

RESOLVE: A new algorithm for aperture synthesis imaging of extended emission in radio astronomy

H. Junklewitz^{1,2}, M. R. Bell¹, M. Selig^{1,2}, and T. A. Enßlin^{1,2}

¹ Max-Planck Institut für Astrophysik, Karl-Schwarzschild-Str. 1, 85748 Garching, Germany
e-mail: henrikju@astro.uni-bonn.de

² Ludwig-Maximilians-Universität München, Geschwister-Scholl-Platz 1, 80539 München, Germany

Received 20 November 2013 / Accepted 30 April 2015

ABSTRACT

We present **RESOLVE**, a new algorithm for radio aperture synthesis imaging of extended and diffuse emission in total intensity. The algorithm is derived using Bayesian statistical inference techniques, estimating the surface brightness in the sky assuming a priori log-normal statistics. **RESOLVE** estimates the measured sky brightness in total intensity, and the spatial correlation structure in the sky, which is used to guide the algorithm to an optimal reconstruction of extended and diffuse sources. During this process, the algorithm succeeds in deconvolving the effects of the radio interferometric point spread function. Additionally, **RESOLVE** provides a map with an uncertainty estimate of the reconstructed surface brightness. Furthermore, with **RESOLVE** we introduce a new, optimal visibility weighting scheme that can be viewed as an extension to robust weighting. In tests using simulated observations, the algorithm shows improved performance against two standard imaging approaches for extended sources, Multiscale-CLEAN and the Maximum Entropy Method.

Key words. methods: data analysis – methods: statistical – techniques: image processing – techniques: interferometric – radio continuum: general

1. Introduction

Aperture synthesis techniques using large interferometers have a long and successful history in radio astronomy (Ryle & Hewish 1960; Thompson et al. 1986; Finley & Goss 2000). While enabling observers to achieve very high resolutions, data processing with large interferometers is considerably more complicated than it is with a single dish instrument. A radio interferometer effectively measures the Fourier transformation of the sky brightness (see, e.g., Thompson et al. 1986). Unfortunately, inverting this relationship to achieve an estimate of the desired source brightness is a nontrivial task since an interferometer only samples a fraction of the Fourier plane, effectively convolving the true image brightness with an observation-dependent pointspread function. A crucial part in data reduction is therefore the imaging, i.e., estimating the sky brightness distribution from the observed data.

The development of new imaging methods is still a field of ongoing research. An important reason is that all widely used imaging algorithms in radio astronomy have a number of drawbacks. The most successful method CLEAN (Högbom 1974; Clark 1980; Schwab 1984) assumes the image to be comprised of uncorrelated point sources and, therefore, is naturally nonoptimal for highly resolved, extended, and diffuse sources. Some of the newest enhancements of CLEAN try to address this problem using a multiscale approach (Cornwell 2008; Rau & Cornwell 2011), but it is not clear, in general, how to properly choose the scales. The maximum entropy method (MEM, Cornwell & Evans 1985), is by design prone to oversmoothing the images. The non-negative-least-squares (NNLS) approach, has been shown to be an improvement over CLEAN only on mildly extended sources (Briggs 1995a; Sault & Oosterloo 2007). The new Adaptive Scale Pixel (ASP) method (Bhatnagar & Cornwell 2004) and very recent approaches using wavelets

within the framework of compressed sensing (Wiaux et al. 2009; Carrillo et al. 2012, 2013) seem promising to overcome many of these problems. For all of these methods, however, no reliable uncertainty estimates for the image reconstruction are available to date (e.g., Thompson et al. 1986; Taylor et al. 1999).

A second incentive for new developments of imaging techniques are recent advances in radio astronomical instrumentation, where new developments in data analysis are required to exploit new capabilities in data acquisition. The new generation of radio telescopes, such as the upgraded VLA, LOFAR, the SKA pathfinder missions or ultimately the SKA itself, are opening new horizons in radio astronomy (see, e.g., Garrett 2012). Their unprecedented capabilities of simultaneous, broadband frequency coverage including previously unexplored wavelength regimes, sensitivity, and wide fields of view will almost certainly advance astrophysical and cosmological sciences (see, e.g., the German SKA white paper, Aharonian et al. 2013).

In this paper, we introduce **RESOLVE** (Radio Extended Sources Lognormal deconvolution Estimator), a novel algorithm for the imaging of diffuse and extended radio sources in total intensity. We take a new approach to the problem, using Bayesian statistics in the framework of information field theory (Enßlin et al. 2009) and based on clearly formulated mathematical principles. The new algorithm is designed to fulfill two main requirements:

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

In its present form **RESOLVE** also comes with mainly two limitations: it is by design nonoptimal for point sources, and the computational costs are fundamentally more demanding than for

many standard methods (in addition, present implementations are rather inefficient but that is not a principal constraint). The first issue can naturally be solved in the Bayesian framework presented (see Sect. 4) and tests have shown that mildly compact sources can be handled to some degree. The latter would need to be addressed on a more fundamental level of algorithmical research (see Appendix C). To some degree, this behavior is typical and expected for Bayesian methods where accuracy often is paid for by higher computational costs (see, e.g., Jaynes 2003). The approach used in is no exception to this.

The main scientific focus of is by construction on extended and diffuse radio sources. Among those are galaxy clusters with their weak diffuse halos and strong extended relic structures, lobes of radio galaxies, giant radio galaxies, supernova remnants, galactic radio halos, and the radio emission from the Milky Way.

Ultimately, we aim to present the new algorithm together with a Bayesian framework (see Sect. 2) which we believe will be advantageous to formulate and solve upcoming and more complex imaging problems in radio data analysis. Among these, for instance, could be multifrequency techniques for GHz-broadband data, direction-dependent calibration problems, unknown beam reconstructions, polarization imaging, and many more. We come back to an outlook in Sect. 4.

2. The algorithm

2.1. Aperture synthesis

In aperture synthesis, we try to connect an array of telescopes in such a way that we can effectively synthesize a combined instrument with significantly improved resolution. Using the van Cittert-Cernike theorem from the theory of optical coherence (Born & Wolf 1999), it can be shown that such a radio interferometer takes incomplete samples of the Fourier transformed brightness distribution I in the sky (Thompson et al. 1986). In the most basic model, taking an observation of I translates into

$$V(u, v, w) = W(u, v, w) \int dl dm \frac{I(l, m)}{\sqrt{1 - l^2 - m^2}} \times e^{-2\pi i(ul + vm + w\sqrt{1 - l^2 - m^2})}. \quad (1)$$

The quantity $V(u, v, w)$ is the visibility function following classical terminology of optical interferometry. The coordinates u , v , and w are vector components describing the distance between a pair of antennas in an interferometric array, where this distance is usually referred to as a baseline. They are given in numbers of wavelengths, with u and v usually parallel to geographic east-west and north-south, respectively, and w pointing in the direction of the center of the image plane (i.e., the phase center). The coordinates l and m are a measure of the angular distance from the phase center along axes parallel to u and v , respectively. $W(u, v, w)$ is a sampling function defined by the layout of the interferometric array. This function is zero throughout most of the u, v, w -space, apart from where measurements have been made, where it is taken to be unity.

For simplicity, we now restrict ourselves to the common approximation of measuring the sky as flat in a plane tangent to the phase center of the observation, such that $w\sqrt{1 - l^2 - m^2} \approx 0$. Nevertheless, this is not a necessary requirement of our formalism (see Sect. 2.2).

With this assumption, (1) simplifies approximatively to a two-dimensional Fourier transformation,

$$V(u, v) \approx W(u, v) \int dl dm I(l, m) e^{-2\pi i(ul + vm)}. \quad (2)$$

Our instrument measures the visibility function, but we are actually interested in the brightness distribution of the source in the sky. This means that we ideally want to invert the relationship (2). Unfortunately, this is not possible, since we have lost all information on the Fourier modes that have not been measured because of the incomplete sampling of the Fourier plane. Thus, an inversion of (2) does not yield the true brightness distribution, rather we find its convolution with the inverse Fourier transform of the sampling function, better known as the point spread function (psf) or, in common radio astronomical terminology, the dirty beam $I_{\text{db}} = \mathcal{F}^{-1}W$, i.e.,

$$I_{\text{D}} = \mathcal{F}^{-1}V = \mathcal{F}^{-1}W\mathcal{F}I = I_{\text{db}} * I. \quad (3)$$

Here, we introduced a symbolic Fourier operator \mathcal{F} , which is strictly defined later, the common notation I_{D} , dirty image, for the simple Fourier inversion of the visibilities, and the symbol $*$ to denote a convolution operation.

Reconstructing the real brightness distribution is therefore an ill-posed inverse problem. In principle, infinitely many signal realizations could have led to the measured visibility function and we have no way to discriminate between them exactly. However, we can find a statistical description that may produce the most probable signal given the measured visibility function.

2.2. Bayesian signal inference in radio astronomy

In the following, we develop a statistical solution to the inverse problem (2) using Bayesian inference techniques. Later, the condition of a spatially extended source brightness distribution, will lead us to the formulation of . Our derivation relies on notation and methods developed within the framework of information field theory (Enßlin et al. 2009; Enßlin 2013).

To start, we comment briefly on our mathematical notation. As in Eq. (3), we generally use a basis-free description of physical quantities and functions by interpreting them as vectors and operators. This is also common in contemporary literature on imaging (e.g., Rau et al. 2009). A detailed comment on the notation can be found in Appendix B.1.

For an illustration of this notation, properly defining the Fourier operator in (3) as $\mathcal{F}_{kx} = \exp(-i(ul + vm))$ with $x = (l, m)$ and $k = (u, v)$, (2) becomes

$$V_k = W_k \int dx \mathcal{F}_{kx} I_x = W\mathcal{F}I. \quad (4)$$

Following the notation of Enßlin et al. (2009), we define two fundamental quantities, the signal s and the data d . The signal is the ideal, true physical quantity we would like to investigate with our observation. The data is what our measurement device has delivered us. In this radio astronomical application, the signal is the true brightness distribution in the sky or at least directly functionally connected to it, $s \leftrightarrow I(x)$ and the data is our visibility function $d := V(k)$. From now on, we use this definition, but occasionally translate equations into traditional radio astronomical notation for a more transparent presentation.

