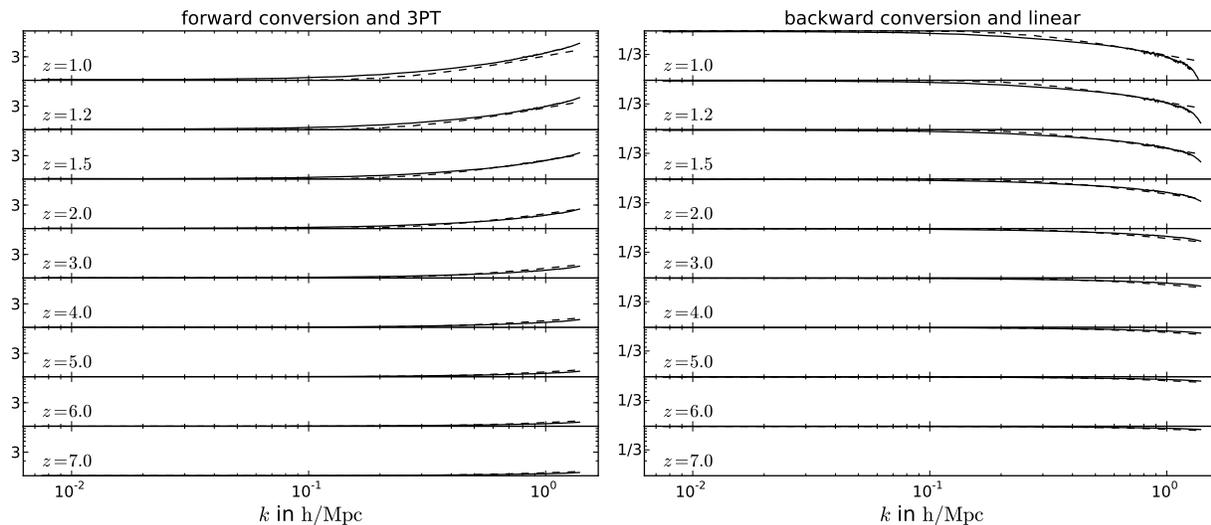


Figure 4.1: The conversion formalism for redshifts 1 to 7. The left panel shows forward converted linear spectra (solid lines) and the corresponding 3PT spectra (dashed lines), which are both divided by the respective linear spectrum for better comparison of the non-linearities. The right panel shows backward converted 3PT spectra (solid lines) and linear spectra (dashed lines), which are both divided by the respective 3PT spectrum.

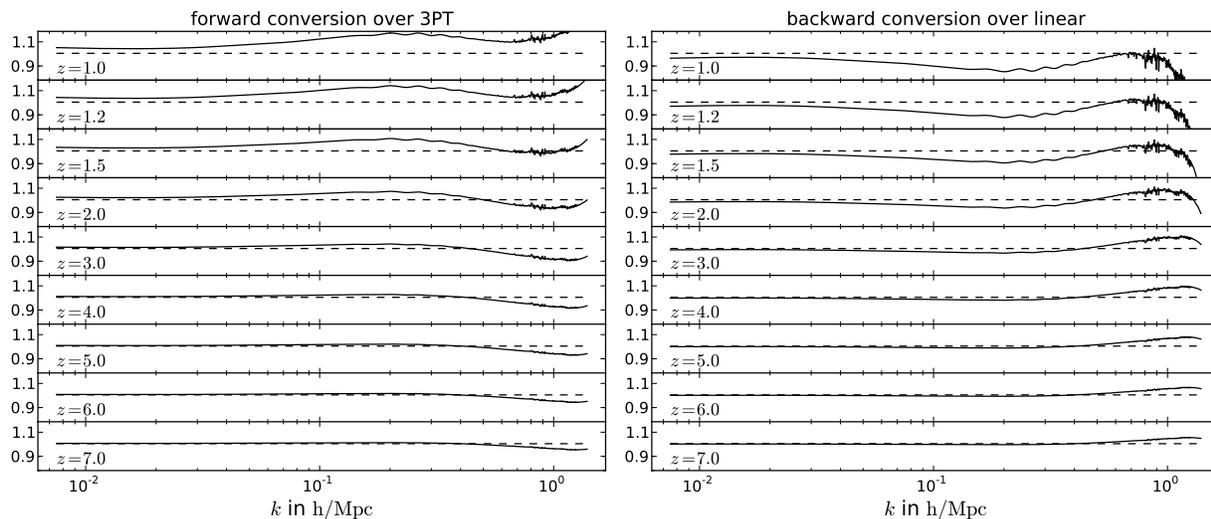


Figure 4.2: The conversion formalism for redshifts 1 to 7. The left panel shows forward converted linear spectra (solid lines) divided by the corresponding 3PT spectra. The right panel shows backward converted 3PT spectra (solid lines) divided by the respective linear spectrum. The dashed line marks 1.

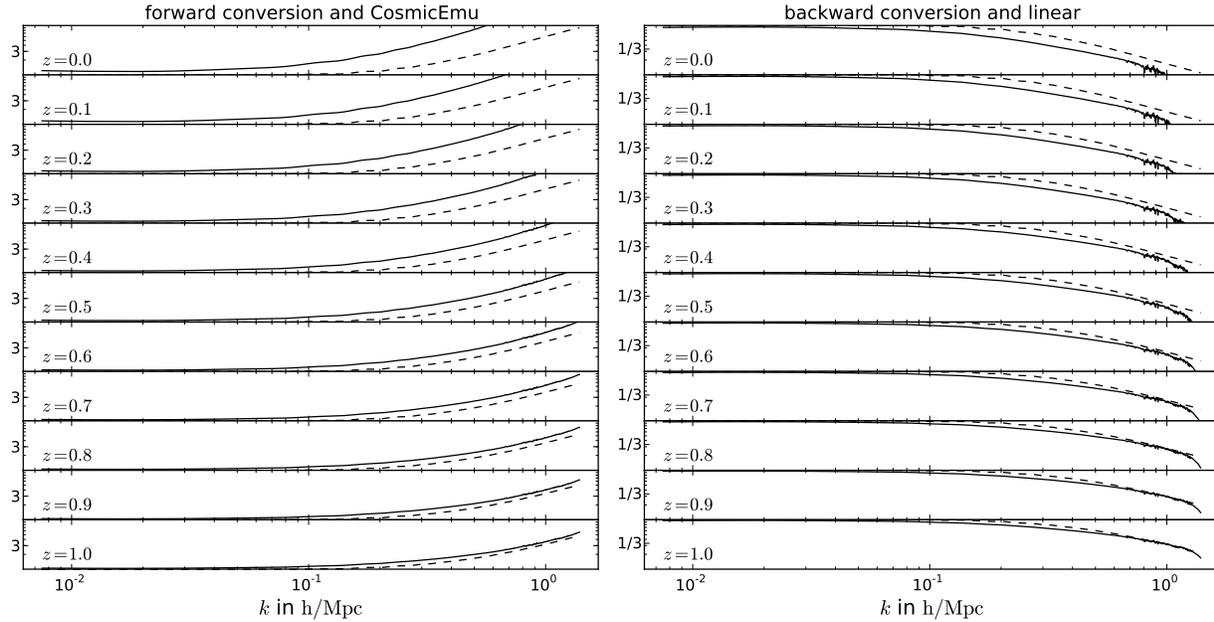


Figure 4.3: The conversion formalism for redshifts 0 to 1. The left panel shows forward converted linear spectra (solid lines) and the corresponding CosmicEmu spectra (dashed lines), which are both divided by the respective linear spectrum for better comparison of the non-linearities. The right panel shows backward converted CosmicEmu spectra (solid lines) and linear spectra (dashed lines), which are both divided by the respective CosmicEmu spectrum.

$z = 0$. For the forward converted spectra it ranges from 0.2 at $z = 1$ to 1.5 at $z = 0$. Clearly, a log-distance of 1.5 (corresponding to a factor of 4.5) is not within an acceptable margin of error. Emulating the non-linear corrections from CosmicEmu by forward conversion breaks down in this regime. The backward conversion performs better than the forward conversion, but a log-distance of 0.8 (corresponding to a factor of 2.2) is far from ideal. However, the backward conversion still reduces the magnitude of non-linear corrections significantly. See Table C.1 in Appendix C.4.1 for a detailed comparison of the log-distances.

4.3.3 Smearing of baryon acoustic oscillations

Another important non-linear effect is the smearing of baryon acoustic oscillations (BAO). Towards lower redshifts the smallest scale wiggles in the matter power spectrum are erased while the large scale wiggles become increasingly damped as redshift decreases. This can be nicely seen in Fig. 4.5, where we depict a comparison between 3PT spectra calculated on the basis of the CAMB power spectrum and 3PT spectra calculated on the basis of

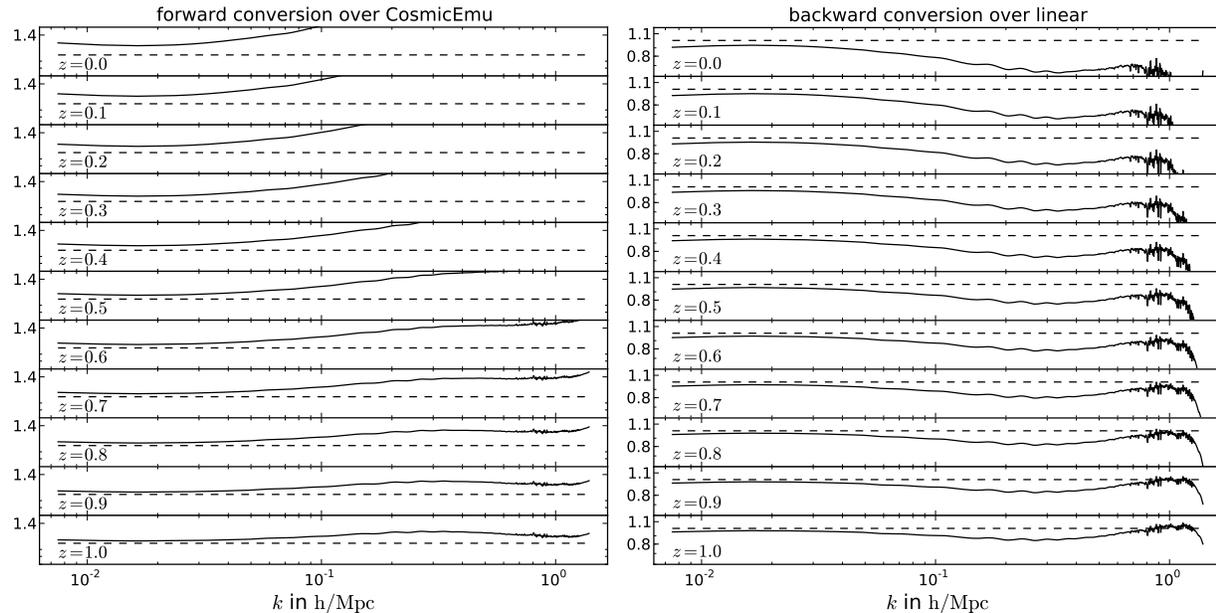


Figure 4.4: The conversion formalism for redshifts 0 to 1. The left panel shows forward converted linear spectra (solid lines) divided by the corresponding CosmicEmu spectra. The right panel shows backward converted CosmicEmu spectra (solid lines) divided by the respective linear spectrum. The dashed line marks 1.

the “no wiggle” power spectrum by Eisenstein & Hu (1998)⁶ in the left panel. The same comparison but using the forward converted linear spectra instead of the 3PT spectra can be seen in the right panel. It is evident that BAO smearing appears only very slightly in the forward converted power spectra, it is much weaker than the smearing calculated using 3PT. BAO smearing appears to be about 3 to 4 times weaker (see Appendix C.4.3) in the forward converted spectra compared to the 3PT spectra, which exhibit a smearing in agreement with N-body simulations.

4.4 Discussion and conclusions

Starting from the assumption of Gaussian statistics for s we presented a formalism to calculate the power spectrum of $s = \log(\rho)$ from the power spectrum of ρ and vice versa. This formalism can for example be used to compare theoretical considerations about logarithmic densities with theoretical or observed power spectra, which are calculated from the linear density. A numerical implementation of these formulas can be found in the NIFTY python package⁷ (Selig et al. (2013)). The formalism is also useful to generate log-normal

⁶The “no wiggle” power spectrum was calculated by putting the cosmological parameters from Planck Collaboration et al. (2014) into the code from <http://www.mpa-garching.mpg.de/~komatsu/CRL/powerspectrum/nowiggle/>.

⁷<http://www.mpa-garching.mpg.de/ift/nifty/>

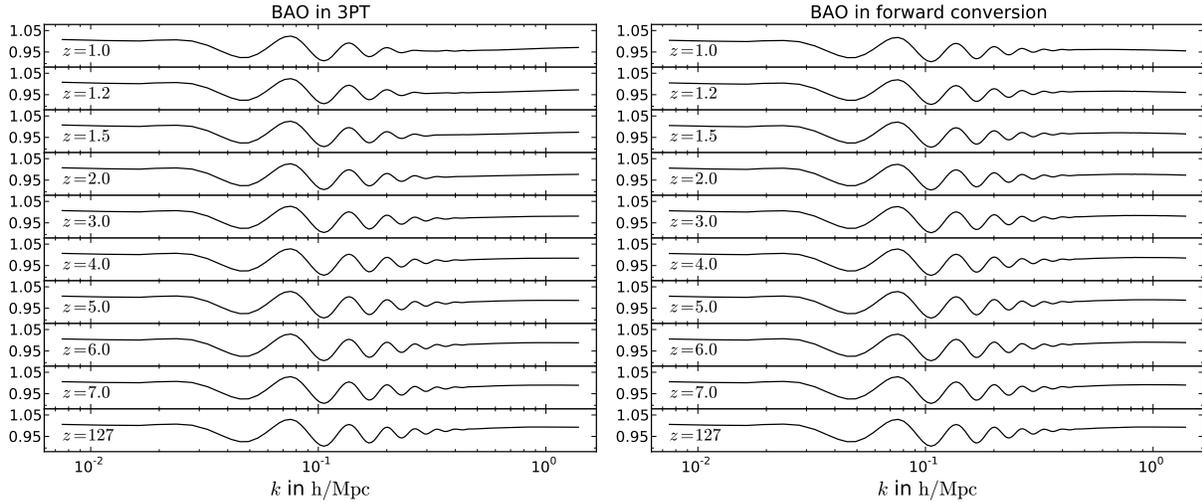


Figure 4.5: BAO smearing. The left panel shows the 3PT spectra with BAO divided by the 3PT spectra without BAO at several redshifts. The right panel shows the forward converted linear spectra with BAO divided by the forward converted linear spectra without BAO. We note the smearing of small-scale wiggles towards lower redshifts in the left panel.

fields with a given power spectrum, as one can simply convert the spectrum, use the resulting spectrum to generate Gaussian random fields, and then exponentiate them (see Appendix C.1).