If we know how to translate the actions of our measurement device into mathematical operations, we can write down

a fundamental data model, connecting signal s and data d with a response operator R as in

$$d = Rs, \quad (5)$$

ignoring measurement noise temporarily.

This is basically Eq. (4), if we identify the response operator with $R = W\mathcal{F}$. We can add more terms to this response operator, slowly introducing more complexity. An inevitable addition is to consider a gridding and de-gridding operation within the sampling $W' = WG$. This is not a feature of the instrument itself, but is needed in its computational representation for purely numerical reasons, to put the visibilities onto a regularly spaced grid to apply the fast Fourier transform algorithm (Cooley & Tukey 1965; Bracewell 1965) to improve computational speed enormously. Henceforth, if not explicitly shown, we drop the prime and consider G to be contained in the sampling operator W .

An important extension is to introduce a mathematical representation of the primary beam pattern A in the response $R = W\mathcal{F}A$. Even more sophisticated instrumental effects like beam smearing, bandwidth effects, or directional dependent sampling could as well be included. Also an extension of the response to noncoplanar baselines, and thus allowing for a non-negligible w -term in Eq. (1), could be directly incorporated without fundamental complication, e.g. in similar form to the w -projection algorithm (Cornwell et al. 2008). For the purpose of this study, none of these instrumental effects are considered explicitly, and do not pose a fundamental problem. For further discussion see Sect. 4.

Another relevant extension is to include multifrequency synthesis by adding a new dimension to signal and data using, e.g., a common spectral model $I(x, \nu) = I(x, \nu_0) \left(\frac{\nu}{\nu_0}\right)^{-\alpha_x}$ yields

$$\begin{aligned} V_{k'} &= \int dx R_{kx} I_{x\nu} \\ &= W_k \int dx \mathcal{F}_{kx} A_x I_{x\nu_0} \left(\frac{\nu}{\nu_0}\right)^{-\alpha_x} \end{aligned} \quad (6)$$

with $k' = k\nu$.

Taking this a step further, a full approach using all four Stokes polarizations is conceivable. In that case, the response representation can in principle be expanded into a full radio interferometer measurement equation (RIME) description, as presented, e.g., by Smirnov (2011a,b). However, both multifrequency and polarization imaging are outside the scope of the present work.

In a real observation, data are always corrupted by measurement noise. This means we have to add this kind of noise contribution n to our data model, as follows:

$$d = Rs + n. \quad (7)$$

As already noted, even without noise, we cannot exactly invert this relationship. We thus instead seek an optimal statistical solution for the signal s given our data d . To find the optimal reconstruction, we regard the signal as a random field following certain basic statistics and being further constrained by the data. In probabilistic terms, we look for an expression of the posterior distribution $\mathcal{P}(s|d)$ of the signal s given the data d . It expresses how the data constrain the space of possible signal realizations by quantifying probabilities for each of them. It comprises all the information we might have obtained through a measurement.

With the posterior distribution, we can in principle estimate the real signal by calculating for instance its posterior mean $\langle s \rangle_{\mathcal{P}(s|d)}$, equivalent to minimizing the

posterior-averaged \mathcal{L}_2 – norm of the quadratic reconstruction error $\arg\min_m \langle \|s - m\|_{\mathcal{L}_2} \rangle_{\mathcal{P}(s|d)}$ (see, e.g., Jaynes 2003). This is the desired type of solution to the ill-posed inverse problem (7).

Probability theory shows that we can calculate $\mathcal{P}(s|d)$ if we have expressions for the likelihood $\mathcal{P}(d|s)$, describing our model of the measurement process and the noise statistics, and for the statistics of the signal alone, the prior distribution $\mathcal{P}(s)$. The renowned Bayes' theorem states this as,

$$\mathcal{P}(s|d) = \frac{\mathcal{P}(d|s)\mathcal{P}(s)}{\mathcal{P}(d)}, \quad (8)$$

where $\mathcal{P}(d)$ is called the evidence distribution. It effectively acts as a normalization factor since it does not depend on s and thus is unimportant for statistical inferences on the signal.

To specify the likelihood for a radio interferometric observation, we only need a valid model for the measurement process. With (7), we see that this involves detailed knowledge of the instrument response R and the statistical properties of the measurement noise n .

Throughout this work, we assume the response representation (2.2) to be exact, or, expressed differently, the data to be fully calibrated. On the perspective of combining calibration and imaging into one inference step, see Sect. 4.

As for the thermal noise of a radio interferometer, it is fair to assume Gaussian statistics, mainly induced by the antenna electronics and independent between measurements at different time steps of the observation (Thompson et al. 1986). Henceforth, the noise field n is assumed to be drawn from a multivariate, zero mean Gaussian distribution of dimension n_d ,

$$\begin{aligned} \mathcal{P}(n) &= \mathcal{G}(n, N) \\ &:= \frac{1}{\det(2\pi N)^{1/2}} \exp\left(-\frac{1}{2}n^\dagger N^{-1}n\right). \end{aligned} \quad (9)$$

The assumption of uncorrelated Gaussian noise leads to a diagonal covariance matrix $N_{kk'} = \delta_{kk'}\sigma_k^2$. For this work, we assume the noise variance σ_k^2 to be known.

We can now derive an expression for the likelihood by marginalizing over the noise field:

$$\begin{aligned} \mathcal{P}(d|s) &= \int \mathcal{D}n \mathcal{P}(d|s, n) \mathcal{P}(n) \\ &= \int \mathcal{D}n \mathcal{P}(d|s, n) \mathcal{G}(n, N) \end{aligned}$$

$$\mathcal{P}(d|s) = \int \mathcal{D}n \delta(n - (d - Rs)) \mathcal{G}(n, N) \quad (10)$$

$$\mathcal{P}(d|s) = \mathcal{G}(d - Rs, N), \quad (11)$$

where the integral is meant to be taken over the infinite space of all possible noise realizations. By inserting the delta function in (10) we stated the implicit assumption that our response (2.2) is exact.

We are left with the crucial question of how to statistically represent our signal prior to the measurement. Until now, the derivation was kept general and we effectively formulated an inference framework for aperture synthesis imaging. Now, we need to specify a prior $\mathcal{P}(s)$, depending on the type of signal field to which the statistical estimation should be optimal.

In the next section, we present a solution to the inference problem with a signal prior chosen to represent the properties of extended and diffuse emission.

2.3. The RESOLVE algorithm

To specify the prior distribution, we choose to follow an approach of least information. The question is: What is the most fundamental, minimal state of knowledge we have about the signal, prior to the measurement and without introducing any specific biases?

We focus on diffuse and extended sources in total intensity. Stating this alone enables us to give a few central assumptions we want to be reflected in the prior distribution:

1. An extended source exhibits a certain, a priori translationally and rotationally invariant (but usually unknown) spatial correlation structure.
2. The signal field must be strictly positive, since it should represent a physical intensity.
3. Typically, signal fields in radio astronomy show high variation in structures across the observed field of view, with a few strong components surrounded by weak extended structure, going over to large regions basically dominated by noise, usually spanning many orders of magnitude in intensity.

Apart from these statements, we assume that we know nothing more specific about our signal, and the prior should be chosen accordingly. For instance, we do not want to include specific source shapes or intensity profiles.

The assumption of translational and rotational invariance is very common and useful in signal inference, where it translates into homogeneity and isotropy of the prior statistics. Given our just stated, restricted prior assumptions, there is no reason, in general, to assume a priori that the correlation of the signal should change under spatial translation or rotation¹. We thus keep this assumption as valid throughout this paper.

The first constraint (1.) urges us to consider how to include the fact that the signal exhibits a spatial correlation of unknown structure. First we might argue to use an uninformative prior, not favoring any particular configuration. However, we do know something, namely that there must be at least some kind of spatial correlation, although its exact structure is obscure to us. Thus, we search for the statistics of a random field about whose correlation we know the least possible, i. e., only the two-point correlation function (equivalent to the second moment of the statistics). Now, the maximum entropy principle of statistics (e.g., [Caticha 2008](#)) states that if we search for such a probability distribution, it must be Gaussian. Of course, a priori, we might not even have any information about the two-point correlation. Nevertheless, it is shown below that the data itself yields this information, which we can extract during the inference procedure.

For the problem of reconstructing a Gaussian signal field with unknown covariance, an optimal solution to the inference problem (7) can actually be found analytically or at least approximately in calculating the posterior mean $\langle s \rangle_{\mathcal{P}(s|d)}$ of the signal. A number of methods have been derived to do this, e.g., the critical filter and variants thereof ([Enßlin & Weig 2010](#); [Enßlin & Frommert 2011](#); [Oppermann et al. 2011b, 2013](#)) or approaches using the method of Gibbs sampling ([Jasche et al. 2010](#); [Sutter et al. 2012](#); [Karakci et al. 2013](#)).

Unfortunately, if we consider the second (2.) and third (3.) constraints from above more closely, we must come to the conclusion that Gaussian signal fields are inappropriate for our

problem since they are neither positive definite nor strongly fluctuating over orders of magnitude in strength.

Instead, we assume that the logarithm of our signal field is Gaussian. If s is now a Gaussian field, $I = e^s$ exhibits all the desired properties (1–3). This is effectively following log-normal statistics. If we adapt the data model (7),

$$d = RI + n = RI_0 e^s + n, \quad (12)$$

we are now faced with a considerably more complicated, non-linear problem. The factor I_0 can be set to account for the right units, w.l.o.g. we set it to one for the rest of this work.

The likelihood $\mathcal{P}(d|s)$ and the signal prior $\mathcal{P}(s)$ take the following form,

$$\begin{aligned} \mathcal{P}(d|s) &= \mathcal{G}(d - Re^s, N) \\ &= \frac{1}{\det(2\pi N)^{1/2}} e^{-\frac{1}{2} (d - Re^s)^\dagger N^{-1} (d - Re^s)}, \end{aligned} \quad (13)$$

$$\begin{aligned} \mathcal{P}(s) &= \mathcal{G}(s, S) \\ &= \frac{1}{\det(2\pi S)^{1/2}} e^{-\frac{1}{2} s^\dagger S^{-1} s}. \end{aligned} \quad (14)$$

Then, the posterior of s

$$\mathcal{P}(s|d) \propto \mathcal{G}(d - Re^s, N) \mathcal{G}(s, S) \quad (15)$$

possibly becomes highly non-Gaussian due to the nonlinearity introduced by (12).

Indeed, the resulting problem cannot be solved analytically. A possible approach would be to separate the quadratic and higher terms in, (15)

$$\mathcal{P}(s|d) \propto e^{-1/2 s^\dagger (S^{-1} + M) s + s^\dagger j + \sum_{n=3}^{\infty} \Lambda_{x_1 \dots x_n}^n s_{x_1} \dots s_{x_n}} \quad (16)$$

where Λ^n is a rank $-n$ tensor, and

$$j = R^\dagger N^{-1} d \quad (17)$$

$$M = R^\dagger N^{-1} R. \quad (18)$$

The higher order terms could be handled either by invoking perturbative methods as known in statistical or quantum field theory ([Huang 1963](#); [Peskin & Schroeder 1995](#)), and already further developed for statistical inference (e.g., [Enßlin et al. 2009](#)), or by using a Monte Carlo Gibbs sampling method ([Hastings 1970](#); [Geman & Geman 1984](#); [Neal 1993](#)). Since these methods are computationally very expensive for this log-normal ansatz and the high dimensionality of the problem, we do not follow them any further in this work.

Instead, we seek an approximate solution m to estimate the signal field that maximizes the posterior,

$$m = \operatorname{argmax}_s \mathcal{P}(s|d) \approx \langle s \rangle_{\mathcal{P}(s|d)}. \quad (19)$$

This method is known as Maximum a posteriori (MAP) in statistical inference and can be interpreted as an approximation to the posterior mean $\langle s \rangle_{\mathcal{P}(s|d)}$ ². For the present problem it leads to a nonlinear optimization problem of a gradient equation for the posterior. With this approach, it is further possible to calculate a consistent uncertainty estimate. In principle, the uncertainty of a signal reconstruction can be estimated by the width of the posterior. In this case, we use the inverse curvature of the posterior

¹ It should be emphasized that this a priori assumption is not in contradiction with a posteriori solution not exhibiting homogeneity and isotropy. Ultimately, if the combination of data and measurement noise allow for a specific source shape, the likelihood dominates the prior and drive the reconstruction in this direction.

² It is not guaranteed to yield a close result, especially not for highly non-Gaussian posterior shapes. Alternatively, it can be derived by minimizing an \mathcal{L}_∞ -norm error measure instead of the \mathcal{L}_2 minimization underlying the posterior mean approach.

at its maximum to approximate the relative uncertainty D (see Appendix A for details).

In this context, we still need to specify how to deal with the unknown correlation structure, i.e., the Gaussian signal covariance $S = \langle s s^\dagger \rangle$. As mentioned earlier, the problem of reconstructing a Gaussian random field with unknown covariance has already been solved (Jasche et al. 2010; Enßlin & Weig 2010; Enßlin & Frommert 2011; Oppermann et al. 2011b; Sutter et al. 2012), and even the respective problem for a log-normal random field has been partly solved before (Oppermann et al. 2013). Unfortunately, none of these methods can be readily applied to the inference problem at hand, since they require the signal response to have a diagonal representation in signal space. This is not necessarily fulfilled for the Fourier response (2.2). We therefore develop a different approach, which, nevertheless, closely follows the previously mentioned works.

Crucially, as explained above, our prior knowledge signal statistics is homogeneous and isotropic. This implies that the unknown signal covariance becomes diagonal in its conjugate Fourier space and can be expressed by its power spectrum $P_s(|k|)$ (see the Wiener-Kinchin theorem in Bracewell 1965),

$$S(k, k') = \langle s(k)s(k')^\dagger \rangle = (2\pi)^{n_s} \delta(k - k') P_s(|k|) \quad (20)$$

where $P_s(|k|)$ is just the Fourier transformation of the homogeneous and isotropic autocorrelation function $C(r) = S(|x - y|)$, where

$$P_s(|k|) = \int dr C(r) \exp(ikr). \quad (21)$$

Because of the assumption of isotropy, the power spectrum only depends on the length $|k|$ of the Fourier vector k . The power spectrum is therefore sensitive to scales but not to full modes in Fourier space. Where the distinction is needed, we will make it explicit using the notation $|k|$.

We now parameterize the unknown covariance S as a decomposition into spectral parameters p_i and positive, disjoint projection operators $S^{(i)}$ onto a number of spectral bands such that the bands fill the complete Fourier domain

$$S = \sum_i p_i S^{(i)}. \quad (22)$$

These parameters can be introduced into the inference problem as a second set of fields to infer.

We therefore add a second MAP algorithm to the signal MAP, solving for these unknown parameters p_i . We then iterate between both solvers until convergence is achieved. The algorithm produces a signal estimate m , an approximation to the reconstruction uncertainty D , and a power spectrum estimate parameter set p_i . At iteration stage n , the equations to be solved are

$$S_{(n-1)}^{-1} m + e^{m_{(n)}} M e^{m_{(n)}} - j e^{m_{(n)}} = 0, \quad (23)$$

$$D_{(n)xy} = S_{(n-1)}^{-1}{}_{xy} + e^{m_{(n)x}} M_{xy} e^{m_{(n)y}} + e^{m_{(n)y}} \int dz M_{xz} e^{m_{(n)z}} - j_x e^{m_{(n)x}} \delta_{xy}, \quad (24)$$

$$p_{(n)i} = \frac{q_i + \frac{1}{2} \text{tr} \left[(m_{(n)} m_{(n)}^\dagger + D_{(n)}) S^{(i)} \right]}{\alpha_i - 1 + \frac{q_i}{2} + (T p)_i}. \quad (25)$$

The two quantities j and M are defined as above, q and α are parameters of a power spectrum parameter prior, q is a measure for the number of degrees of freedom of each Fourier band,

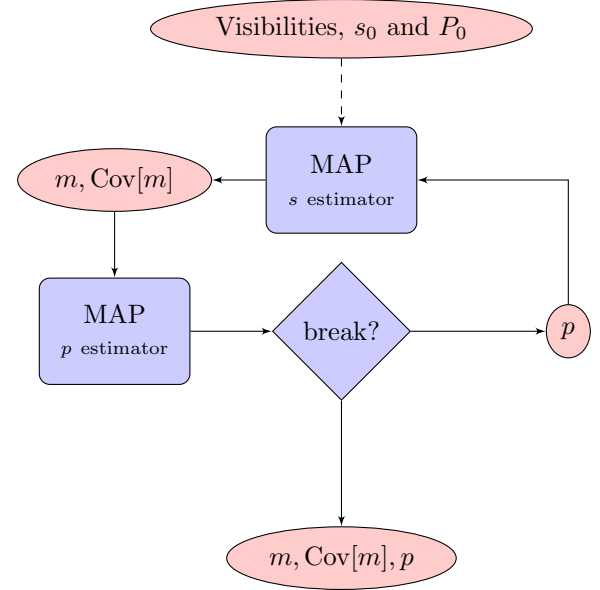


Fig. 1. Flow chart, illustrating the basic workflow of the algorithm.

and T is an operator, which enforces a smooth solution of the power spectrum p_i . A thorough derivation and explanation of all these terms can be found in Appendix A. Equation (23) is the fix point equation that needs to be solved numerically to find a Maximum a posteriori signal estimate $m_{(n)}$ for the current iteration. The second Eq. (24) results from calculating the second derivative of the posterior for the signal estimate $m_{(n)}$, its inverse serves as an approximation to the signal uncertainty $D_{(n)}$ at each iteration step. The last Eq. (25) represents an estimate for the signal power spectrum using the current signal uncertainty $D_{(n)}$ to correct for missing signal power in the current estimate $m_{(n)}$. The iteration is stopped after a suitable convergence criterion is met (see Appendix C). The final estimate for the sky brightness is then $I = e^m \pm \sqrt{e^{2m} [e^D - 1]}$ (see Appendix A for details) using the last estimates for m and D . The whole algorithm is visualized in a flow chart in Fig. 1.

It should be noted that solving these equations can be relatively time-consuming compared to, e.g., MS-CLEAN, depending on the complexity of the problem at hand, since it involves a nonlinear optimization scheme (23) and the numerical inversion and random probing of an implicitly defined matrix (24)³ (for details, see Appendix C). We call the combined algorithm (Radio Extended SOURces Lognormal deconVolution Estimator).

2.4. Properties of RESOLVE

2.4.1. Image weighting and resolution

As derived in A.4, naturally converges to a robust-like image weighting (see Briggs 1995a). It effectively weights all visibilities by the ratio of the reconstructed power spectrum to the noise power spectrum. This is conceptually similar to an

³ The overall computational costs go roughly with $N_{\text{global}} N_{\text{pr}} \mathcal{O}(\sqrt{n_s} n_d)$ in the limit of a large number of visibility measurements n_d . The n_s are the number of pixels in image space, N_{pr} is the number of used random probing vectors to estimate matrix traces, and N_{global} is the global number of iterations needs to converge (see Appendix C).

optimal noise weighting in Wiener Filtering. It is thus unnecessary to set the image weighting by hand and W in (2.2) really only contains the sampling operation and no further weighting.

Since the weighting depends on the converged power spectrum, this also means that the image resolution is determined by the algorithm and cannot simply be predicted beforehand. After the algorithm has been applied, the achieved resolution can be estimated from the reconstruction (see B.2).

2.4.2. Deconvolution

To the first order, the process of image deconvolution with \mathcal{F}^{-1} can be understood considering the multiplicative term e^m in the fix-point Eq. (23). It acts effectively as a convolution kernel in Fourier space, which is exploited by the algorithm for extrapolating the measured visibilities into the regions of uv -space without direct measurements. In this way, \mathcal{F}^{-1} is also capable of achieving some degree of superresolution. For pure extended emission, the tests in Sect. 3 strongly indicate that \mathcal{F}^{-1} deconvolves at least as effectively as standard methods. For a more detailed explanation, see B.2.