We used this formalism to investigate whether the non-linear corrections to the matter spectrum are reduced for the logarithmic density and whether the non-linearities can be emulated by applying the growth factor to the log-density instead of the density. To that end we compared emulated linear (CAMB) and non-linear (third-order perturbation theory and CosmicEmu) spectra with the in- and outputs of our formulas. For the mildly non-linear regime (redshift 1-7) we find that the log-transformed non-linear spectra agree with the linear spectra rather well (less than 20% difference between them up to $k = 1.0 h \text{ Mpc}^{-1}$). This enables one to easily generate a log-normal field, which follows the appropriate non-linear power spectrum, even for position dependent redshifts. We describe this procedure in detail in Appendix C.4.5. For lower redshifts the agreement decreases but the non-linearities are still reduced significantly. The smearing of baryon acoustic oscillations can not be emulated by applying the growth factor to the log-density. There is some smearing due to mode coupling, but it is about 3 to 4 times weaker than in the more exact third-order perturbation theory.

We performed the calculation on four different grids leading to different but within the achieved precision comparable results. The differences between the results on different grids indicate that the agreement between our model and the non-linear spectra from literature decreases if the grid allows scales corresponding to $k \geq 1.4 h \text{ Mpc}^{-1}$, since mode coupling to the small scales overestimates the non-linear corrections. This problem arises since the total power of the matter spectrum diverges as its spectral index is above -3 ,

which makes some sort of cut-off in power necessary in a non-linear transformation such as the exponential and logarithmic functions. Here, this cut-off is imposed by the resolution of the grid. Ultimately, one would have to find a physically justifiable way to regularize the integral in Eq. 4.13.

Acknowledgements

The authors would like to thank Niels Oppermann, Philipp Wullstein, Sebastian Dorn, and Marco Selig for fruitful discussions and support. We would also like to thank the anonymous referee for the constructive review and useful comments. All numerical calculations were done using the NIFTY python package by Selig et al. (2013).

Chapter 5

Conclusions and outlook

We have developed two imaging algorithms applicable in the field of radio astronomy. The developed tomographic algorithm was shown to be able to reconstruct the Galactic free electron density using simulated pulsar dispersion measures and distances that were generated according to the expected results from the upcoming Square Kilometre Array (SKA). The degree of details to be expected in an reconstruction depends not only on the number of available pulsar dispersion measures and the accuracy of the measured pulsar distances but also on the dynamic range or contrast of the free electron density. Since the real dynamic range is not known a definitive answer to which level of detail will be reconstructible using our algorithm is prohibited. However, the algorithm has been shown to be capable of three-dimensional tomography with minimal prior assumptions and could be applied to the data available at present. The reason why this has not been done in this thesis was the more promising scientific output of the fastRESOLVE project and the time it demanded.

The next logical step in Galactic tomography would be to apply the tomography algorithm to all pulsar data available at present. Even though the resulting electron density map would lack the detail of models such as the NE2001 model, it would provide an honest assessment of the information contained in pulsar dispersion measures when combined with minimal assumptions. To increase the level of detail one could include more sources of information, such as free-free emission. If the algorithm is extended to a joint reconstruction with the Galactic magnetic field an additional source of information would be the Faraday depths of intra- and extragalactic sources (and other sources connected to the Galactic magnetic field such as synchrotron emission and dust polarization). This would of course be a very challenging project, but also a major leap forward for research of the interstellar medium.

fastRESOLVE, the developed imaging algorithm for aperture synthesis data, was shown to be 100 times faster than its predecessor, RESOLVE. fastRESOLVE also introduced the functionalities of point source separation and Bayesian likelihood variance estimation. The variance estimation technique has not been applied in any aperture synthesis algorithm before and could prove a vital ingredient to Bayesian imaging in the field. fastRESOLVE introduces all of these advantages while loosing only a small amount of precision compared

to RESOLVE. It was shown to be superior to the standard imaging algorithm CLEAN in a number of ways on observational data of the Galaxy cluster Abell 2199.

Currently, fastRESOLVE cannot handle wide-field observations in which the sky cannot be assumed to be flat. To that end, a gridding scheme for the so-called w -term would have to be developed. This would enable fastRESOLVE to produce images from the data of the next generation of radio interferometers such as the SKA. Another desired feature would be the processing of polarization data. Furthermore, a more rigorous way of point source separation should be developed. For all of these extensions there are ongoing research projects building on top of the RESOLVE algorithm. One feature that is already available for RESOLVE is the simultaneous treatment of multiple frequencies. Once all of these research efforts are finished, it is relatively straight-forward to merge these extensions with fastRESOLVE to build a complete and performant imaging package for radio interferometry.

Additionally, we presented a conversion formalism to relate the power spectrum of a positive definite field to the power spectrum of its logarithm. We showed this to be a useful tool in evolving the matter power spectrum with respect to redshift, comparing its precision to third order perturbation theory and the emulator CosmicEmu. We found that while evolution of the matter power spectrum via our conversion method lacks the precision of perturbation theory or CosmicEmu our formula provides an easy way to generate random realizations of the large-scale matter distribution spanning several redshifts. Such random realizations can be very useful for feasibility studies to plan cosmological observations.

The conversion formalism for power spectra could only emulate the evolution of the matter power spectra to a limited degree of precision. However, for a local transformation its performance was impressive. A potential topic of future research would be to find a better local transformation than the exponential/logarithmic mapping used here.

Appendix A

Appendix to tomography of the Galactic free electron density with the Square Kilometer Array

A.1 Parameters of the power spectrum prior

The inverse gamma distribution is defined as

$$\mathcal{I}(p_k; \alpha_k, q_k) = \frac{1}{q_k \Gamma(\alpha_k - 1)} \left(\frac{p_k}{q_k} \right)^{-\alpha_k} \exp\left(-\frac{q_k}{p_k}\right). \quad (\text{A.1})$$

The mean and variance of this distribution are

$$\begin{aligned} \langle p_k \rangle_{(p_k)} &= q_k / (\alpha_k - 2) \quad \text{for } \alpha > 2 \\ \langle p_k^2 \rangle_{(p_k)} - \langle p_k \rangle_{(p_k)}^2 &= \frac{q_k^2}{(\alpha_k - 3)(\alpha_k - 2)^2} \quad \text{for } \alpha > 3. \end{aligned} \quad (\text{A.2})$$

There are three properties of the prior that we wish to fulfill by choosing α_k and q_k . The prior of the monopole p_0 , which corresponds to the variance of a global prefactor of the density, should be close to Jeffreys prior, that is, the limit $\alpha_0 \rightarrow 1$, $q_0 \rightarrow 0$. The reason for this is that we need the algorithm to stay consistent if the units are changed, for instance, from pc to kpc. These changes introduce a global prefactor in front of the density. Jeffreys prior has no preferred scale because it is flat for $\log p_0$ and therefore all prefactors are equally likely a priori. For other k -bins the parameters should favor, but not enforce, falling power spectra. Furthermore, since $p(k)$ is the average power of many independent Fourier components, its a priori variance should be inversely proportional to ϱ_k (the amount of degrees of freedom in the respective k -bin), while the a priori mean should be independent of ϱ_k . We therefore set the parameters as

$$q_k = f \varrho_k \quad \text{and} \quad \alpha_k = 1 + \frac{k}{100k_{\min}} \varrho_k, \quad (\text{A.3})$$

where k_{\min} is the first non-zero k -value and f is a prefactor, which defines a lower cut-off of the power spectrum calculated in Eq. (2.31). The choice of f does not influence the result as long as it is suitably low, but higher f accelerate the convergence of the algorithm. The denominator of $100k_{\min}$ before ϱ_k is chosen to introduce a preference for falling power spectra starting two orders of magnitude from the fundamental mode k_{\min} (we note that 1 is subtracted in Eq. (2.27)). As long as the denominator is not too small, it has very little influence on the result of the algorithm, but smaller denominators increase the convergence speed. We found $100k_{\min}$ to be a good compromise.

The parameter σ_p in Eq. (2.27) describes how much the power spectrum is expected to deviate from a power law. We chose $\sigma_p = 1$. If the power spectrum is locally described by a power law, $\sigma_p = 1$ means that the typical change of the exponent within a factor of e in k should be on the order of 1.

A.2 Functional derivatives of the Hamiltonian

To minimize the Hamiltonian in Sect. 2.2.4, the first derivative with respect to s is needed. It is

$$\frac{\delta}{\delta s^\dagger} \mathcal{H}(s, DM|p, \tilde{N}) = S^{-1}s + \widehat{(e^s)}M(e^s) - \widehat{(e^s)}j, \quad (\text{A.4})$$

where the hat converts a field to a diagonal operator in position space, for example, $\widehat{\xi}(\vec{x}, \vec{y}) = \xi(\vec{x})\delta(\vec{x} - \vec{y})$, and we used the shorthand notations

$$\begin{aligned} S^{-1} &\equiv \sum_k S^{(k)} p_k^{-1} \\ M &\equiv \widehat{\Delta} \tilde{R}^\dagger \tilde{N}^{-1} \tilde{R} \widehat{\Delta} \\ j &\equiv \widehat{\Delta} \tilde{R}^\dagger \tilde{N}^{-1} DM. \end{aligned} \quad (\text{A.5})$$

The second derivative in Eq. (2.49) is

$$\begin{aligned} \frac{\delta^2}{\delta s \delta s^\dagger} \mathcal{H}(s, DM|p, \tilde{N}) &= S^{-1} + \widehat{(e^s)}M\widehat{(e^s)} \\ &\quad + \widehat{(e^s)}\widehat{M}\widehat{(e^s)} - \widehat{(e^s)}\widehat{j}. \end{aligned} \quad (\text{A.6})$$

The last term in the second derivative can be problematic because it can break the positive definiteness¹ of the second derivative, which is crucial for efficiently applying inversion techniques such as the conjugate gradient method. However, a closer inspection of the last two terms (omitting the hats for readability)

$$M(e^s) - j = \widehat{\Delta} \tilde{R}^\dagger \tilde{N}^{-1} \left(\tilde{R} \widehat{\Delta} e^s - DM \right) \propto \widetilde{DM} - DM \quad (\text{A.7})$$

¹Mathematically, the second derivative has to be positive definite at the minimum, but in high dimensional parameter spaces this is not guaranteed in numerical practice.

shows that their contribution is proportional to the difference between the real dispersion data DM and the idealized data generated by the map, $\widetilde{DM} = \widetilde{R}\widehat{\Delta}e^s$. These two terms counteract each other at the minimum, and we therefore omit them to gain numerical stability. Hence the second derivative is approximated as

$$\frac{\delta^2}{\delta s \delta s^\dagger} \mathcal{H}(s, DM|p, \widetilde{N}) \approx S^{-1} + \widehat{(e^s)} M \widehat{(e^s)}. \quad (\text{A.8})$$

A.3 Convergence

In this section, we display the convergence behavior of the power spectrum, effective errors, and profile functions. For the sake of brevity, we limit the discussion to the reconstruction of the Galaxy model with fluctuations of medium strength using the data set of 5000 pulsars with a distance error of 5%. In Appendix A.5 we show the reconstruction of this model and data set using the real power spectrum and profile functions as a benchmark on how well our iterative estimation of them does.

In Fig. A.1 we show the convergence of the power spectrum. As is evident, the power moves away from the initial guess to a fixed point. Compared to the spectrum of the logarithmic input model, the converged spectrum misses power in the large- and small-scale regime. The loss of power in the large-scale regime arises because the profile field absorbs large features, in the small-scale regime it is due to the general loss of small-scale power.

The loss of small-scale power comes from two effects. First, the dispersion measure data sample the density sparse and irregularly. Without the regularization imposed by the prior, this would lead to severe aliasing, as is commonly known from Fourier analysis because the prior typically suppresses aliasing from large to small scales and the algorithm consequently interprets the power as noise. Aliasing from small to large scales is negligible, since the input model is spatially correlated and therefore has a falling power spectrum. Second, some loss of power is due to the distance uncertainties. These make the the likelihood less informative about small-scale structures, which are in consequence suppressed by the prior. This effect yields no aliasing, but smoothens the resulting map (which is desired to avoid overfitting). For further details about the loss of power in filtering algorithms such as the one in this paper we refer to Enßlin & Frommert (2011).

The fixed point power spectrum falls as a power law with index -5.5 for $k > 2 \times 10^{-4}$. Our algorithm allows for spectral indices of up to -5.5 . Without this limit, the power spectrum would fall² to a minimal value of q_k/q_k for $k \gtrsim 3 \times 10^{-4}$. However, introducing a hard limit speeds up the convergence, and using a slope of 5.5 made no difference toward the lower limit for the resulting maps in our tests. This hard limit can be thought of as part of the power spectrum prior.

² This means that the algorithm underestimates the power on scales that are not sufficiently probed by the data set. This does not influence the quality of the density map too much, but it causes the algorithm to underestimate the posterior uncertainty.

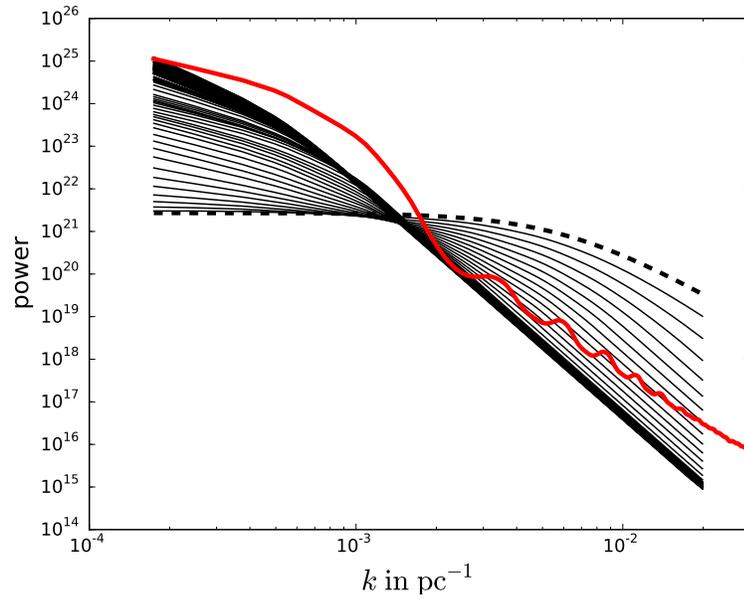


Figure A.1: Plot of the power spectrum changing with the iterations. The thick dashed line is the initial guess, the bulge of black lines is where the algorithm converges. The thick solid red line is the power spectrum of the logarithmic Galaxy model with medium fluctuations. The power spectrum is in arbitrary units.

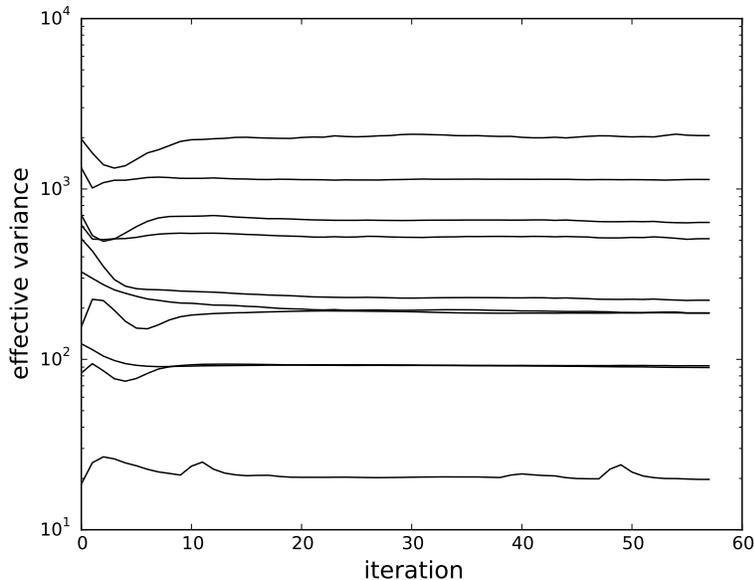


Figure A.2: Propagated distance variance of ten data points changing with the iterations. The units of the variances are $\left(\frac{\text{pc}}{\text{cm}^3}\right)^2$.

In Fig. A.2 we show the convergence of the propagated distance variance (Eq. 2.20) of a random selection of ten data points. The behavior seen in this plot is qualitatively the same for all data points we investigated. Most data points clearly reach convergence rather quickly, but there are also outliers. In this plot, the lowest line exhibits a kink after it had seemingly already converged. This behavior is unfortunately not entirely suppressible, but it appears to have very little effect on the resulting map because only a small fraction of data points shows this behavior.

In Figs. A.3 and A.4 we show the convergence of the profile functions, where we shifted the functions by a global value to line them up at $\beta(|z|=0)$ and $\alpha(r=0)$. We note that the functions α and β are degenerate with respect to a global addition in their effect on the Galactic profile field and degenerate with the monopole of s as well. This is why a shift by a constant for plotting purposes is reasonable. The z -profile function β seems to reach a fixed point for $|z| < 2400$ pc. For higher $|z|$ the profile function reaches no clear fixed point. However, for the Galactic profile, where the profile function is exponentiated, this makes only a small difference, since β is already three e-foldings below its values at $|z| = 0$. The radial profile function α seems to only correct the initial guess mildly, and it is not clear whether the result is independent of the initial guess. However, it appears that α does reach a fixed point.

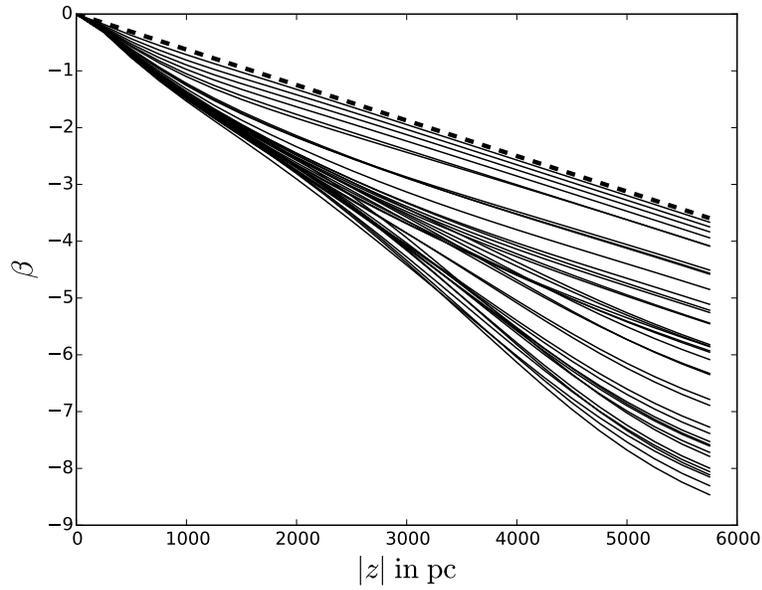


Figure A.3: z -profile function of $\log n_e$ changing with the iterations. The thick dashed line is the initial guess.

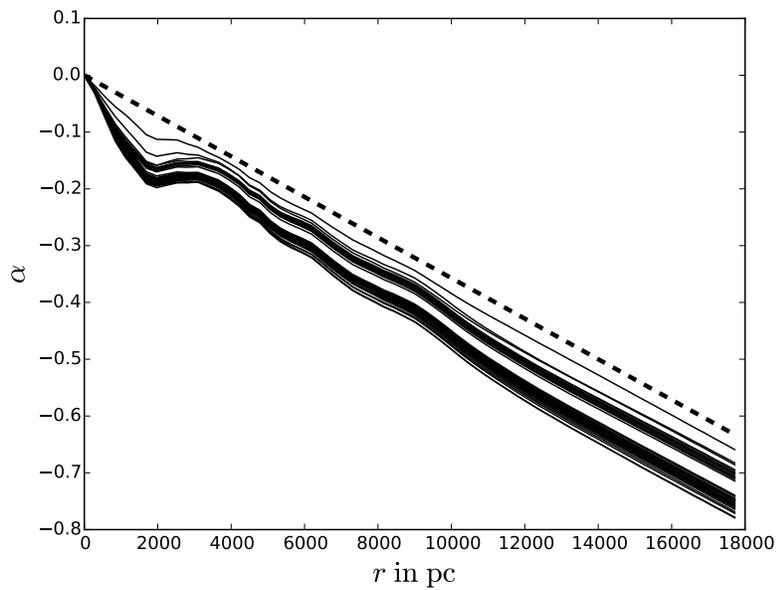


Figure A.4: Radial profile function of $\log n_e$ changing with the iterations. The thick dashed line is the initial guess.

Table A.1: Reduced χ^2 values corresponding to the maps shown in Fig. 2.5.

data set	null model	profile alone	full map
5000 PSR @ 25%	17.0	1.53	0.89
10000 PSR @ 25%	16.9	1.48	0.89
1000 PSR @ 15%	44.4	3.20	1.21
5000 PSR @ 15%	44.7	3.78	1.24
10000 PSR @ 15%	44.6	3.48	1.11
1000 PSR @ 5%	345	23.2	3.34
5000 PSR @ 5%	345	23.5	2.56

A.4 Goodness of fit (χ^2) of the reconstructions

In this section, we discuss the goodness of fit characterized by the reduced χ^2 value,

$$\chi^2 = \frac{1}{N_{\text{data}}} \sum_{i=1}^{N_{\text{data}}} \frac{(DM_i - \widetilde{DM}_i)^2}{\sigma_i^2}, \quad (\text{A.9})$$

where \widetilde{DM} is the dispersion data reproduced by applying the response to our reconstruction,

$$\widetilde{DM}_i = \left(\tilde{R}m^{(\rho)} \right)_i, \quad (\text{A.10})$$

and σ_i is the crudely propagated distance uncertainty given by

$$\sigma_i = \frac{\sqrt{\text{Var}[d_i]}}{d_i} DM_i. \quad (\text{A.11})$$

To set the χ^2 value into perspective, we compared them with the χ^2 values of the null model ($m^{(\rho)} = 0$) and the Galactic profile alone ($m^{(\rho)} = \Delta$). The reduced χ^2 values corresponding to the maps shown in Fig. 2.5 are shown in Table A.1. The reconstruction of the mixed data set has a reduced χ^2 of 1.31. In Table A.2 we show the reduced χ^2 values of the maps shown in Fig. 2.7. For the reconstruction of the contrast-enhanced input model shown in Fig. 2.8 the reduced χ^2 value is 345 for the null model, 110 for the profile alone, and 2.6 for the full reconstruction.

It is evident from the tables that our reconstructions account for a large part of the data variance in all cases. The Galactic profile without local fluctuations also accounts for a large part of the variance, especially if the distance uncertainties are high and the fluctuation strength of the input model is weak. For our reconstructions the χ^2 values are close to 1 for the 25% and 15% data sets. Therefore, we assume that our inference mechanism resolved the most relevant information in the data sets and that the prior assumptions are not too restrictive for these data sets. For the 5% reconstructions the χ^2 values are around 3. This is a hint that the data might contain more information than the reconstruction resolves and that more elaborate prior assumptions might yield a better

Table A.2: χ^2 values corresponding to the maps shown in Fig. 2.7.

data set	null model	profile alone	full map
weak @ 15%	44.7	2.67	1.05
medium @ 15%	44.7	3.78	1.24
strong @ 15%	44.7	6.44	1.54
weak @ 5%	345	16.2	2.34
medium @ 5%	345	23.5	2.56
strong @ 5%	345	47.1	3.01

map. However, how to achieve this is a non-trivial question that we do not aim to answer in this work.

A.5 Reconstruction with real power spectrum and profile functions

The posterior map our algorithm finds depends on the prior power spectrum, the effective errors, and the profile functions, all of which are simultaneously estimated from the data. To benchmark the efficiency of this joint estimation, we investigated the case where the real power spectrum and the real profile functions are known, that is, only iterating the effective errors. The resulting map serves as an indicator whether our ansatz with unknown hyper parameters is sensible or if the problem is too unrestricted in this setting. We depict the map resulting from the real hyper parameters in Fig. A.5. The morphology of the result obviously does not change. More of the small-scale structure is resolved, and the intensity of the overdensity between the Sun and the Galactic center, which belongs to the ring the original model, is more pronounced. Consequently, this map has a better reduced χ^2 value of 1.95 compared to the value of 2.56 of our map with unknown hyperparameters. But considering the amount of unknowns, this is a satisfactory result. We therefore regard our estimation procedure for the hyper parameters as sensible.

A.6 Uncertainty map

Here, we discuss the 1σ uncertainty map of reconstructing the Galaxy model with medium fluctuations using the data set with 5000 pulsars and 5% distance uncertainty. We compare the uncertainty map for unknown profile and power spectrum and the uncertainty map for a known profile and power spectrum (see Appendix A.5) with the corresponding absolute errors. The density maps are shown in Fig. A.5. The uncertainty estimates $\sigma^{(\rho)}$ (see Eq. (2.55)) are shown in Fig. A.6. These uncertainties are underestimated because they are calculated from the curvature of the negative log-posterior around its minimum (see

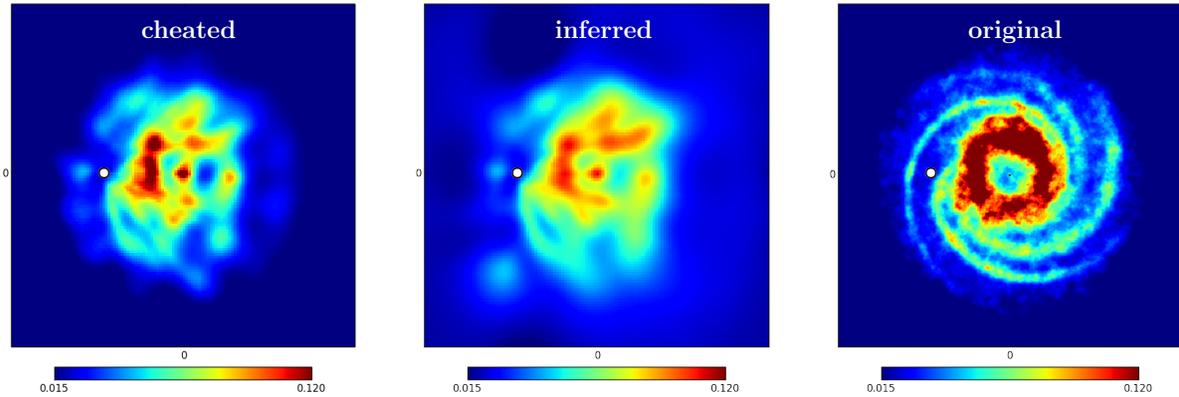


Figure A.5: Top-down view of the reconstructed electron densities in the Galactic plane (in units of cm^{-3}), if we use the real power spectrum and Galactic profile functions (“cheated”), the results from our algorithm (“inferred”), and the original input model.

Eq. (2.35)), not from the full distribution. By visual comparison with the absolute error³, $|m^{(\rho)} - \rho|$, we estimate that the uncertainty estimates are underestimated by a factor of roughly 3. However, their morphology seems to be reliable.

A.7 Uncertainties of the vertical fall-off

In principle, the posterior variance of α and β is the diagonal of the operator $D_{(\xi)}$ (see Eq. (2.45)). However, this diagonal is too large because α and β are completely degenerate with respect to a constant shift ($\alpha + c$ and $\beta - c$ yield the same profile as α and β). This degeneracy yields a large point variance, which is not instructive for quantifying the uncertainty of the vertical fall-off. Therefore, we projected out the eigenvector corresponding to the constant shift before calculating the diagonal of $D_{(\xi)}$. This corrected diagonal is the squared 1σ uncertainty that we plot in Fig. 2.10.

³ To calculate the absolute error, the original Galaxy model is downsampled to the resolution the algorithm uses.