2.4.3. Residual images

Residual images are usually defined as the inverse Fourier transform of the difference between the visibility data and the reconstructed image $I_{\text{res}} := \mathcal{F}^{-1}(d - \mathcal{F}m)$ and are frequently used to judge the image quality. For the test simulations presented later in Sect. 3, \mathcal{F}^{-1} provides noise-like residuals as usually expected. However, it should be noted that without taking the uncertainty properly into account, a residual image alone might not be the best measure for image quality with \mathcal{F}^{-1} (for details see B.2).

2.4.4. Image rms-noise and dynamic range

Of course, image noise and dynamic ranges can be calculated for images. For a meaningful comparison with standard results, however, calculating an rms and a peak value should be done considering the uncertainty. For a conservative estimate upper and lower bounds can be used instead of simple image values (see B.2).

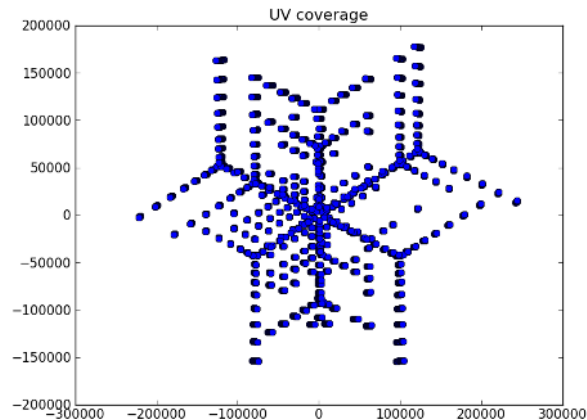
2.4.5. Compact emission

As presented, compact emission cannot be handled optimally using \mathcal{F}^{-1} . Extensions for a combined algorithm are foreseeable (see Sect. 4), but for all practical purposes, \mathcal{F}^{-1} in its present form will need to be combined with a previous step of point-source subtraction for best results.

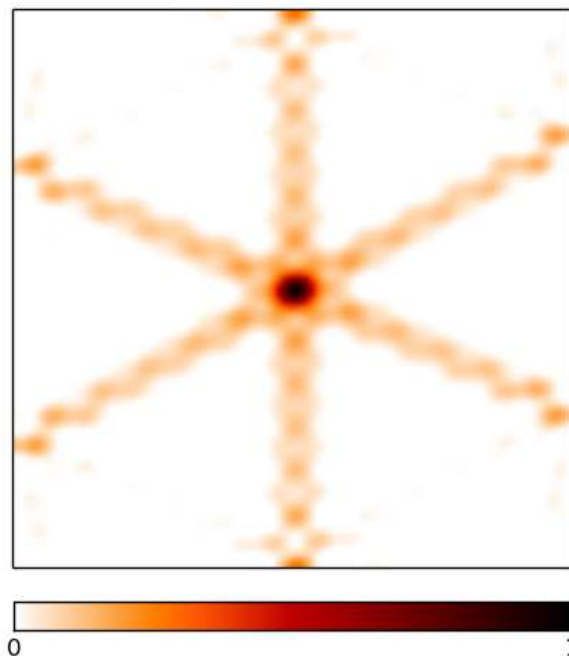
3. Test simulations

In what follows, we present a range of tests of \mathcal{F}^{-1} using simulated data. We have implemented the algorithm⁴ in Python using the versatile signal inference library NIFT (Selig et al. 2013). For all details of the implementation, we refer to Sect. 2 and Appendix C. We also show comparisons to CLEAN and MEM to benchmark the performance and fidelity of our algorithm.

⁴ To get access to the code prior to its envisaged public release, please contact henrikju@mpa-garching.mpg.de or enssllin@mpa-garching.mpg.de



(a) uv -coverage in units of # of wavelengths.



(b) Point spread function.

Fig. 2. Point spread function uv -coverage for the simulated 20 min snapshot observation in VLA-a configuration. The image of the point spread function is 100^2 pixels large, the pixel size corresponds to roughly 0.2 arcsec.

For all tests, we constructed simulated observations with the tool \mathcal{F}^{-1} ⁵ using a realistic uv -coverage from a VLA observation in its A-Configuration. We thus simulated an approximately 20 min snapshot observation with a total of 42 120 visibility measurements at a single central frequency of 1 GHz (see Fig. 2). This setting leads to an especially sparse sampling of the uv -plane. For ease of code development and testing, we have not used longer observations. On the other hand, if we can solve the more demanding cases of sparse uv -coverage, we certainly can handle better-suited data.

For the next two Sects. (3.1 and 3.2), the signals were drawn from a log-normal distribution, exactly meeting our prior

⁵ See <http://www.lofar.org/wiki/lib/exe/fetch.php?media=software:makems.pdf>

assumptions. In Sect. 3.3, we go beyond that and illustrate the validity of our statistical model by using a signal derived from a CLEAN image of a real source.

Through all simulations, we varied thermal visibility noise levels. The variance of the complex Gaussian input noise in uv -space is defined equal for all visibilities. Low noise refers to $\sigma_{\text{ln}}^2 = 10^{-3} \text{ Jy}^2$, whereas high noise denotes $\sigma_{\text{hn}}^2 = 10^5 \text{ Jy}^2$. This translates into an average visibility signal-to-noise ratio of roughly 10^3 – 10^4 and 0.1–1., respectively. These numbers are of course somewhat arbitrary, and are only chosen for demonstrational reasons as extreme cases. They are not intended to necessarily reflect realistic visibility noise values in every possible aspect, but to serve as examples for particularly low- or high-noise cases.

To give a quantitative account of the accuracy of the reconstructions, we use a relative \mathcal{L}_2 – norm measure of the difference of signal to map

$$\delta = \sqrt{\frac{\sum (e^s - e^m)^2}{\sum (e^s)^2}}, \quad (26)$$

where the sums are taken over all pixels of the reconstruction. This choice is motivated by the fact that the inference approach underlying approximates a reconstruction that is optimal in the sense of minimizing this error measure (see Sect. 2 and Eq. (19) therein).

In Sects. 3.1–3.3, we focus exclusively on the reconstruction of the signal, i.e., the sky brightness distribution. The reconstruction of the power spectrum is discussed separately in Sect. 3.5.

3.1. Main test results

Here, we describe the main test results for the reconstruction of a simulated signal using

In Fig. 4, an artificial log-normal signal is shown alongside the results from for observations with low- and high-noise. The error measures are $\delta_{\text{ln}} = 0.12$ and $\delta_{\text{hn}} = 0.3$ for the low and high noise case respectively.

We can recover all the structures of the original surface brightness, down to even very small features in the low-noise case and at least all main features in the high-noise case. All strong effects of the point spread function have been successfully removed, thus showing that is effective in deconvolving the dirty image.

In fact, the reconstruction is expected to be smoothed out on the smaller scales because of the inherent image weighting (see Sect. 2.4.1). All information in the power spectrum gets lost for powers comparable to the noise variance.

On a sidenote, it can be seen that mildly compact emission, for instance in the strongest emission regions of the simulated signal, can be handled by as well. Further tests seem to indicate that even some purely compact emission can be reconstructed by , but further work is clearly needed (see Sect. 4).

For convenience and comparison, in Fig. 3 we show a residual map for the low-noise reconstruction. It qualitatively reveals an almost noise-like structure with mainly gridding artifacts in the background, which usually would be expected for a close reconstruction. However, more remnant substructure in the residual would be consistent with the reconstructed uncertainty as further tests have shown. With the original signal available in the presented simulations, the difference maps are nevertheless a more reliable way to judge the quality of the reconstruction.

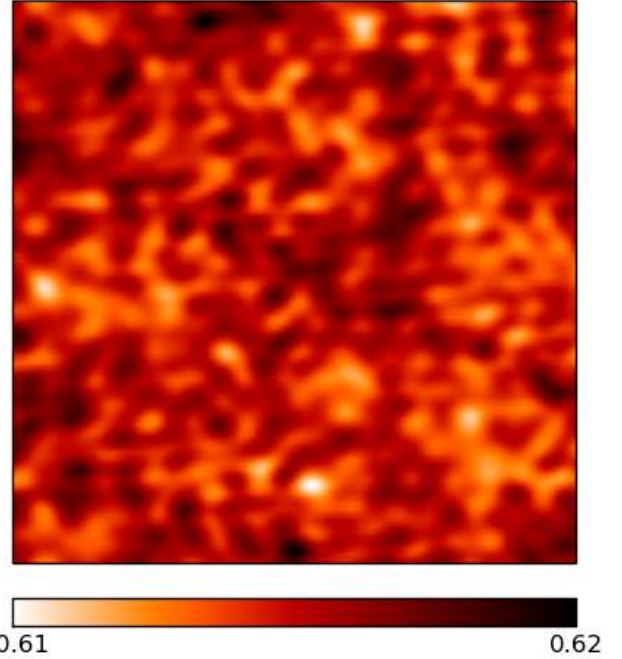


Fig. 3. Residual map for the low noise reconstruction.

3.2. Comparison to standard imaging methods

In this section, we briefly introduce common imaging algorithms in radio interferometry and show comparisons to . We focus on two of them, MS-CLEAN and MEM, which are probably the most widespread methods to date.

In addition, we should mention recent developments in the application of Compressed Sensing (Candes et al. 2006; Donoho 2006, CS) to radio imaging, most notably the development of the sparsity averaging reweighted analysis algorithm (Carrillo et al. 2012, SARA). Another recent approach applied Gibbs sampling methods to imaging in radio interferometry (Sutter et al. 2014), also within the framework of Bayesian inference, but restricted to pure Gaussian priors. Yet another proposed method is the ASP algorithm (ASP). A direct comparison to either SARA, Gibbs Sampling methods, or ASP is out of the scope of this work mainly due to unavailability of robust public implementations, but we discuss possible ways to include the CS approach into our Bayesian framework in the conclusions (see Sect. 4).

For CLEAN, we used the implementation in the radio astronomical software package CASA (Reid & CASA Team 2010); for MEM we utilized the task VTESS from the software package AIPS (Greisen 1990).

3.2.1. Comparison to CLEAN

The CLEAN algorithm was first presented by (Högbom 1974) and is undoubtedly the most widely used deconvolution algorithm in radio astronomy. It works around the major assumption that the image is comprised of point sources. In its simplest variant, it iteratively finds the highest peak in the dirty map, subtracts a psf-convolved fraction of a delta function fitted to the peak, and saves the delta components in a separate image. After some noise threshold is reached, the algorithm stops and reconvolves the components with a so-called clean beam, usually the

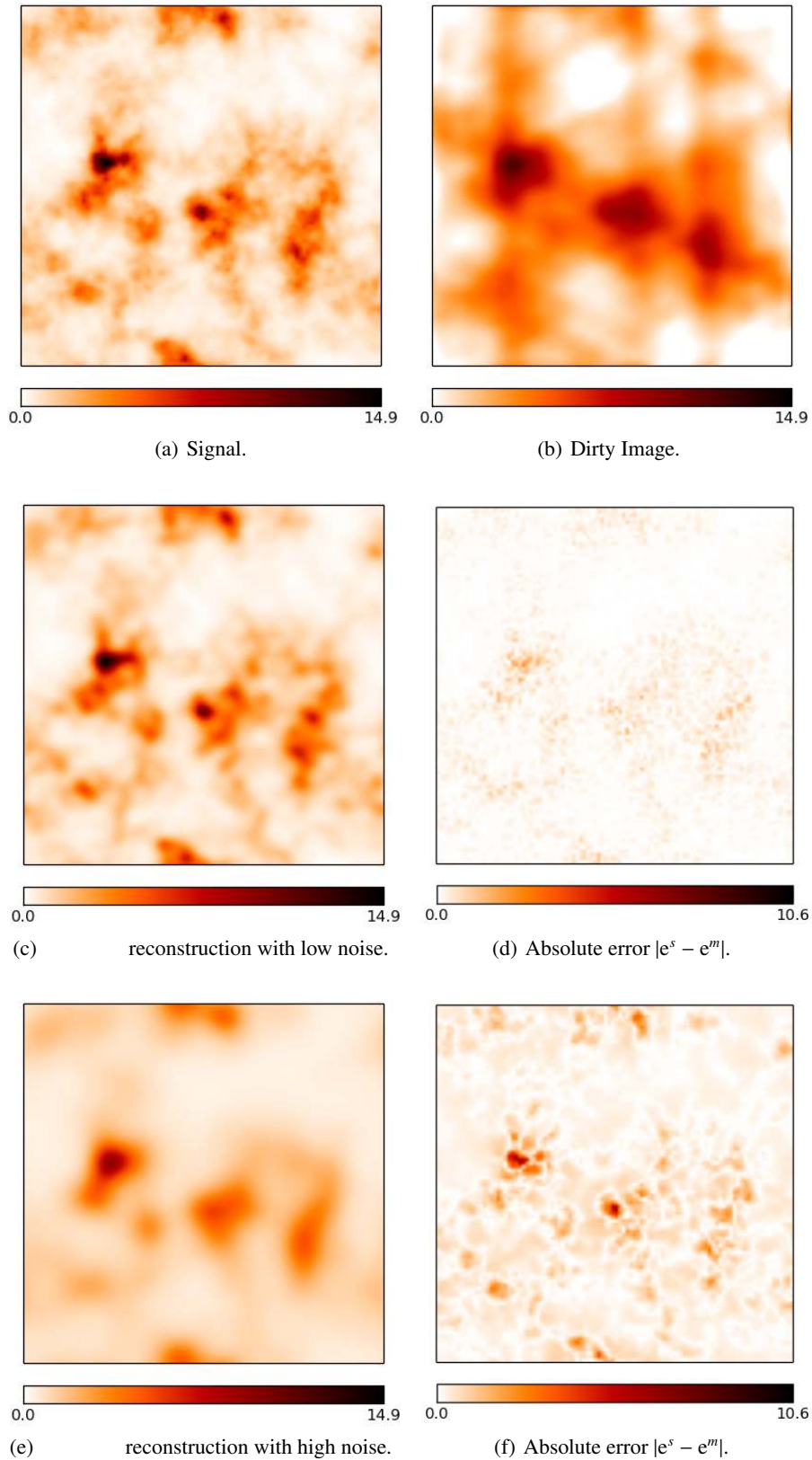


Fig. 4. Reconstruction of a log-normal signal field, observed with a sparse uv -coverage from a VLA-A-configuration and different noise levels. The images are 100^2 pixels large, the pixel size corresponds to roughly 0.2 arcsec. The brightness units are in Jy/px. The ridge-like structures in the difference maps simply stem from taking the absolute value and mark zero-crossings between positive and negative errors. *First row left:* signal field. *First row right:* dirty map. *Second row left:* reconstruction with low noise. *Second row right:* absolute per-pixel difference between the signal and the reconstruction with low noise. *Third row left:* reconstruction with high noise. *Third row right:* absolute per-pixel difference between the signal and the reconstruction with high noise.

main lobe of the point spread function or a broader version of it to downgrade resolution.

Over time, many variants of CLEAN have been developed (Clark 1980; Schwab 1984; Sault & Wieringa 1994). Among those, multiscale CLEAN (MS-CLEAN; Cornwell 2008) was constructed to better reflect extended emission by subtracting Gaussians of various shapes instead of pure point sources. We will compare the results of `msclean` to MS-CLEAN.

In Fig. 5, a comparison is shown between the results of `msclean` and MS-CLEAN. For this test, the same simulated low-noise data were used as in Sect. 3.1. We compare to two different CLEAN reconstructions with natural and uniform weighting. We also compared to robust weighting with a robust parameter of $r = 0$, which yields an intermediate result between the other two schemes. Since the results are only mildly different from uniform weighting, we have left them out. We used a very small noise threshold and a standard gain factor of 0.1. In total, we choose to run the algorithm interactively for around 1000 iterations. We used approximately ten different scales for the multiscale settings, ranging from a single pixel to enough to roughly match the scales found in the signal. Together with the reconstructions, we show maps of the squared difference to the signal $(e^s - m)^2$ for each of them. The \mathcal{L}_2 – error measures and dynamic range values are shown in Table 1. For the `msclean` dynamic range, the most conservative and the most optimistic values are given considering the measurement uncertainty as explained in Sect. 2.4.1. For the most conservative estimate, `msclean` achieves a dynamic range roughly 1.5 times higher than the best CLEAN result.

Both quantitative analysis and visual comparison show that `msclean` clearly outperforms MS-CLEAN in this case. Its result is closer to the signal in the \mathcal{L}_2 error measure sense and it is clearly superior in reconstructing the detailed extended structure of the surface brightness signal. In particular, the very weak emission around all the brighter sources is much better resolved and denoised than in the MS-CLEAN images. The reconstruction with natural weighting is overestimating the flux scales considerably, while uniform and robust weighting roughly find the same correct solution as `msclean`. However, at least for natural weighting, this is a somewhat biased comparison, since the natural weighting scheme is by construction enhancing point-source sensitivity, while preserving larger side-lobe structures (Briggs 1995b), and thus not the optimal choice for resolving extended emission.

3.2.2. Comparison to the maximum entropy method

The maximum entropy method (MEM) is an imaging algorithm introduced into radio astronomy by (Cornwell & Evans 1985). It actually goes back to earlier developments in statistical inference, connected to the broad field of entropic priors (Gull & Daniell 1979; Skilling et al. 1979). It should not be confused with the maximum entropy principle of statistics mentioned earlier, which describes how to update probability distributions when new information has to be included (Caticha 2008; Enßlin & Weig 2010, see also Sect. 2.3).

MEM aims to maximize a quantity called image entropy S_{im} , which is defined for strictly positive signal images s as

$$S_{\text{im}} = - \int dx s(x) \log(s(x)/m(x)) \quad (27)$$

where $m(x)$ is a model image of the observed signal, thus allowing us to introduce some kind of prior information into the

Table 1. \mathcal{L}_2 error measures and dynamic ranges for `msclean`, MS-CLEAN and MEM for the low-noise simulation and the reconstruction shown in Figs. 5 and 6.

Algorithm	δ	DNR
	0.12	73–113
MS-CLEAN, natural	1.46	13
MS-CLEAN, uniform	0.67	49
MEM	0.99	45

problem. The data enter this formalism as a constraint for the maximization problem. Since both, MEM and `msclean` were designed toward extended emission, an analysis of MEM within the presented Bayesian inference framework together with a theoretical comparison to `msclean`, illustrating their significant differences, can be found in Appendix B.3. In short, `msclean` is better suited to represent structured extended emission, because of its implicit reconstruction of the signal correlation as opposed to maximally smoothed reconstructions.

In Fig. 6, a comparison is shown between the results of `msclean` and MEM as implemented in the VTESS task from the radio astronomical software package AIPS. Again, the same simulated low-noise data were used as in Sect. 3.1. As a model image, we used an MS-CLEAN reconstruction with uniform weighting. We again show maps of the squared error $(e^s - m)^2$ for the reconstruction with `msclean` and MEM respectively. The \mathcal{L}_2 error measures are shown in Table 1.

It can be clearly seen that `msclean` also outperforms MEM, as reflected by the ℓ_2 – norm analysis. The overall structure is reconstructed roughly correctly, though some fine structure is clearly missing. Additionally, MEM underestimates the image peak values in general, which is expected because of the specific smoothing MEM prior (see Appendix B.3).

3.3. Comparison with a real signal

So far we have only shown reconstructions of signals that were drawn from log-normal statistics, using the exact assumptions that we use to specify the prior distribution. It is expected that `msclean` should be optimal for these simulated signals.

To further demonstrate the validity of our assumptions, we have conducted a test, in which we did not use a signal drawn from log-normal statistics. Instead, we took an MS-CLEAN image, obtained from real data of the galaxy cluster Abell 2256 (Clarke & Enßlin 2006) and reused this as a signal for the simulated observation using the same VLA configuration as before. The original data were taken with the VLA at 1.369 GHz in D-configuration. The surface brightness values are not in the original range but chosen arbitrarily in our simulation, effectively given in Jy/px. The signal (i.e., the adapted CLEAN image of Abell 2256) and the reconstruction from `msclean` are shown in Fig. 8.

Although this time we have at no point introduced log-normal statistics into the simulation process, the prior assumption still seems to be valid and leads to results comparable in exactness to the tests using explicit log-normal signals.