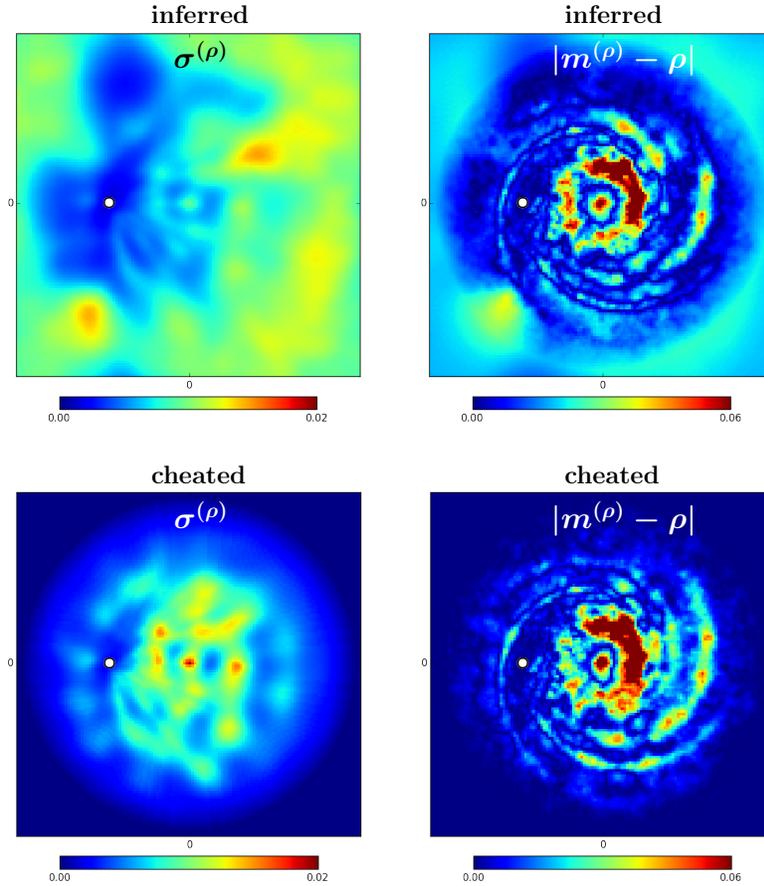


Figure A.6: Top-down view on the Galactic plane, showing the uncertainty estimate ($\sigma^{(\rho)}$, left panels) and the absolute error ($|m^{(\rho)} - \rho|$, right panels) for our reconstruction $m^{(\rho)}$ in units of cm^{-3} . The input density ρ has fluctuations of medium strength and is sampled by 5000 pulsars with 5% distance uncertainty. The top row shows the scenario with unknown power spectrum and Galactic profile, the bottom row shows the scenario with known power spectrum and profile. We note that the left and right panels have different color bars.

Appendix B

Appendix to fastRESOLVE: fast Bayesian imaging for aperture synthesis in radio astronomy

B.1 Aliasing

The visibilities are calculated by a Fourier transform of the product between intensity and primary beam Eq. (3.1). In practice we are working on a pixelized grid where the integrals become sums,

$$V(u, v) = \sum_{x \in X} \Delta x e^{-2\pi i u x} \sum_{y \in Y} \Delta y e^{-2\pi i v y} B(x, y) I(x, y), \quad (\text{B.1})$$

with

$$\begin{aligned} X &= \{j\Delta x \mid -N_x/2 \leq j < N_x/2\}, \\ Y &= \{j\Delta y \mid -N_y/2 \leq j < N_y/2\}. \end{aligned} \quad (\text{B.2})$$

Inserting a discrete Fourier transformation and inverse transformation before $B(x, y) I(x, y)$ yields

$$\begin{aligned} V(u, v) &= \sum_{x \in X} \Delta x e^{-2\pi i u x} \sum_{k \in K} \Delta k e^{2\pi i k x} \\ &\quad \sum_{y \in Y} \Delta y e^{-2\pi i v y} \sum_{q \in Q} \Delta q e^{2\pi i q y} \tilde{I}(k, q) \end{aligned} \quad (\text{B.3})$$

with

$$\tilde{I}(k, q) = \sum_{x \in X} \Delta x \sum_{y \in Y} \Delta y e^{-2\pi i (kx + qy)} B(x, y) I(x, y) \quad (\text{B.4})$$

and

$$\begin{aligned} K &= \{j\Delta k \mid -N_x/2 \leq j < N_x/2\}, \quad \Delta k = \frac{1}{N_x\Delta x}, \\ Q &= \{j\Delta q \mid -N_y/2 \leq j < N_y/2\}, \quad \Delta q = \frac{1}{N_y\Delta y}. \end{aligned} \tag{B.5}$$

Summing over x and y first makes this

$$V(u, v) = \sum_{k \in K} \Delta k \sum_{q \in Q} \Delta q G(u, v; k, q) \tilde{I}(k, q), \tag{B.6}$$

with the gridding operator

$$\begin{aligned} G(u, v; k, q) &= \Delta x \frac{2i \sin(\pi(k-u)/\Delta k)}{e^{2i\pi(k-u)/(N_x\Delta k)} - 1} \times \\ &\quad \Delta y \frac{2i \sin(\pi(q-v)/\Delta q)}{e^{2i\pi(q-v)/(N_y\Delta q)} - 1}. \end{aligned} \tag{B.7}$$

To emphasize the difference between the unstructured set of uv -points and the structured set of k and q points, we shorten the notation of G to

$$G_j(k, q) \equiv G(u_j, v_j; k, q). \tag{B.8}$$

If N_x and N_y are large, G can be approximated as

$$G_j(k, q) \approx \frac{1}{\Delta k \Delta q} \text{sinc}\left(\frac{k-u_j}{\Delta k}\right) \text{sinc}\left(\frac{q-v_j}{\Delta q}\right), \tag{B.9}$$

with $\text{sinc}(x) := \sin(\pi x)/(\pi x)$. In this representation it is easy to see, that $G_j(k, q)$ falls off with the inverse distance from the maximal pixel in each direction. This behavior is independent of the pixelization as the pixel edge lengths are divided out in the sinc terms. To project the visibilities $V_j \equiv V(u_j, v_j)$ into the regular space the adjoint of G has to be applied:

$$V(k, q) = \sum_j G_j^*(k, q) V_j. \tag{B.10}$$

The inverse (fast) Fourier transform of $V(k, q)$ yields the same result as the sum of the inverse (direct) Fourier transform of V_j .

B.2 Kullback-Leibler Divergence exremization

In this section we derive the information theoretically optimal approximation of a Gaussian likelihood with linear measurement by a diagonal measurement precision operator. The likelihood we want to approximate is of the form

$$\mathcal{P}(d|s) = \frac{1}{|2\pi N|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} (d - Rs)^\dagger N^{-1} (d - Rs)\right). \tag{B.11}$$

Combined with a Gaussian prior it yields the posterior

$$\begin{aligned} \mathcal{P}(s|d) &= \left| 2\pi (S^{-1} + R^\dagger N^{-1} R)^{-1} \right|^{-\frac{1}{2}} \\ &\times e^{-\frac{1}{2}(s - R^\dagger N^{-1} d)^\dagger (S^{-1} + R^\dagger N^{-1} R) (s - R^\dagger N^{-1} d)}. \end{aligned} \quad (\text{B.12})$$

The likelihood enters the posterior via two quantities: the information source $j = R^\dagger N^{-1} d$ and the measurement precision $M = R^\dagger N^{-1} R$. Therefore, a Gaussian likelihood with a linear response always yields a posterior of the form

$$\mathcal{P}(s|d) = \left| 2\pi (S^{-1} + M)^{-1} \right|^{-\frac{1}{2}} e^{-\frac{1}{2}(s-j)^\dagger (S^{-1} + M) (s-j)}. \quad (\text{B.13})$$

We want to find a diagonal approximation \tilde{M} to the measurement precision. The information loss by replacing M with \tilde{M} is to be minimal. It is quantified by the Kullback-Leibler divergence,

$$D_{\text{KL}}[\tilde{\mathcal{P}}(s|d)|\mathcal{P}(s|d)] = \int \mathcal{D}s \tilde{\mathcal{P}}(s|d) \ln \frac{\tilde{\mathcal{P}}(s|d)}{\mathcal{P}(s|d)}. \quad (\text{B.14})$$

For now we also introduce an approximate \tilde{j} to see if this implies corrections to \tilde{M} . We will equate it with j later. Dropping all terms independent of \tilde{M} (denoted by $\hat{=}$) the Kullback-Leibler divergence is

$$\begin{aligned} D_{\text{KL}} \hat{=} & \frac{1}{2} \text{tr} [R^\dagger N^{-1} R (D \tilde{j} \tilde{j}^\dagger D + D)] - \frac{1}{2} d N^{-1} R D \tilde{j} \\ & - \frac{1}{2} \tilde{j}^\dagger D R^\dagger N^{-1} d - \frac{1}{2} \text{tr} [\tilde{M} (D \tilde{j} \tilde{j}^\dagger D + D)] \\ & + \frac{1}{2} \tilde{j}^\dagger D \tilde{j} + \frac{1}{2} \text{tr} [\ln D^{-1}], \end{aligned} \quad (\text{B.15})$$

where $D = (S^{-1} + \tilde{M})^{-1}$. By setting $S^{-1} \rightarrow 0$ (corresponding to a flat prior) we remove the influence of the prior. Since \tilde{M} is chosen to be diagonal, $\tilde{M}_{kq} = \delta(k - q) g_q$, we have $D^{-1} = \delta(k - q) g_q$ and $D = \delta(k - q) g_q^{-1}$. This simplifies the Kullback-Leibler divergence further,

$$\begin{aligned} D_{\text{KL}} \hat{=} & \frac{1}{2} \int dq \left[(R^\dagger N^{-1} R)_{qq} (|\tilde{j}_q|^2 / g_q^2 + g_q^{-1}) \right] \\ & - \int dq \Re \left[(R^\dagger N^{-1} d)_q^* (\tilde{j}_q / g_q) \right] \\ & - \frac{1}{2} \int dq \left[g_q (|\tilde{j}_q|^2 / g_q^2 + g_q^{-1}) \right] \\ & + \frac{1}{2} \int dq \frac{|\tilde{j}_q|^2}{g_q} + \frac{1}{2} \int dq \ln g_q, \end{aligned} \quad (\text{B.16})$$

where $\Re(\cdot)$ denotes the real part of a complex value. Its functional derivative with respect to g is

$$\begin{aligned} \frac{\delta}{\delta g_q} D_{KL} = & -\frac{1}{2} \left[(R^\dagger N^{-1} R)_{qq} (2|\tilde{j}_q^2|/g_q^3 + g_q^{-2}) \right] \\ & + \Re \left[(R^\dagger N^{-1} d)_q^* (\tilde{j}_q/g_q^2) \right] \\ & + \frac{1}{2} |\tilde{j}_q^2|/g_q^2 - \frac{1}{2} |\tilde{j}_q^2|/g_q^2 + \frac{1}{2g_q}. \end{aligned} \quad (\text{B.17})$$

Setting this derivative to zero and solving for g yields

$$\begin{aligned} g_q = & \frac{1}{2} (R^\dagger N^{-1} R)_{qq} - \Re \left[(R^\dagger N^{-1} d)_q^* \tilde{j}_q \right] \\ & \pm \left\{ \frac{1}{4} (R^\dagger N^{-1} R)_{qq}^2 + \Re \left[(R^\dagger N^{-1} d)_q^* \tilde{j}_q \right]^2 \right. \\ & \left. - (R^\dagger N^{-1} R)_{qq} \Re \left[(R^\dagger N^{-1} d)_q^* \tilde{j}_q \right] + 2 (R^\dagger N^{-1} R)_{qq} |\tilde{j}_q^2| \right\}^{\frac{1}{2}}. \end{aligned} \quad (\text{B.18})$$

Finally, by setting¹ $\tilde{j} = j = R^\dagger N^{-1} d$ we arrive at

$$g_q = \frac{1}{2} (R^\dagger N^{-1} R)_{qq} - |j_q^2| \pm \left[\frac{1}{2} (R^\dagger N^{-1} R)_{qq} + |j_q^2| \right] \quad (\text{B.19})$$

where the $+$ corresponds to minimal information loss.

The final result of this derivation is therefore that the measurement operator of Gaussian likelihood with linear response can be approximated by its diagonal,

$$\tilde{M}_{kq} = \delta(k - q) (R^\dagger N^{-1} R)_{qq}. \quad (\text{B.20})$$

For the likelihood in this paper we choose to approximate the measurement precision operator after factoring out the primary beam², thus setting

$$\tilde{F}(k, q; k', q') = \delta(k - k') \delta(q - q') F(k, q; k, q). \quad (\text{B.21})$$

B.3 Approximate propagator

The filter equation (3.24) involves the evaluation of the propagator D . It is defined implicitly via its inverse (Junklewitz et al. 2016, Eq. (28)),

$$\begin{aligned} D^{-1}(\vec{x}, \vec{x}') = & S^{-1}(\vec{x}, \vec{x}') + e^{m(\vec{x})} M(\vec{x}, \vec{x}') e^{m(\vec{x}')} \\ & + \delta(\vec{x} - \vec{x}') e^{m(\vec{x})} (M e^m - j)(\vec{x}). \end{aligned} \quad (\text{B.22})$$

¹ We could also minimize D_{KL} with respect to \tilde{j} . This would yield $\tilde{j} = \tilde{M} M^{-1} j$. In our tests the inversion of M was numerically problematic and we achieved better results using $\tilde{j} = j$.

²This corresponds to interpreting the primary beam as part of the signal for the sake of the derivation in this Appendix.

Calculating the trace of D is numerically expensive since a numerical inversion method like the conjugate gradient method has to be applied several times. The conjugate gradient method is efficient in inverting symmetric positive definite operators. Mathematically, the second derivative has to fulfill both conditions at the minimum. In numerical practice positive definiteness can be violated by the $(Me^m - j)$ term. This term is proportional to the difference between real data and reconstructed data. Setting it to zero is therefore a reasonable approximation that ensures positive definiteness,

$$D^{-1}(\vec{x}, \vec{x}') \approx S^{-1}(\vec{x}, \vec{x}') + e^{m(\vec{x})} M(\vec{x}, \vec{x}') e^{m(\vec{x}')}. \quad (\text{B.23})$$

This form of D is used to calculate the uncertainty map described in Sec. 3.2.4.

For the power spectrum estimation in Eq. (3.24) we can approximate D even further to avoid the expensive numerical inversion altogether:

$$D^{-1}(\vec{x}, \vec{x}') \approx S^{-1}(\vec{x}, \vec{x}') + \left\langle \hat{B}e^m \right\rangle_{(\vec{x})} \tilde{F}(\vec{x}, \vec{x}') \left\langle \hat{B}e^m \right\rangle_{(\vec{x})}, \quad (\text{B.24})$$

with $\langle \cdot \rangle_{(\vec{x})}$ denoting the spatial average,

$$\left\langle \hat{B}e^m \right\rangle_{\vec{x}} := \int d\vec{x} B(\vec{x}) e^{m(\vec{x})} / \int d\vec{x} 1. \quad (\text{B.25})$$

This is of course a strong approximation to Eq. (B.22), but in numerical tests Eq. (3.24) still provided the desired fixed point. The advantage is that since both, S and \tilde{F} are diagonal in Fourier space, D^{-1} is diagonal as well and the inversion is trivial.

B.4 Likelihood variance estimation

In this section we show how the variances of a Gaussian likelihood can be estimated. To that end we follow the derivation of Oppermann et al. (2011). We assume that the noise is signal independent and that the diagonal basis of the noise covariance matrix is known,

$$\begin{aligned} d &= R(s) + n, \\ \mathcal{P}(n|s) &= \mathcal{P}(n) = \mathcal{G}(n, N), \\ N_{ij} &= \delta_{ij} \sigma_i^2. \end{aligned} \quad (\text{B.26})$$

The diagonal entries σ_i^2 are unknown. Since they are variances, they have to be positive. Therefore, a convenient choice for their prior is an inverse Gamma distribution,

$$\mathcal{P}(\sigma^2) = \prod_i \frac{1}{r_i \Gamma(\beta_i - 1)} \left(\frac{\sigma_i^2}{r_i} \right)^{-\beta_i} \exp\left(-\frac{r_i}{\sigma_i^2} \right). \quad (\text{B.27})$$

The inverse Gamma distribution is defined by two parameters, r and β . We will discuss canonical choices for them later on. It is furthermore beneficial to parametrize σ_i^2 by its logarithm,

$$\begin{aligned}\eta_i &:= \ln(\sigma_i^2), \\ \mathcal{P}(\eta) &= \mathcal{P}(\sigma^2)|_{\sigma^2=e^\eta} \prod_i e^{\eta_i}\end{aligned}\tag{B.28}$$

An estimate for η can now be derived using a joint maximum a posteriori Ansatz, i.e. we maximize the joint probability of η and s given the data. The joint maximum yields the same estimate for s as if η was known and the estimate for $\sigma^2 \equiv e^\eta$ is

$$\sigma_i^2 = \frac{r_i + \frac{1}{2} |(R(s) - d)_i|^2}{\beta_i - \frac{1}{2}}.\tag{B.29}$$

The prior for σ^2 can be made uninformative by choosing $r_i \rightarrow 0$ and $\beta_i \rightarrow 1$. In this limit the inverse Gamma prior becomes Jeffrey's prior, which is flat in η ,

$$\mathcal{P}(\eta)|_{r=0, \beta=1} = \text{const.}\tag{B.30}$$

Under this choice the estimate for σ^2 simplifies further to

$$\sigma_i^2 = |(R(s) - d)_i|^2.\tag{B.31}$$

This means that the σ^2 are chosen to set the χ^2 value of each data point to 1. To avoid the singularity at $(R(s) - d)_i = 0$, we regularize σ^2 by

$$\sigma_i^2 \rightarrow t \sigma_i^2 + (1 - t) \frac{1}{N_{\text{data}}} \sum_j \sigma_j^2,\tag{B.32}$$

with $t \in [0, 1]$. In practice we choose 0 for well-calibrated data sets and $t = 0.9$ otherwise.

B.5 Point source procedure

The point source contribution I_{point} is determined by iteratively collecting the brightest pixels of the maximum a posteriori map and subtracting their contribution from the data. The iterative procedure is stopped when the amount of pixels containing a point source reaches a cut-off criterion,

$$\sum_{\vec{x}} \Theta(I_{\text{point}}(\vec{x}) > 0) > N_{\text{point}},\tag{B.33}$$

where Θ is the indicator function,

$$\Theta(I_{\text{point}}(\vec{x}) > 0) = \begin{cases} 1 & \text{if } I_{\text{point}}(\vec{x}) > 0 \\ 0 & \text{else.} \end{cases}\tag{B.34}$$

The cut-off criterion N_{point} should be chosen according to the expected amount of point sources in the image and the width of the point spread function. We are not aware of a canonical way to choose it.

Assuming a power spectrum p and noise variances σ_i^2 start with an initial I_{point} (e.g. 0):

1. calculate an estimate for $I = e^m$ by solving Eq. (3.22)
2. identify the maximal intensity $I_{\text{max}} = \max_{\vec{x}} e^{m(\vec{x})}$
3. point sources locations are $\{\vec{x} \mid I(\vec{x}) > I_{\text{max}}/t_{\text{point}}\}$
4. at every point source location I_{point} is increased by $I(\vec{x}) \times (1 - 1/t)$
5. if the cut-off criterion Eq. (B.33) is reached undo point 4 and skip point 6
6. change the data to $d \rightarrow d - R(I_{\text{point}})$ and repeat from point 1

The regularization factor t_{point} has to be larger than 1, but should ideally be small. A good compromise is 1.75 in our experience.

B.6 Simulation tests

B.6.1 Diffuse emission

In this section we validate the behavior of fastRESOLVE using simulated signal and data. The simulated signal is the same as in the original publication of RESOLVE (Junklewitz et al. 2016, Sec. 3). This allows for comparing the performance of these two RESOLVE variants. As can be seen in Fig. B.1 the results using fastRESOLVE are not quite as good as the ones using RESOLVE, but recover the most significant structure in a comparable quality. This is due to the approximations made in the derivation of fastRESOLVE. However, fastRESOLVE detected automatically that there were no point sources present. Furthermore, RESOLVE was provided with the measurement uncertainties while fastRESOLVE had to reconstruct them from the data. We can report that the reconstructed measurement uncertainty is mostly overestimated and reaches an upper limit even if the data are noise-free (using $t = 0$). This leads to a more conservative final image.

The RESOLVE runs used relatively loose convergence criteria and took roughly 1.9 hours each using two cores of an AMD Opteron 6376 @ 2.3Ghz. The fastRESOLVE runs took 2.8 minutes each using relatively strict convergence criteria and running on a single core of an Intel Xeon E5-2650 v3 @ 2.3GHz. With loose convergence criteria, the run-time of fastRESOLVE dropped to 1.1 minutes. Even though it is not straight-forward to compare the convergence criteria of Resolve and fastResolve (as mentioned in Sec. 3.3.1), this leads us to confirm the speed-up factor of roughly 100 we found in Sec. 3.3.1.

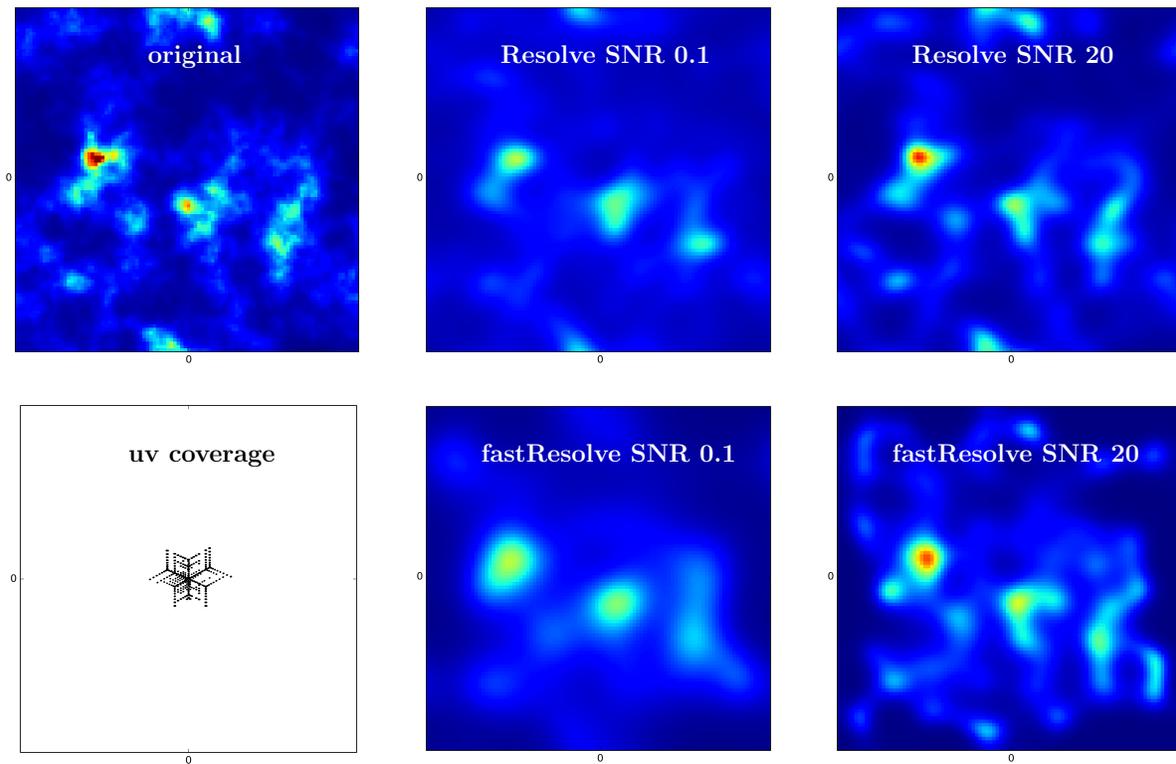


Figure B.1: Simulation results for purely diffuse emission. All panels apart from the uv coverage use the same color scheme. The panel size of uv coverage corresponds to the Fourier scales populated by the mock signal.

B.6.2 point-like emission

To test the performance of fastRESOLVE in the presence of point sources, we add three point sources to the simulated diffuse emission used in Appendix B.6.1. The strongest peak of diffuse emission is at 14.9Jy/px in this test. The three point sources are at 5700Jy/px, 1000Jy/px, and 770Jy/px covering exactly one pixel at the cellsize of the simulation (10^{-6} rad). The total flux of is 1.3×10^4 Jy for the diffuse component and 7.5×10^3 Jy for the point-like emission. In Fig. B.2 we show the simulated signal as well as the fastRESOLVE reconstructions at $SNR = 0.1$ and $SNR = 20$. The signal-to-noise value is with respect to the diffuse emission here. We use exactly the same noise contributions as in Appendix B.6.1. As can be seen from the figure, the point-like emission is smoothed due to the point-spread-function, but some deconvolution is performed depending on the signal-to-noise ratio. The structure of the diffuse emission can be recovered in a comparable quality compared to the purely diffuse test in Appendix B.6.1 for this strength of point-like emission. We performed another simulation in which the brightness and total flux of the point-sources were 10 times greater (keeping all other parameters fixed). In that case only the most dominant features of the diffuse flux could be reconstructed. Most of it was obscured by the point sources. However, as long as the total flux of the point-like emission was smaller than or comparable to the diffuse flux, all our tests showed satisfactory results.

B.7 Signal to noise

In this paper we define the signal-to-noise ratio as

$$SNR = \frac{1}{N_{\text{data}}} \sum_i \frac{|(RI)_i|^2}{\sigma_i^2}. \quad (\text{B.35})$$

There are other definitions of signal-to-noise, we chose this one for its simplicity. It can be estimated by inserting the reconstructed I or directly from the data as

$$SNR \approx \frac{1}{N_{\text{data}}} \sum_i \frac{|d_i|^2}{\sigma_i^2} - 1. \quad (\text{B.36})$$

Due to the way fastRESOLVE estimates the likelihood variances σ_i^2 both estimates typically agree to a reasonable precision. By inserting the corresponding components into Eq. (B.35) we can provide SNR estimates for the point-like and diffuse flux separately.

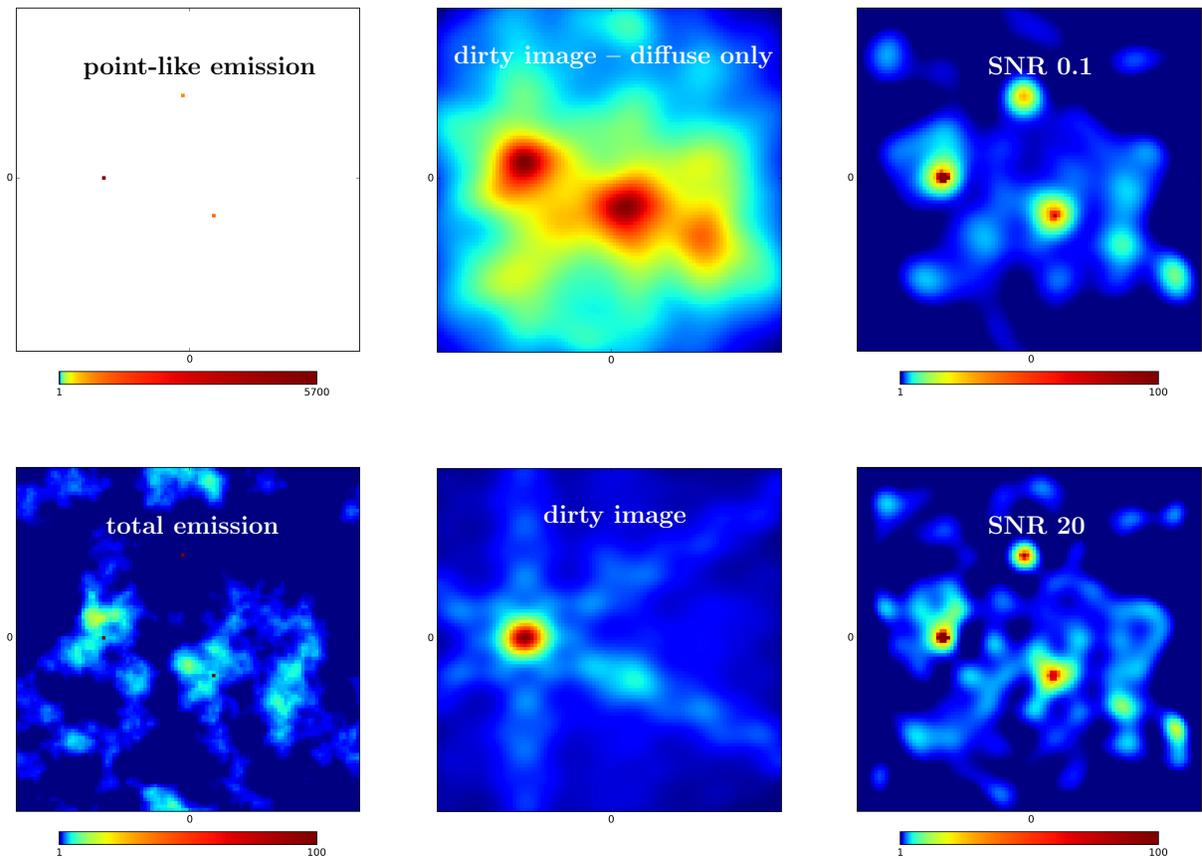


Figure B.2: Simulation results for diffuse and point-like emission. The diffuse component and the uv-coverage were the same as in Fig. B.1. Left panels show the simulated emission, middle panels show the dirty images of the noiseless diffuse emission (top) and noiseless total emission (bottom), right panels show the fastRESOLVE reconstructions.

Appendix C

Appendix to log-transforming the matter power spectrum

C.1 Generating log-normal fields from a given power spectrum

We can use the backward conversion (Eqs. (4.17) & (4.18)) to generate log-normal random fields that follow a given linear power spectrum. This method has already been presented by Percival et al. (2004) (see Sect. 3.2 therein). For illustrative purposes we demonstrate the procedure for an isotropic case, but the same procedure can be applied for an anisotropic power spectrum.