3.4. Signal uncertainty

As already stated in Sect. 2.3, `msclean` provides also an estimate of the uncertainty of the signal reconstruction. The algorithm uses the inverse second derivative D of the posterior, evaluated at the specific signal estimate m , to approximate the posterior

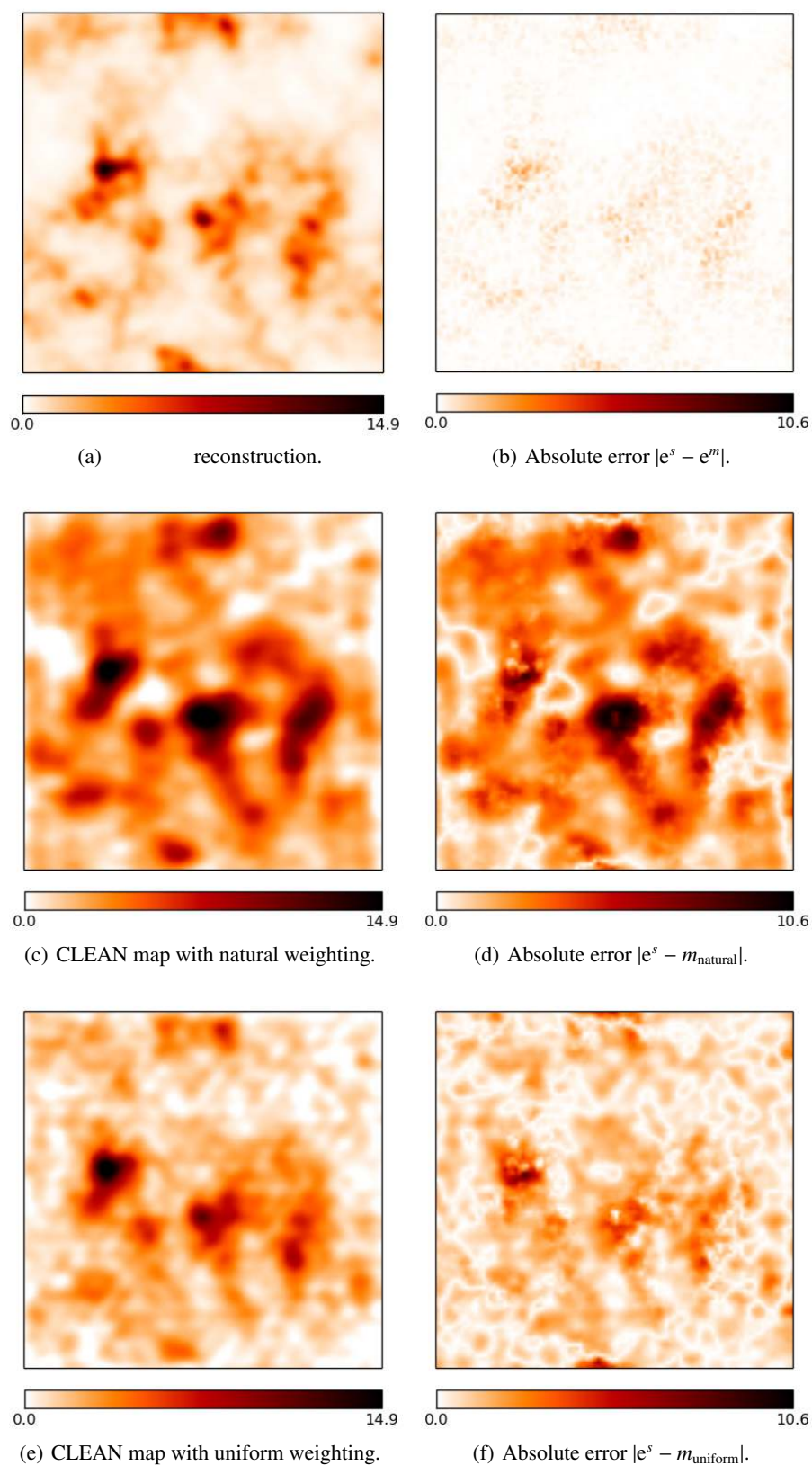


Fig. 5. Comparison of m_{natural} with MS-CLEAN for the simulated low-noise observation of Sect. 3.1. The images are 100^2 pixels large, the pixel size corresponds to roughly 0.2 arcsec. The brightness units are in Jy/px. The ridge-like structures simply stem from taking the absolute value and mark zero-crossings between positive and negative errors. *From first to last row:* MS-CLEAN with uniform weighting, MS-CLEAN with natural weighting, MS-CLEAN with uniform weighting.

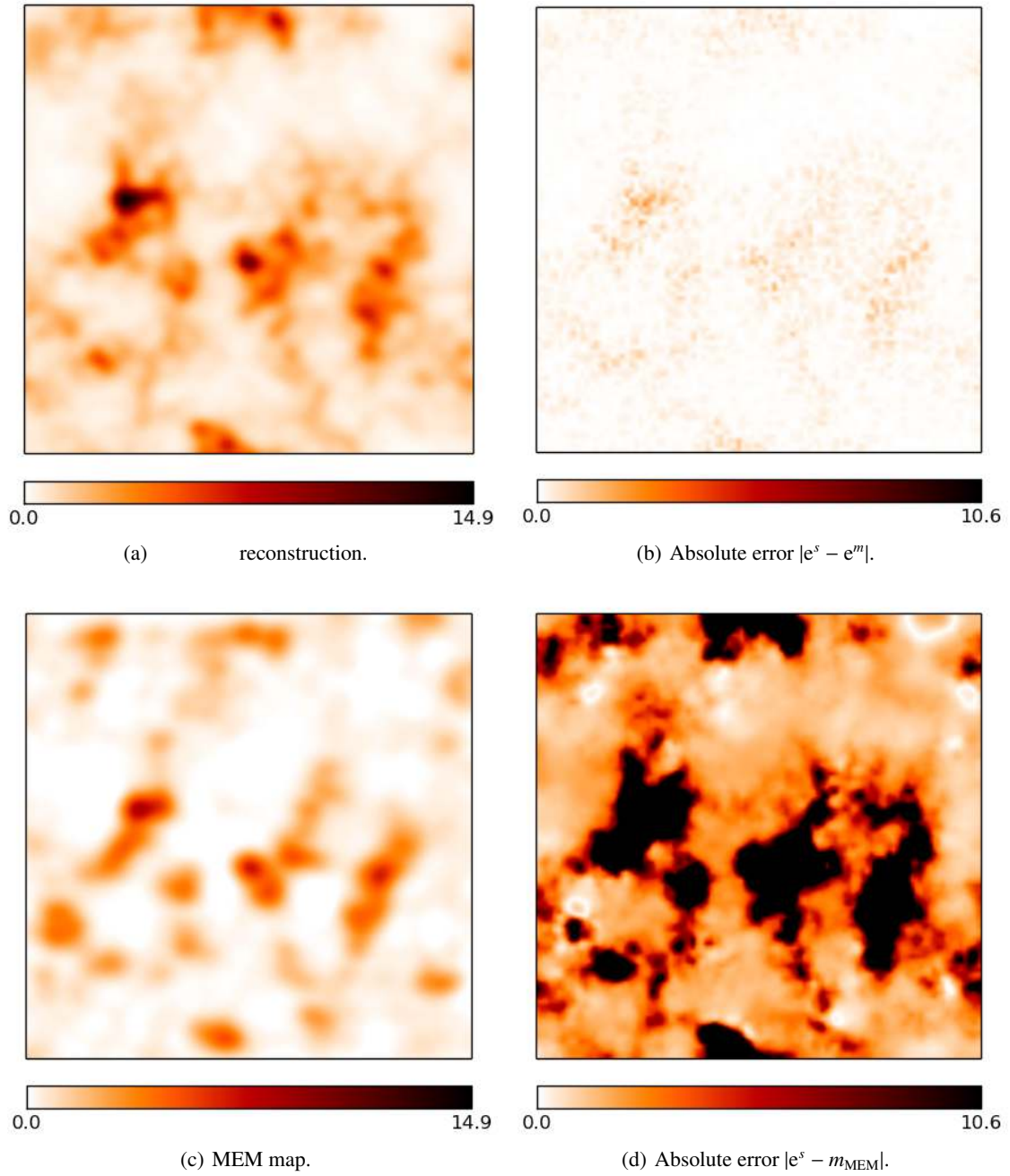


Fig. 6. Comparison of e^m with MEM for the simulated low-noise observation of Sect. 3.1. The images are 100^2 pixels large, the pixel size corresponds to roughly 0.2 arcsec. The brightness units are in Jy/px. The ridge-like structures simply stem from taking the absolute value and mark zero-crossings between positive and negative errors. *First row left:* e^m reconstruction. *First row right:* absolute per-pixel difference between the signal and the e^m reconstruction. *Second row left:* MEM reconstruction using the radio astronomical software package CASA. *Second row right:* absolute per-pixel difference between the signal and the MEM reconstruction.

covariance. In Appendix A.2, it is shown that a full signal estimate taking approximative uncertainty into account leads to

$$I \approx e^{m_x} \pm \sqrt{e^{2m_x} [e^{D_{xx}} - 1]}. \quad (28)$$

In Fig. 7, we present the following example of the approximated relative uncertainty

$$\sqrt{\frac{\langle (e^{s_x})^2 \rangle_{\mathcal{G}(m,D)} - \langle e^{s_x} \rangle_{\mathcal{G}(m,D)}^2}{\langle e^{s_x} \rangle_{\mathcal{G}(m,D)}^2}} = \sqrt{[e^{D_{xx}} - 1]} \quad (29)$$

for the low noise reconstruction of Sect. 3.1, together with the signal estimate, and absolute and relative difference map between signal and estimate. The subscripts indicate that our approach effectively approximates the full posterior with a Gaussian $\mathcal{G}(m, D)$ centered on the signal estimate and with a covariance of D (see Appendix A.2).