Suppose the power spectrum $P_\rho(k)$ is known within some range $k_{\min} < k < k_{\max}$ and the mean $\langle \rho \rangle$ is also known. First, we set up a discretized space in which k_{\min} corresponds to the minimal non-zero mode and k_{\max} is the highest supported mode. We construct the monopole according to Eq. (4.19) and apply Eqs. (4.17) and (4.18) to the power spectrum. Setting $P_s(k=0) = 0$ in Eq. (4.18) we end up with a power spectrum $P_s(k)$ and a mean m . We can now construct log-normal fields by generating Gaussian random fields from $P_s(k)$ and exponentiating the sum of s and m . The resulting field $\rho = e^{s+m}$ follows (in the statistical average) the power spectrum $P_\rho(k)$.

We illustrate this in Fig. C.1. Starting with the power spectrum of ρ and $\langle \rho \rangle = 1$, we calculate the power spectrum and mean of s . The mean is -0.76 . With these quantities we can draw log-normal fields that follow the original power spectrum. One random field of that kind is plotted in Fig. C.2.

C.2 The discretized conversion formula

The discretized conversion formulas are implemented in the NIFTY python package by Selig et al. (2013) (version 0.8.4 or higher).

In discretizing the forward and backward conversion formulas from Sect. 4.2.2 and 4.2.3

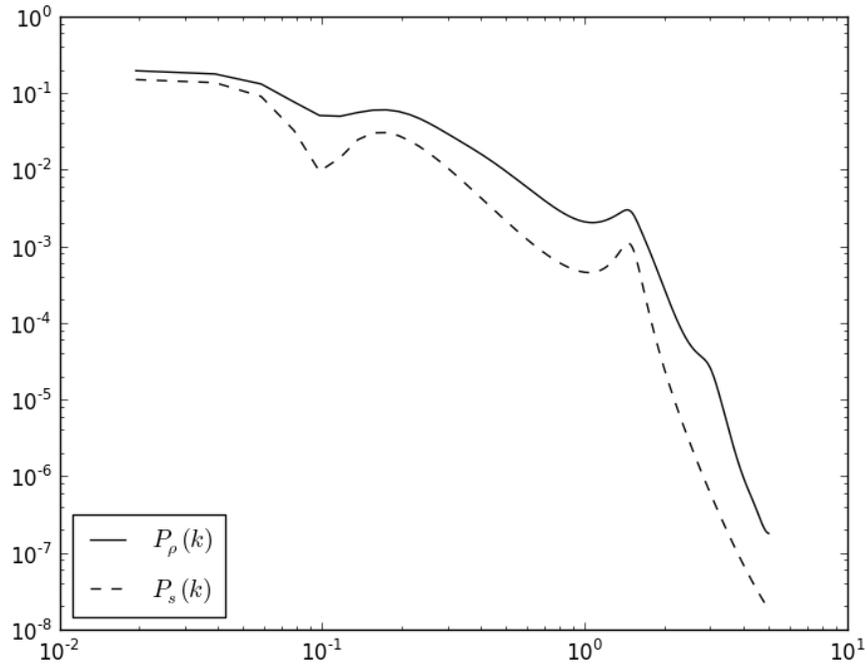


Figure C.1: The backward conversion applied to a one-dimensional power spectrum $P_\rho(k)$ (solid line) yields the power spectrum $P_s(k)$ (dashed line).

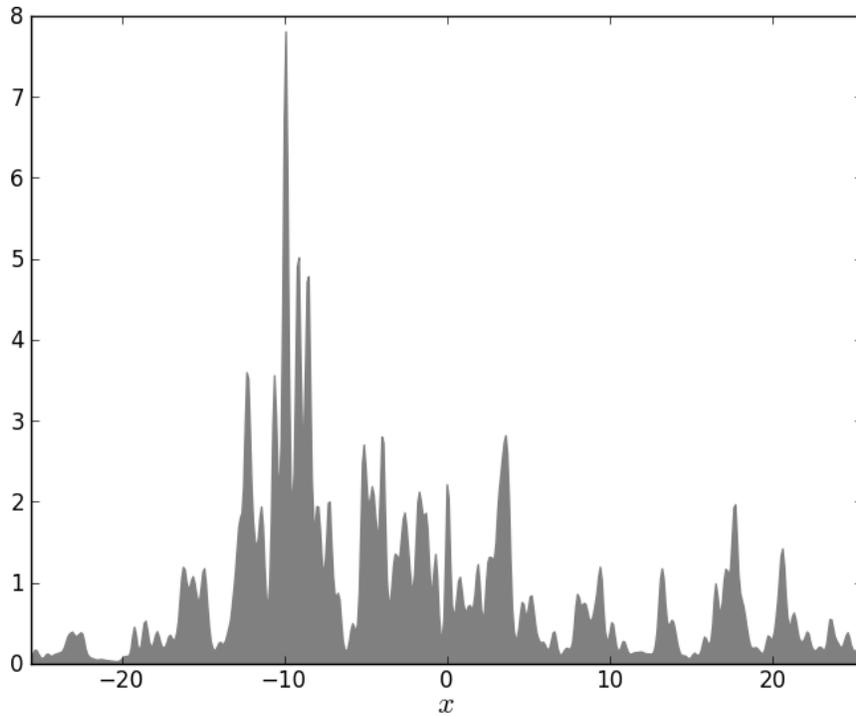


Figure C.2: A one-dimensional log-normal random field following the power spectrum $P_\rho(k)$ from Fig. C.1.

one needs to pay careful attention to volume factors arising from the discretization of the integrals. We define the discrete Fourier transform as

$$s(\vec{k}) = \sum_{\vec{x}} V_x e^{i2\pi \vec{k} \cdot \vec{x}} s(\vec{x}) \quad (\text{C.1})$$

We note that the definition of the Fourier vector has changed with respect to the main text, $\vec{k} = \vec{k}/(2\pi)$, as this is more common in a numerical setting. The vector \vec{x} can only take discrete values,

$$\begin{aligned} \vec{x}_j &= n_j \Delta x_j \quad \text{with} \\ n_j &\in \{-N_j/2, -N_j/2 + 1, \dots, N_j/2 - 1\}, \end{aligned} \quad (\text{C.2})$$

where the dimension j is discretized into N_j pixels with edge length Δx_j . Therefore, our discretized space consists of $N = \prod_j N_j$ pixels that have the volume

$$V_x = \prod_j \Delta x_j. \quad (\text{C.3})$$

This defines a conjugate set of discrete values for \vec{k} ,

$$\begin{aligned} \vec{k}_j &= n_j \Delta k_j \quad \text{with} \\ n_j &\in \{-N_j/2, -N_j/2 + 1, \dots, N_j/2 - 1\}, \end{aligned} \quad (\text{C.4})$$

where $\Delta k_j = 1/(N_j \Delta x_j)$. Therefore, the conjugate space consists of N pixels that have the volume $V_k = 1/(N V_x)$. The power spectrum is defined analogously to Eq. (4.7) by

$$\sum_{\vec{x} \vec{y}} e^{i2\pi \vec{k} \cdot \vec{x} - i2\pi \vec{q} \cdot \vec{y}} \langle s(\vec{x}) s(\vec{y}) \rangle_{\mathcal{P}(s)} = \frac{1}{V_k} \delta_{\vec{k} \vec{q}} P_s(\vec{k}), \quad (\text{C.5})$$

where

$$\delta_{\vec{k} \vec{q}} = \begin{cases} 1 & \text{if } \vec{k} = \vec{q} \\ 0 & \text{otherwise} \end{cases}. \quad (\text{C.6})$$

Defining $\rho(\vec{x}) := e^{m+s(\vec{x})}$ one can derive¹ the forward conversion formula as

$$P_\rho(\vec{k}) = \sum_{\vec{x}} V_x e^{i2\pi \vec{k} \cdot \vec{x}} e^{2m} \exp \left(\sum_{\vec{q}} V_k \left(e^{-i2\pi \vec{x} \cdot \vec{q}} + 1 \right) P_s(\vec{q}) \right). \quad (\text{C.7})$$

The backward conversion formula is

$$P_s(\vec{k} \neq \vec{0}) = \sum_{\vec{x}} V_x e^{i2\pi \vec{k} \cdot \vec{x}} \log \left(\sum_{\vec{q}} V_k e^{-i2\pi \vec{q} \cdot \vec{x}} P_\rho(\vec{q}) \right) \quad (\text{C.8})$$

¹Assuming Gaussianity and statistical homogeneity of s .

and for the monopole

$$P_s(\vec{k} = \vec{0}) = \sum_{\vec{x}} V_x \log \left(\sum_{\vec{q}} V_k e^{-i2\pi \vec{q} \cdot \vec{x}} P_\rho(\vec{q}) \right) - \frac{1}{2} \frac{1}{V_k} \log \left(\sum_{\vec{q}} V_k P_\rho(\vec{q}) \right) - \frac{1}{V_k} m. \quad (\text{C.9})$$

C.3 The conversion for spherical harmonics

The spherical harmonics conversion formulas are implemented in the NIFTY python package by Selig et al. (2013) (version 0.8.4 or higher).

We define the spherical harmonics as

$$Y_l^m(\theta, \varphi) = \sqrt{\frac{(2l+1)(l-m)!}{4\pi(l+m)!}} P_l^m(\cos \theta) e^{im\varphi}, \quad (\text{C.10})$$

where we will write Ω as a short-hand notation for θ, φ with $\Omega = 0$ corresponding to $\theta = 0, \varphi = 0$. In the spherical harmonics basis we define the power spectrum as

$$P_s(l) \delta_{ll'} \delta_{mm'} = S_{lm}^{l'm'} = \langle s_{lm} s_{l'm'}^* \rangle_{\mathcal{P}(s)}, \quad (\text{C.11})$$

where we assumed statistical homogeneity and isotropy. In this section, δ_{ij} denotes the Kronecker delta and $*$ denotes complex conjugation. The covariance matrix in pixel space is related to the power spectrum by

$$S(\Omega, \Omega') = \sum_{l=0}^{\infty} \sum_{m=-l}^l \sum_{l'=0}^{\infty} \sum_{m'=-l'}^{l'} Y_l^m(\Omega) Y_{l'}^{m'}(\Omega') \times P_s(l) \delta_{ll'} \delta_{mm'}, \quad (\text{C.12})$$

which in the case of $\Omega' = 0$ is

$$S(\Omega, 0) = \sum_{l=0}^{\infty} \sum_{m=-l}^l Y_l^m(\Omega) \delta_{m0} \sqrt{\frac{2l+1}{4\pi}} P_s(l). \quad (\text{C.13})$$

Because of statistical homogeneity and isotropy, this stays the same for all angles that are separated by Ω .

Following a similar calculation as in Sects. 4.2.2 and 4.2.3 we derive the forward conversion as

$$P_\rho(l) = \sqrt{\frac{4\pi}{2l+1}} \int d\Omega Y_l^0(\Omega)^* e^{2m} \times \exp \left(\sum_{l'=0}^{\infty} \left(Y_{l'}^0(\Omega) + \sqrt{\frac{2l'+1}{4\pi}} \right) \sqrt{\frac{2l'+1}{4\pi}} P_s(l') \right) \quad (\text{C.14})$$

and the backward conversion as

$$P_s(l \neq 0) = \sqrt{\frac{4\pi}{2l+1}} \int d\Omega Y_l^0(\Omega)^* \log \left(\sum_{l'=0}^{\infty} Y_{l'}^0(\Omega) \sqrt{\frac{2l'+1}{4\pi}} P_\rho(l') \right) \quad (\text{C.15})$$

and

$$P_s(l=0) + 4\pi m = \int d\Omega \log \left(\sum_{l'=0}^{\infty} Y_{l'}^0(\Omega) \sqrt{\frac{2l'+1}{4\pi}} P_\rho(l') \right) - \frac{1}{2} \sqrt{4\pi} \log \left(\sum_{l'=0}^{\infty} \frac{2l'+1}{4\pi} P_\rho(l') \right) \quad (\text{C.16})$$

C.4 Supplements to Sect. 4.3

C.4.1 Log-distances

We calculate the maximal log-distance between two spectra as

$$\max_k |\log(P_1(k)) - \log(P_2(k))|. \quad (\text{C.17})$$

This distance gives a quantitative measure about the goodness of approximating the non-linear spectra by forward converted spectra and about the reduction of non-linearities in the backward converted spectra. In Sects. 4.3.1 & 4.3.2 we give an overview over the log-distances for $k \leq 1.0 h \text{ Mpc}$. Here, we list all of their values in Table C.1, which also shows the log-distances between the linear and non-linear spectra. As one can see the converted spectra always reduce the log-distance.

C.4.2 The effect of different grids

The calculation in Sect. 4.3 is performed on four different grids which all cover the same physical volume, but with different resolution. The Figures in Sect. 4.3 show the results of the second grid (grid B). The lowest non-zero mode we want to cover is $k = 0.0075 h \text{ Mpc}^{-1}$ for all grids. The highest mode is different for the three grids, the coarsest one having $k_{\text{max}} = 1.0 h \text{ Mpc}^{-1}$. The minimum spectral length covered by a three-dimensional Cartesian box with the origin in the middle is

$$k_{\text{min}} = 2\pi (\max\{L_x, L_y, L_z\})^{-1}, \quad (\text{C.18})$$

where L_x , L_y , and L_z are the total edge lengths of the box. The maximum spectral length is

$$k_{\text{max}} = \pi \sqrt{\Delta x^{-2} + \Delta y^{-2} + \Delta z^{-2}}, \quad (\text{C.19})$$

Table C.1: Highest log-distance (for $k \leq 1.0 h \text{Mpc}^{-1}$) between the forward converted spectra and the emulated spectra (3PT and CosmicEmu, respectively) and the backward converted emulated spectra and the linear spectra. We also list the log-distance between the linear spectra and the emulated spectra for comparison.

redshift	forward & emulated	backward & linear	emulated & linear
7	0.04	0.04	0.14
6	0.05	0.05	0.17
5	0.07	0.07	0.23
4	0.08	0.08	0.31
3	0.09	0.10	0.45
2	0.09	0.11	0.69
1.5	0.10	0.10	0.88
1.2	0.14	0.13	1.0
1.0	0.17 (3PT) 0.2 (Emu)	0.15 (3PT) 0.17 (Emu)	1.1 (3PT) 1.2 (Emu)
0.9	0.2	0.19	1.2
0.8	0.3	0.2	1.3
0.7	0.4	0.2	1.3
0.6	0.4	0.3	1.4
0.5	0.6	0.3	1.5
0.4	0.7	0.4	1.5
0.3	0.8	0.4	1.6
0.2	1.0	0.5	1.7
0.1	1.3	0.7	1.7
0.0	1.5	0.8	1.8

Table C.2: Grids used for conversion of matter density spectra. Units of Δx are Mpc/h , units of k are $h \text{Mpc}^{-1}$. For grid D the power spectra are cut before the conversion at $k = 1.4 h \text{Mpc}^{-1}$ (setting all higher modes to zero).

	grid A	grid B	grid C	grid D
N_{pix}	152^3	214^3	522^3	766^3
Δx	5.51	3.91	1.60	1.09
k_{min}	0.0075	0.0075	0.0075	0.0075
k_{max}	1.0	1.4	3.4	5.0, cut at 1.4

where Δx , Δy , and Δz are the edge lengths of one pixel. This leaves us with an infinite number of possible pixelizations. We restrict ourselves to cubic pixels and an equal number of pixels in each dimension leaving us with only two parameters: the number of pixels per dimension N_{pix} and the pixel edge length Δx .

The first grid (grid A) consists of 152^3 , the second grid (grid B) of 214^3 , the third grid (grid C) of 522^3 , and the fourth grid (grid D) of 766^3 cubic pixels. The pixel edge lengths are chosen in a way that the lowest non-zero value of k is $0.0075 h \text{Mpc}^{-1}$. We summarize the properties of these grids in Table C.2. The upper limit of the spectral range of grid D is higher than $1.4 h \text{Mpc}^{-1}$. To investigate the effect of the pixel size on the conversion we cut the power spectrum at $k = 1.4 h \text{Mpc}^{-1}$ setting all higher modes to zero before the conversion. Therefore, any difference in the results of grid B and D originates from the different choice of pixel sizes, and because grid D covers a sphere in Fourier space whereas grid B covers a cube, since no individual component of the wavevector in grid B can have a value above $0.8 h \text{Mpc}^{-1}$ and wavevectors with a length above that can only be reached in the corners of the cube.

To see the effect of the grid on the conversion we plot a selection of forward converted spectra using all four grids divided by the corresponding emulated spectra in Fig. C.3 and the backward converted spectra divided by the corresponding linear spectra in Fig. C.4. In grid D the power spectrum is only filled up to $k = 1.4 h \text{Mpc}^{-1}$ before the conversion, higher modes are set to zero. In the backward conversion the power spectra start to differ around $k = 0.9 h \text{Mpc}^{-1}$, where the result from grid D starts to overshoot the others. In the forward conversion the differences are much more prominent. For redshift 0 the spectrum from grid C exhibits a clear increase in broad-band power on all scales, because at $z = 0$ the region with $1.4 h \text{Mpc}^{-1} < k \leq 3.4 h \text{Mpc}^{-1}$ contains more power than the region with $k \leq 1.4 h \text{Mpc}^{-1}$. This additional power couples to all scales in the forward conversion. However, for redshifts greater than 1 the spectra appear to be consistent up to $k = 0.8 h \text{Mpc}^{-1}$ and grids A, B, and D appear to be consistent at all redshifts. Table C.3 lists the log-distances up to $k = 1.0 h \text{Mpc}^{-1}$ for grids A, C, and D with respect to grid B. One can see, that the log-distances are comparable with the log-distances of grid B with respect to the emulated and linear spectra listed in Table C.1 in Appendix C.4.1. We therefore conclude that the calculation is consistent between the investigated grids within the precision of our model. There is however a systematic trend apparent in Fig. C.3

Table C.3: Log-distances for $k \leq 1.0 h \text{ Mpc}^{-1}$ between the converted spectra at different grids. The distances are with respect to grid B (the grid used in Sect. 4.3); “(fwd)” indicates forward conversion, “(bwd)” indicates backward conversion.

redshift	0	0.5	1.0	1.5	2	3
grid A (fwd)	0.4	0.16	0.07	0.04	0.05	0.04
grid C (fwd)	1.9	0.7	0.3	0.12	0.07	0.03
grid D (fwd)	0.18	0.2	0.19	0.15	0.12	0.08
grid A (bwd)	0.8	0.18	0.12	0.09	0.06	0.03
grid C (bwd)	0.3	0.2	0.2	0.10	0.07	0.04
grid D (bwd)	0.7	0.4	0.2	0.18	0.13	0.07

that an increase in total power due to a greater covered k -range leads to an increase in broad-band power in the forward converted spectra. We therefore expect the validity of our model to break down for higher high dynamic ranges as the total power of the Cosmic matter spectrum diverges, since its spectral index towards high k is above -3 . The problem with a spectral index above -3 can be seen best in the angle integrated conversion formula Eq. (4.13).

C.4.3 Strength of the smearing of BAO

We estimate the strength of the smearing of baryon acoustic oscillations by comparing to a smoothing (with respect to $\log(k)$) of the linear spectrum using a Gaussian kernel,

$$g(\log(k); \sigma_g) = \frac{1}{\sqrt{2\pi\sigma_g^2}} \exp\left(-\frac{1}{2} \left(\frac{\log(k)}{\sigma_g}\right)^2\right), \quad (\text{C.20})$$

where σ_g is the smoothing length in terms of e -folds. σ_g is a dimensionless distance in $\log(k)$. The smoothed spectrum is therefore

$$P_{\text{sm}}(k; \sigma_g) = \int d\log(q) g(\log(k/q); \sigma_g) P_{\text{lin}}(q). \quad (\text{C.21})$$

We can compare $P_{\text{nl}}/P_{\text{nl,nw}}$ with $P_{\text{sm}}(\sigma_g)/P_{\text{lin,nw}}$ in order to find a σ_g where the smoothing appears to be similar. The higher σ_g , the stronger the smoothing. Here, P_{nl} denotes the non-linear power spectrum with BAO, $P_{\text{nl,nw}}$ the non-linear power spectrum without BAO, and $P_{\text{lin,nw}}$ the linear power spectrum without BAO. The best value for σ_g at each redshift is found by visual comparison in the absence of a more rigorous criterion. We list the best fit values for different redshifts for the 3PT spectra as well as the forward converted spectra in Table C.4. The values for the CosmicEmu spectra are missing because the CosmicEmu code did not allow us to calculate non-linear power spectra without BAO.

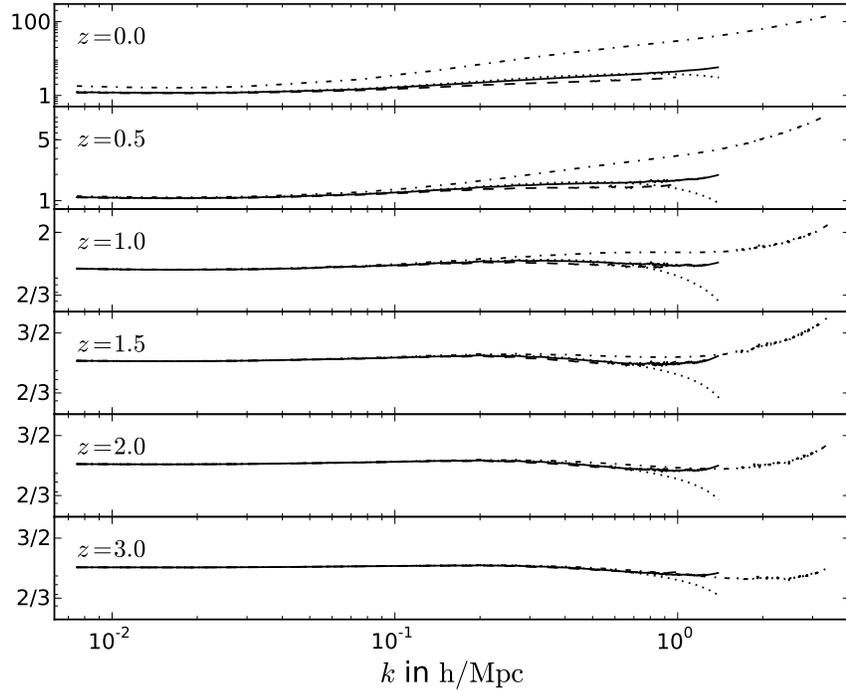


Figure C.3: The forward converted spectra using different grids at a selection of redshifts. The panels show the ratio between the forward converted spectrum and the emulated spectrum (CosmicEmu for redshifts 0, 0.5, and 1 and 3PT else). The solid line corresponds to grid B (the grid used in Sect. 4.3), the dashed line to grid A, the dot-dashed line to grid C, and the dotted line to grid D.

Table C.4: Best fit smoothing scales of a Gaussian smoothing on log-scale that emulate the smearing of BAO for the 3PT spectra and the forward converted spectra at different redshifts. A higher smoothing scale means stronger smoothing. The 3PT result is smoothed stronger than the forward-converted spectrum at all redshifts.

redshift	3PT smoothing scale	forward conversion smoothing scale
7	0.03	0.01
6	0.03	0.01
5	0.035	0.01
4	0.035	0.01
3	0.045	0.015
2	0.055	0.015
1.5	0.07	0.02
1.2	0.07	0.02
1.0	0.075	0.025

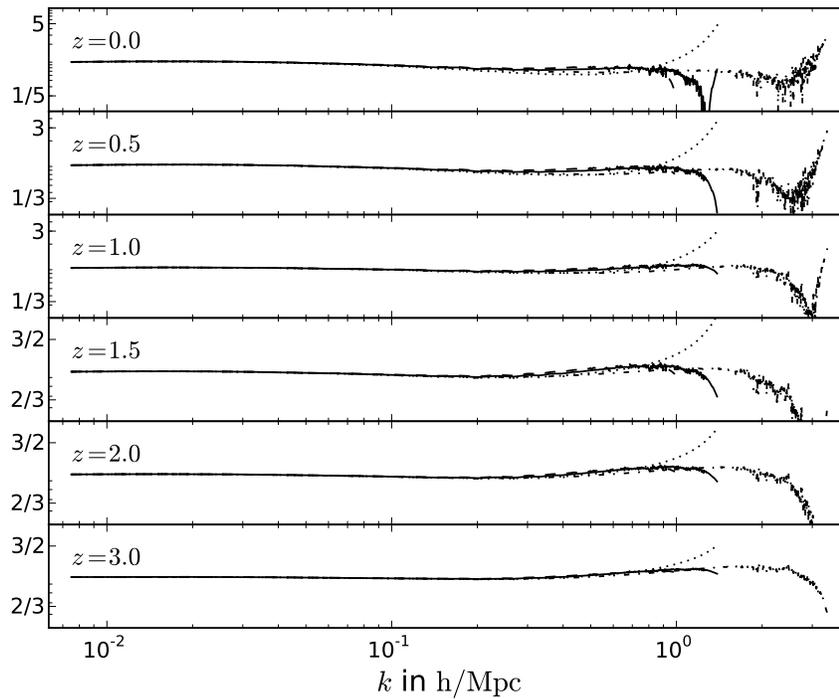


Figure C.4: The backward converted spectra using different grids at a selection of redshifts. The panels show the ratio between the backward converted emulated spectrum (emulators as in Fig. C.3) and the linear spectrum. The solid line corresponds to grid B (the grid used in Sect. 4.3), the dashed line to grid A, the dot-dashed line to grid C, and the dotted line to grid D.

C.4.4 Large-scale bias

Neyrinck et al. (2009) find a bias factor between the the power spectra of the density contrast and the logarithmic density, which they fit to $\exp(-\text{Var}[\log(1 + \delta)])$ with high accuracy up to $z \approx 1.2$. If the density field follows log-normal statistics the correlation functions of the density and log-density are related by

$$\begin{aligned} \langle \rho(\vec{x})\rho(\vec{y}) \rangle_{\mathcal{P}(\rho)} &= \exp\left(\frac{1}{2} \langle s(\vec{x})^2 \rangle_{\mathcal{P}(s)} + \frac{1}{2} \langle s(\vec{y})^2 \rangle_{\mathcal{P}(s)}\right) \times \\ &\quad \exp\left(\langle s(\vec{x}) s(\vec{y}) \rangle_{\mathcal{P}(s)}\right), \end{aligned} \quad (\text{C.22})$$

which simplifies under the assumption of statistical homogeneity and a zero mean for s to

$$\begin{aligned} \langle \rho(\vec{x})\rho(\vec{y}) \rangle_{\mathcal{P}(\rho)} &= \exp(\text{Var}[s]) \times \\ &\quad \exp\left(\langle s(\vec{x}) s(\vec{y}) \rangle_{\mathcal{P}(s)}\right), \end{aligned} \quad (\text{C.23})$$

where $\text{Var}[s] = \langle s(\vec{x})^2 \rangle_{\mathcal{P}(s)}$. This prefactor is the square of the expectation value of ρ ,

$$\langle \rho(x) \rangle_{\mathcal{P}(\rho)} = \exp\left(\frac{1}{2} \langle s(\vec{x})^2 \rangle_{\mathcal{P}(s)}\right) = \sqrt{\exp(\text{Var}[s])}. \quad (\text{C.24})$$

In Sect. 4.3 of this work, we compared the spectra of the δ and $e^s / \langle e^s \rangle$ since the density contrast is defined around a mean of 1. In the backward conversion, the mean of the resulting log-density field was not zero. Eq. 4.18 determined it as

$$m = -\frac{1}{2} \langle s(\vec{x})^2 \rangle_{\mathcal{P}(s)} \quad (\text{C.25})$$

for all redshifts (to 0.1% precision). However, this prefactor should not appear in the work of Neyrinck et al. (2009), since they compare the power spectra of δ and $s = \log(1 + \delta)$, where $\delta = \rho/\rho_0 - 1$. This means that the prefactor in Eq. (C.23) is already divided out.

There is an additional bias apart from this factor, which can be seen best in the low redshift panels in the left part of Fig. 4.4. Here the forward converted power spectrum exceeds the original power spectrum even at the lowest k -bin. This is not due to the prefactor presented in the previous paragraph, but simply to mode coupling in the forward conversion. The inverse of this effect is present in the backward conversion, where the backward (i.e., log-transformed) power spectrum undershoots the linear power spectrum at the lowest k -bin. This factor is, however, much weaker than what Neyrinck et al. (2009) find. For completeness, we list the factors between the lowest k -bins, i.e., $P_{\text{backward}}(k \rightarrow 0)/P_{\text{emulated}}(k \rightarrow 0)$, in Table C.5. In conclusion, the log-normal model offers no insight to the bias factor found by Neyrinck et al. (2009).