Figure 7 shows that the uncertainty follows the structure of the reconstruction. Where the signal is strong, the relative uncertainty is much lower than in regions that are mainly dominated by noise. A comparison between the estimated relative uncertainty and the real relative difference map shows the

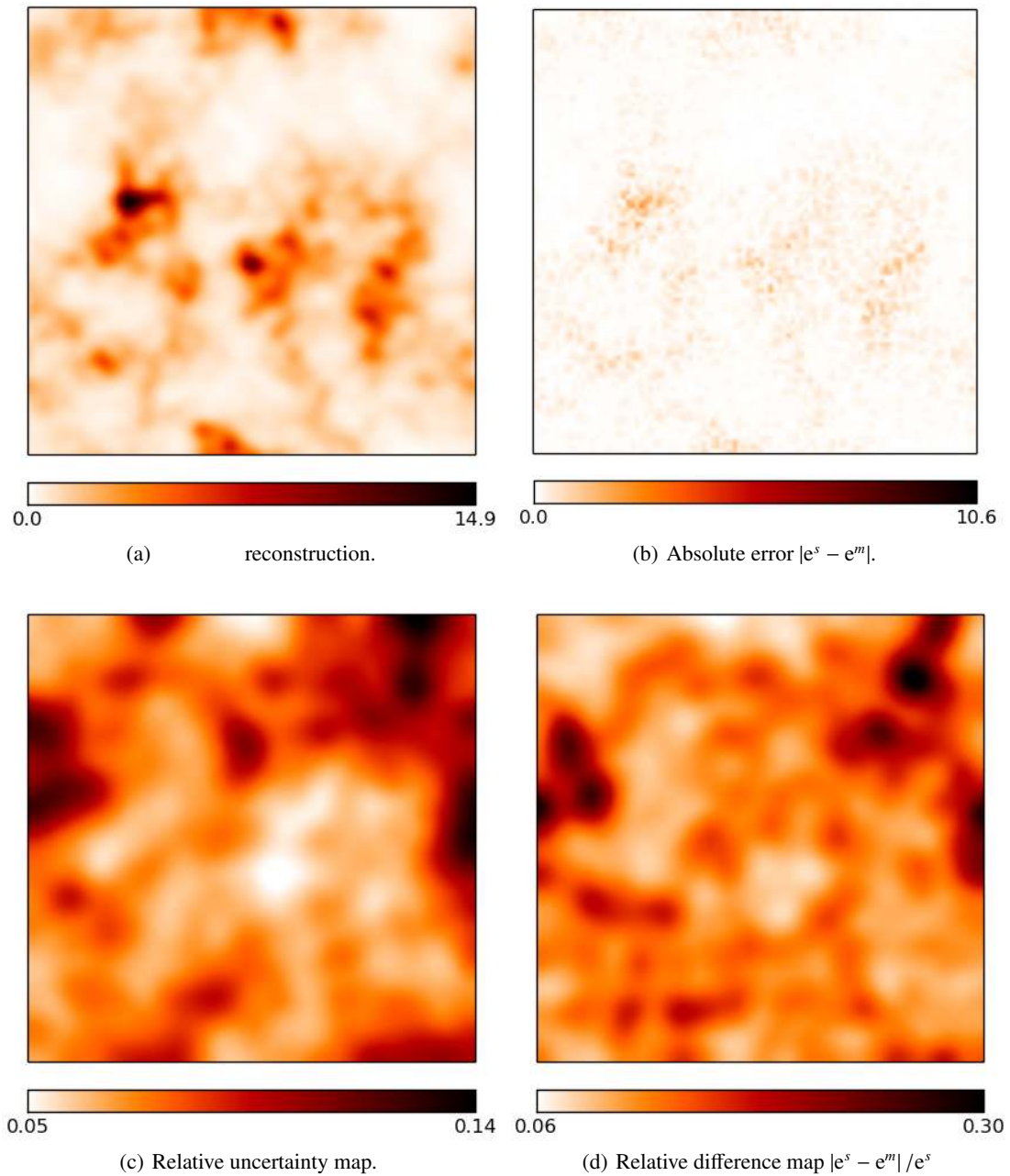


Fig. 7. *First row left:* reconstruction for the low-noise reconstruction of Sect. 3.1. *First row right:* absolute per-pixel difference between the signal and the reconstruction. The ridge-like structures simply stem from taking the absolute value and mark zero-crossings between positive and negative errors. *Second row left:* relative Uncertainty map derived from the reconstruction. *Second row right:* relative difference map between signal and reconstruction.

approximative nature of the theoretical estimate. While both maps agree nicely in structure, they do not fully match in terms of values. Overall, the theoretical uncertainty underestimates the real relative difference. However, the deviations between both maps are much stronger in the outer regions, where the signal is only weak. In the center of the map, where the source mainly is located, both agree relatively well.

If we further use (28) to calculate the absolute uncertainty for the low-noise reconstruction of Sect. 3.1, we find that roughly 40% of the original signal values lie within a 1σ region, and roughly 70% within a 2σ region. Although this result deviates from pure Gaussian expectations, this is a reasonable

outcome. Since the posterior is in general non-Gaussian, the assumption of posterior Gaussianity needed to exactly define (28) can only result in an approximation.

Calculating the uncertainty to a very high precision is computationally expensive⁶. It involves the probing of an implicitly defined matrix and a numerical algorithm to invert this matrix (see Appendix C). In this case, we stopped the stochastic probing

⁶ The estimation of the uncertainty goes roughly with $N_{\text{pr}}(O(\sqrt{n_s n_d}) + O(\sqrt{n_s n_s} \log(n_s)))$, where N_{pr} is the number of probes, n_d the number of visibility measurements, and n_s the number of pixels in image space (see Appendix C).

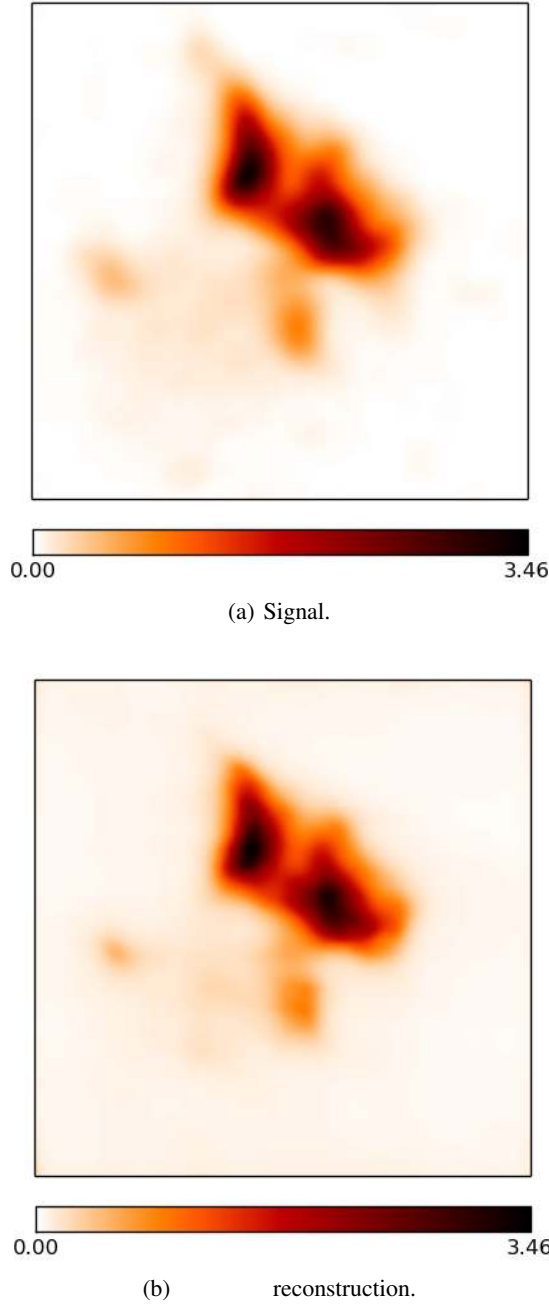


Fig. 8. Reconstruction of a signal field that was obtained from a CLEAN image of the real extended emission of Galaxy cluster Abell 2256. For the simulation, the same setup with low noise was used as in Sect. 3.1.

of D at some point for computational reasons and smoothed the outcome a bit to obtain Fig. 7. This might add to the deviations from pure Gaussian expectations on the absolute uncertainty, which we mentioned earlier. However, since the matrix representation of D theoretically enforces smoothness, this procedure should to some degree be an acceptable way to overcome numerical artifacts.

3.5. Power spectrum reconstructions

Until now, we have focused entirely on the reconstruction of signal maps. Now we discuss the reconstruction of the signal power spectrum that achieves automatically to infer the

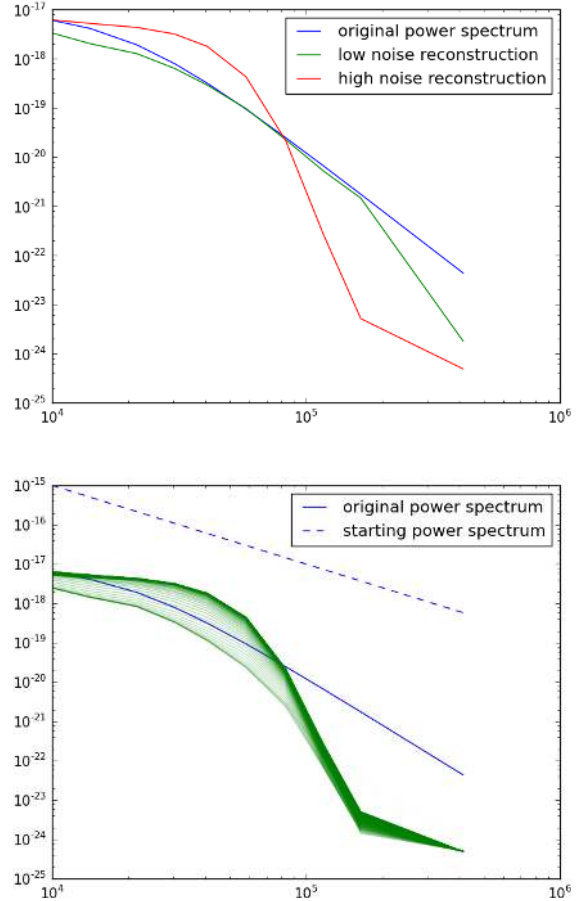


Fig. 9. *First panel:* power spectrum reconstruction for the simulated low-noise and high-noise observations of Sect. 3.1. *Second panel:* evolution of the high-noise power spectrum reconstruction over 80 iterations. The iteration process is indicated from transparent to full green.

best signal solution. The signal power spectrum is defined as the Fourier transformation of the autocorrelation function of the signal, assuming translationally and rotationally invariant statistics. We find

$$P(|k|) = \int dr C(r) \exp(ikr). \quad (30)$$

(for more details, see Sect. 2.3).