Table C.5: Large-scale bias factors due to mode coupling. Listed are the values of the backward converted power spectra for the lowest k -bin divided by the power in the lowest k -bin in the respective emulated spectrum.

redshift	$P_{\text{backward}}(k \rightarrow 0) / P_{\text{emulated}}(k \rightarrow 0)$	$\text{Var}[\log(1 + \delta)]$
7	1.00	0.095
6	1.00	0.124
5	1.00	0.168
4	0.99	0.24
3	0.99	0.38
2	0.98	0.65
1.5	0.97	0.89
1.2	0.97	1.10
1.0	0.96	1.27
0.9	0.96	1.34
0.8	0.95	1.44
0.7	0.95	1.55
0.6	0.95	1.66
0.5	0.94	1.78
0.4	0.94	1.90
0.3	0.93	2.04
0.2	0.93	2.17
0.1	0.92	2.32
0.0	0.91	2.47

C.4.5 Generating matter densities in spaces spanning in redshift

In Sect. 4.3 we have established that the power spectra of the exponentiated linear density contrast agree with the emulated spectra to a reasonable accuracy down to redshift 1 and $k \leq 1.0 h \text{ Mpc}^{-1}$. Therefore, the log-density at different redshifts is (in a statistical average) related by a simple global prefactor. This enables us to formulate a local function which translates the density between different redshifts to better accuracy than linear theory. We let $G(z; z_0)$ be the growth factor between redshift z_0 and redshift z and $\delta(z)$ the density contrast at a given redshift. Then we have

$$1 + \delta(z) \approx \exp\left\{-\frac{1}{2}G(z; z_0)^2 \text{Var}[\log(1 + \delta(z_0))]\right\} \times \exp\{G(z; z_0) \log(1 + \delta(z_0))\}, \quad (\text{C.26})$$

where $\text{Var}[\cdot]$ is the variance in one cell.

Using this formula one can easily generate a lognormal field that behaves like the matter density contrast to a good accuracy. One simply takes z_0 to be sufficiently high so that $\text{Var}[\delta(z_0)] \ll 1$ and $\log(1 + \delta(z_0)) \approx \delta(z_0)$. At such a redshift a Gaussian random field generated from the matter power spectrum describes the statistics of $\delta(z_0)$ very well. By applying Eq. (C.26) using a position dependent redshift $z(\vec{x})$,

$$\delta(\vec{x}) \approx \exp\left\{-\frac{1}{2}G(z(\vec{x}); z_0)^2 \text{Var}[\delta(z_0)]\right\} \times \exp\{G(z(\vec{x}); z_0) \delta(z_0)\} - 1, \quad (\text{C.27})$$

one can now generate a log-normal matter density contrast that follows the appropriate non-linear matter power spectrum in each position (or redshift slice).

Bibliography

- Aharonian, F., Arshakian, T. G., Allen, B., et al. 2013, ArXiv e-prints [[arXiv]1301.4124]
- Armstrong, J. W., Rickett, B. J., & Spangler, S. R. 1995, ApJ, 443, 209
- Bates, S. D., Lorimer, D. R., Rane, A., & Swiggum, J. 2014, MNRAS, 439, 2893
- Bennett, C. L., Hill, R. S., Hinshaw, G., et al. 2003, ApJS, 148, 97
- Bennett, C. L., Larson, D., Weiland, J. L., et al. 2013, The Astrophysical Journal Supplement Series, 208, 20
- Berry, M., Ivezić, Ž., Sesar, B., et al. 2012, ApJ, 757, 166
- Bhatnagar, S. & Cornwell, T. J. 2004, A&A, 426, 747
- Bildsten, L., Chakrabarty, D., Chiu, J., et al. 1997, ApJS, 113, 367
- Brentjens, M. A. & de Bruyn, A. G. 2005, A&A, 441, 1217
- Briggs, D. S., Schwab, F. R., & Sramek, R. A. 1999, in Astronomical Society of the Pacific Conference Series, Vol. 180, Synthesis Imaging in Radio Astronomy II, ed. G. B. Taylor, C. L. Carilli, & R. A. Perley, 127
- Burns, J. O., Schwendeman, E., & White, R. A. 1983, ApJ, 271, 575
- Byrd, R. H., Lu, P., & Nocedal, J. 1995, SIAM Journal on Scientific and Statistical Computing, 16, 1190
- Carron, J. 2011, ApJ, 738, 86
- Chamel, N., Haensel, P., Zdunik, J. L., & Fantina, A. F. 2013, International Journal of Modern Physics E, 22, 1330018
- Clark, B. G. 1980, A&A, 89, 377
- Coles, P. & Jones, B. 1991, MNRAS, 248, 1
- Cordes, J. M. & Lazio, T. J. W. 2002, ArXiv Astrophysics e-prints [arXiv:astro-ph/0207156]

- Cornwell, T. J. 2008, *IEEE Journal of Selected Topics in Signal Processing*, 2, 793
- Deller, A. T., Brisken, W. F., Chatterjee, S., et al. 2011, in 20th Meeting of the European VLBI Group for Geodesy and Astronomy, held in Bonn, Germany, March 29-30, 2011, Eds: W. Alef, S. Bernhart, and A. Nothnagel, Institut für Geodäsie und Geoinformation, Rheinischen Friedrich-Wilhelms-Universität Bonn, p. 178-182, ed. W. Alef, S. Bernhart, & A. Nothnagel, 178–182
- Eisenstein, D. J. & Hu, W. 1998, *ApJ*, 496, 605
- Enßlin, T. A. & Frommert, M. 2011, *Phys. Rev. D*, 83, 105014
- Enßlin, T. A., Frommert, M., & Kitaura, F. S. 2009, *Phys. Rev. D*, 80, 105005
- Enßlin, T. A. & Weig, C. 2010, *Phys. Rev. E*, 82, 051112
- Fanti, C., Fanti, R., de Ruiter, H. R., & Parma, P. 1986, *A&AS*, 65, 145
- Feretti, L., Comoretto, G., Giovannini, G., Venturi, T., & Wehrle, A. E. 1993, *ApJ*, 408, 446
- Ferrière, K. M. 2001, *Reviews of Modern Physics*, 73, 1031
- Finley, D. G. & Goss, W. M., eds. 2000, *Radio interferometry : the saga and the science*
- Frigo, M. & Johnson, S. G. 2005, *Proceedings of the IEEE*, 93, 216, special issue on “Program Generation, Optimization, and Platform Adaptation”
- Gaensler, B. M., Madsen, G. J., Chatterjee, S., & Mao, S. A. 2008, *PASA*, 25, 184
- George Casella, E. I. G. 1992, *The American Statistician*, 46, 167
- Giovannini, G., Cotton, W. D., Feretti, L., Lara, L., & Venturi, T. 1998, *ApJ*, 493, 632
- Gold, T. 1968, *Nature*, 218, 731
- Greiner, M. & Enßlin, T. A. 2015, *A&A*, 574, A86
- Greiner, M., Schnitzeler, D. H. F. M., & Ensslin, T. A. 2016a, *A&A*, 590, A59
- Greiner, M., Vacca, V., Junklewitz, H., & Enßlin, T. A. 2016b, *ArXiv e-prints* [[arXiv]1605.04317]
- Hamilton, A. J. S. 1985, *ApJ*, 292, L35
- Hastings, W. K. 1970, *Biometrika*, 57, 97
- Heger, A., Fryer, C. L., Woosley, S. E., Langer, N., & Hartmann, D. H. 2003, *ApJ*, 591, 288

- Heitmann, K., Higdon, D., White, M., et al. 2009, *ApJ*, 705, 156
- Heitmann, K., White, M., Wagner, C., Habib, S., & Higdon, D. 2010, *ApJ*, 715, 104
- Högbom, J. A. 1974, *A&AS*, 15, 417
- Hubble, E. 1934, *ApJ*, 79, 8
- Jansson, R. & Farrar, G. R. 2012a, *ApJ*, 757, 14
- Jansson, R. & Farrar, G. R. 2012b, *ApJ*, 761, L11
- Jasche, J. & Wandelt, B. D. 2013, *MNRAS*, 432, 894
- Jaynes, E. T. & Bretthorst, G. L. 2003, *Probability Theory*, 758
- Jeong, D. & Komatsu, E. 2006, *ApJ*, 651, 619
- Junklewitz, H., Bell, M. R., Selig, M., & Enßlin, T. A. 2016, *A&A*, 586, A76
- Kalberla, P. M. W. & Kerp, J. 2009, *ARA&A*, 47, 27
- Kayo, I., Taruya, A., & Suto, Y. 2001, *ApJ*, 561, 22
- Lallement, R., Vergely, J.-L., Valette, B., et al. 2014, *A&A*, 561, A91
- Lawrence, E., Heitmann, K., White, M., et al. 2010, *ApJ*, 713, 1322
- Lewis, A., Challinor, A., & Lasenby, A. 2000, *ApJ*, 538, 473
- Lieb, E. H. & Yau, H.-T. 1987, *ApJ*, 323, 140
- Lorimer, D. R., Faulkner, A. J., Lyne, A. G., et al. 2006, *MNRAS*, 372, 777
- Manchester, R. N., Hobbs, G. B., Teoh, A., & Hobbs, M. 2005, *VizieR Online Data Catalog*, 7245, 0
- Maoz, D. 2016, *Astrophysics in a Nutshell*, Second Edition (Princeton University Press)
- Markevitch, M. & Vikhlinin, A. 2007, *Phys. Rep.*, 443, 1
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. 1953, *J. Chem. Phys.*, 21, 1087
- Neal, R. M. 2012, *ArXiv e-prints* [[arXiv]1206.1901]
- Neyrinck, M. C., Szapudi, I., & Szalay, A. S. 2009, *ApJ*, 698, L90
- Nordlund, Å. K. & Padoan, P. 1999, in *Interstellar Turbulence*, ed. J. Franco & A. Car-
raminana, 218

- Oppermann, N., Junklewitz, H., Greiner, M., et al. 2015, *A&A*, 575, A118
- Oppermann, N., Junklewitz, H., Robbers, G., et al. 2012, *A&A*, 542, A93
- Oppermann, N., Robbers, G., & Enßlin, T. A. 2011, *Phys. Rev. E*, 84, 041118
- Oppermann, N., Selig, M., Bell, M. R., & Enßlin, T. A. 2013, *Phys. Rev. E*, 87, 032136
- Oppermann, N., Selig, M., Bell, M. R., & Enßlin, T. A. 2013, *Phys. Rev. E*, 87, 032136
- Pacini, F. 1968, *Nature*, 219, 145
- Passot, T. & Vázquez-Semadeni, E. 1998, *Phys. Rev. E*, 58, 4501
- Percival, W. J., Verde, L., & Peacock, J. A. 2004, *MNRAS*, 347, 645
- Planck Collaboration, Adam, R., Ade, P. A. R., et al. 2015, *ArXiv e-prints* [[arXiv]1502.01588]
- Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2014, *A&A*, 571, A16
- Politzer, H. D. & Wise, M. B. 1984, *ApJ*, 285, L1
- Rau, U. & Cornwell, T. J. 2011, *A&A*, 532, A71
- Robbins, H. 1956, in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics* (Berkeley, Calif.: University of California Press), 157–163
- Ryle, M. & Hewish, A. 1960, *MNRAS*, 120, 220
- Sale, S. E. & Magorrian, J. 2014, *MNRAS*, 445, 256
- Sault, R. J. & Oosterloo, T. A. 2007, *ArXiv Astrophysics e-prints* [astro-ph/0701171]
- Sault, R. J. & Wieringa, M. H. 1994, *A&AS*, 108
- Schlickeiser, R. 2002, *Cosmic Ray Astrophysics*
- Schnitzeler, D. H. F. M. 2012, *MNRAS*, 427, 664
- Schwab, F. R. 1984, *AJ*, 89, 1076
- Selig, M., Bell, M. R., Junklewitz, H., et al. 2013, *A&A*, 554, A26
- Selig, M. & Enßlin, T. A. 2015, *A&A*, 574, A74
- Skilling, J. 2004, in *American Institute of Physics Conference Series, Vol. 735, American Institute of Physics Conference Series*, ed. R. Fischer, R. Preuss, & U. V. Toussaint, 395–405

- Smits, R., Kramer, M., Stappers, B., et al. 2009, *A&A*, 493, 1161
- Smits, R., Tingay, S. J., Wex, N., Kramer, M., & Stappers, B. 2011, *A&A*, 528, A108
- Springel, V., White, S. D. M., Jenkins, A., et al. 2005, *Nature*, 435, 629
- Sun, X.-H. & Reich, W. 2010, *Research in Astronomy and Astrophysics*, 10, 1287
- Sun, X. H., Reich, W., Waelkens, A., & Enßlin, T. A. 2008, *A&A*, 477, 573
- Taylor, G. B., Carilli, C. L., & Perley, R. A., eds. 1999, *Astronomical Society of the Pacific Conference Series*, Vol. 180, *Synthesis Imaging in Radio Astronomy II*
- Taylor, J. H. & Cordes, J. M. 1993, *ApJ*, 411, 674
- Thompson, A. R., Moran, J. M., & Swenson, G. W. 1986, *Interferometry and synthesis in radio astronomy*
- Vacca, V., Murgia, M., Govoni, F., et al. 2012, *A&A*, 540, A38
- Vacca, V., Oppermann, N., Ensslin, T., et al. 2015, *ArXiv e-prints* [[arXiv]1509.00747]
- Verbiest, J. P. W., Lorimer, D. R., & McLaughlin, M. A. 2010, *MNRAS*, 405, 564
- Wang, X., Neyrinck, M., Szapudi, I., et al. 2011, *ApJ*, 735, 32

Danksagung

Zunächst möchte ich meinem Betreuer Torsten Enßlin danken. Seine Begeisterung für das Thema der Signalrekonstruktion und seine Visionen von noch jungen Projekten waren ansteckend und motivierend. Seine Zuversicht und sein Vertrauen in seine Studenten hat ihn zu einem sehr entspannten aber gleichzeitig effektiven Betreuer für mich gemacht. Ich bin sehr froh über meine Entscheidung, sowohl meine Masterarbeit als auch meine Doktorarbeit bei ihm gemacht zu haben. In meiner Zeit in seiner Gruppe hatte ich viele tolle Kommilitonen und Kollegen die mich anfangs betreut und mit denen ich später zusammengearbeitet habe. In meiner Anfangszeit sind da besonders Niels Oppermann und Marco Selig hervorzuheben. Später vor allem Sebastian Dorn, Valentina Vacca, Henrik Junklewitz und Theo Steininger. Außerhalb der Enßlin Gruppe bin ich vor allem Martin Reinecke und Jörg Knoche aus der Planck Gruppe und Frederik Beaujean und Jens Jasche vom Exzellenzcluster Universe zu Dank verpflichtet. Des Weiteren danke ich Marat Gilfanov und Thorsten Naab für ihre Beratung als mein *PhD committee*. Die vielen weiteren Kollegen, die mir in dieser Zeit immer wieder mit Rat und Tat zur Seite gestanden haben bleiben hier unerwähnt, aber nicht vergessen.

Meiner Familie, meinen Freunden und meiner Lebensgefährtin, die mein Leben außerhalb des Instituts so wunderbar machen, möchte ich hier nur kurz danken. Ihr wisst wer ihr seid.