Qualitatively, it can be understood as decomposing the signal autocorrelation into its different contributions from various scales. High power on low Fourier modes means strong correlations on larger scales and high power on high Fourier modes means strong correlations on smaller scales.

In the first row of Fig. 9, we show the reconstruction of power spectra for the low- and high-noise reconstructions of Sect. 3.1. The figure shows the original power spectrum, which defines the correlation structure of the signal field, and the final results of after 6 iterations in the low, and 80 iterations in the high noise case. It can be seen that, with more noise, the reconstruction loses sensitivity for the smaller scales. This is reflected in the high-noise map reconstruction in Fig. 4, where the smallest scales are smoothed out by the algorithm.

The second row of Fig. 9 serves as an example for the actual reconstruction process, where all of the 80 iterations for the high-noise power spectrum are shown, together with the starting

guess, which was a simple and generic power law $P_{\text{sg}} \propto k^{-2}$. The power spectrum dropped first, and then slowly rose again. This is a consequence of a numerical procedure to ensure the convergence of the underlying nonlinear optimization routines, where a constant diagonal is first added to the uncertainty estimate D^{-1} used in the power spectrum reconstruction, and then suppressed again with converging iterations (see Appendix C).

We emphasize that an accurate power spectrum reconstruction can be a scientific result on its own and should not be regarded as a mere by-product. Since this is a rather unusual topic for observations of radio total intensity, it might be in place to explain a little further its meaning and to outline possible scientific merits.

The most typical physical source of extended emission in radio astronomy is synchrotron radiation. By spelling the power spectrum of the total intensity from some astronomical synchrotron source we effectively measure its correlation structure. Since synchrotron intensity is in part determined by the magnetic field strength (Rybicki & Lightman 1985) in the source, we automatically gather valuable scientific information on the magnetic field statistics as well, which gives

$$C_I(r) = \langle I(x)I(x+r) \rangle \propto \langle B(x)^2 B(x+r)^2 \rangle. \quad (31)$$

Detailed derivations of this and related statistical quantities, together with many discussions on its scientific use, mostly in the context of analyzing turbulent magnetic fields, can be found in a series of astrophysical papers (e.g., Spangler 1982, 1983; Eilek 1989; Waelkens et al. 2009; Junklewitz & Enßlin 2011; Oppermann et al. 2011a; Lazarian & Pogosyan 2012)

For future observations, it might be especially interesting to use these results from [Eq. \(31\)](#) to compare data of specific astrophysical synchrotron sources, e.g., supernova remnants or radio halos of galaxies and clusters, to simulations thereof. In simulations, the inputs are under control, and (31) can actually be calculated and compared with real data⁷

4. Conclusions

We presented a new approach to signal inference and imaging in radio astronomy and especially radio interferometry. The inference algorithm [Eq. \(22\)](#) is targeted to be optimal for the imaging of extended and diffuse radio sources in total intensity. In simulations, [Eq. \(22\)](#) demonstrated to produce high fidelity reconstructions of these extended signals, drawn from pure log-normal statistics or from real data. Comparisons showed that [Eq. \(22\)](#) can outperform current imaging algorithms in these tasks.

Furthermore, [Eq. \(22\)](#) is capable of producing an approximate uncertainty estimate for the inferred image through consistent propagation of measurement uncertainty. This is not possible with current imaging algorithms.

In addition to the inferred signal reconstruction, also estimates the power spectrum of the signal, i.e., its two-point correlation structure. The power spectrum is used for the signal reconstruction, but can be regarded as a new scientific outcome by itself. For instance, it opens opportunities to study the statistical properties of magnetic fields that lead to observed synchrotron emission. At the same time it offers a unique tool

⁷ It should be noted that for this, a log-normal power spectrum needs to be calculated from the reconstructed spectrum of the Gaussian field s . This can be done in a straight-forward way, (see e.g. Greiner & Enßlin 2015).

to compare simulations of turbulent, magnetoionic media in extended radio sources to observations.

It was shown that instead of using classical visibility weights directly, [Eq. \(22\)](#) chooses these internally, according to the ratio of reconstructed signal power to noise power. This is much in the spirit in which the robust weighting approach was originally conceived by Briggs (Briggs 1995b,a).

It should be noted, however, that obtaining all results with high accuracy, especially producing the uncertainty map, can be significantly more time consuming than traditional imaging methods because of the complicated numerical procedures necessarily involved to solve Eqs. ((23), (24)). Thus, more work is needed to obtain a more efficient implementation of the algorithm, examples include using a major/minor-cycle prescription as in standard imaging software, relaxing the usage of gridding operations in the response, using the most efficient libraries for all optimization algorithms, and developing a parallelized version for computer clusters or GPUs.

We only analyzed simulated data and reviewed the fundamental principles underlying [Eq. \(22\)](#). To simplify the analysis, we omitted some typical complexities of radio interferometers. However, the response operator R (see Eq. (2.2)), describing the act of observation, can easily be expanded to cover more effects, thereby adapting to the needs of the actual observational situation.

It is most straightforward to include the effects of a primary beam, as long as it is known accurately for the instrument in question. Also a direction- or time-dependent point spread function can be included without any further fundamental complications, although computational complexity would be considerably higher.

Furthermore, it should be highlighted that the inclusion of single dish data is almost readily possible. A radio interferometer is not sensitive to the largest scales of the sky brightness because it cannot measure at arbitrarily small uv -points, leaving a gap in the center of the uv -plane. This problem can in principle be overcome by combining the radio interferometric data with single dish observations on the same source. When using CLEAN-derived imaging algorithms, there always is a problem with the choice of the correct restoring beam, since it is not possible to trivially use the point spread function of the radio interferometer for the combined data. There is no problem like this with the imaging approach presented in this work.

The extension to multifrequency synthesis (see Eq. (6)) and polarization imaging is already being worked on and will be the subject of upcoming publications.

Another future topic is the possible inclusion of calibration into the framework. A first step could be to include the calibration errors into the error budget and use an approach similar to the extended critical filter (Oppermann et al. 2011b), where the noise covariance is subject to the inference itself. In principle, calibration itself can be understood as a reconstruction problem for which the presented methods could be useful. In the long run, the distinction between *calibration* and *imaging* is somewhat artificial and should ideally be merged into one step of complete reconstruction (see also Smirnov 2011a,b).

Finally, a future goal should be to extend the imaging algorithm [Eq. \(22\)](#) to a broader approach that can handle diffuse emission and point sources simultaneously (see, e.g., Selig & Enßlin 2015, for an example from photon count imaging). It could be worthwhile to think about merging the approaches of compressed sensing, where optimal imaging strategies for sparse signals are already known, with the presented Bayesian approach into which they could be included in form of a Laplacian prior.

Acknowledgements. We like to thank Niels Oppermann and Maksim Greiner for many helpful comments and discussions on statistics and numerics, Ashmeet Singh for extensive help with CASA, and Martin Reinecke and Jörg Knoche for computational support. H. Junklewitz thanks Rick Perley for the initial introduction to radio interferometry and its problems. Furthermore, we thank Oleg Smirnov, Annalisa Bonafede, and Chris Hales for suggestions and discussions on the radio astronomical aspects of the work and especially Tracy C. Clarke for the same, and for kindly providing the data on the galaxy cluster Abell 2256. This work was partly conducted within the DFG Research Unit 1254 “Magnetisation of Interstellar and Intergalactic Media” and profited from the framework of the Magnetism Key Science Project of LOFAR.

References

- Aharonian, F., Arshakian, T. G., Allen, B., et al. 2013, ArXiv e-prints [[arXiv:1301.4124](https://arxiv.org/abs/1301.4124)]
- Beatty, P. J., Nishimura, D. J., & Pauly, J. M. 2005, *IEEE Trans. Med. Imaging*, 24, 799
- Bhatnagar, S., & Cornwell, T. J. 2004, *A&A*, 426, 747
- Born, M., & Wolf, E. 1999, *Principles of Optics* (Cambridge, UK: Cambridge University Press)
- Bracewell, R. 1965, *The Fourier Transform and its applications* (USA: The McGraw-Hill Companies)
- Briggs, D. S. 1995a, in *BAAS*, 27, Am. Astron. Soc. Meet. Abstr., 112.02
- Briggs, D. S. 1995b, Ph.D. Thesis, New Mexico Tech, Socorro, USA
- Candes, E. J., Romberg, J. K., & Tao, T. 2006, *Comm. Pure Appl. Math.*
- Carrillo, R. E., McEwen, J. D., & Wiaux, Y. 2012, *MNRAS*, 426, 1223
- Carrillo, R. E., McEwen, J. D., & Wiaux, Y. 2013, *MNRAS*, 439, 3591
- Caticha, A. 2008, ArXiv e-prints [[arXiv:0808.0012](https://arxiv.org/abs/0808.0012)]
- Clark, B. G. 1980, *A&A*, 89, 377
- Clarke, T. E., & Enßlin, T. A. 2006, *AJ*, 131, 2900
- Cooley, J. W., & Tukey, J. W. 1965, *Math. Comp.*, 19, 297
- Cornwell, T. J. 2008, *IEEE Journal of Selected Topics in Signal Processing*, 2, 793
- Cornwell, T. J., & Evans, K. F. 1985, *A&A*, 143, 77
- Cornwell, T. J., Golap, K., & Bhatnagar, S. 2008, *IEEE Journal of Selected Topics in Signal Processing*, 2, 647
- Donoho, D. L. 2006, *IEEE Trans. Information Theory*, 52, 1289
- Eilek, J. A. 1989, *Bull. Am. Phys. Soc.*, 34, 1286
- Enßlin, T. 2013, in *AIP Conf. Ser.* 1553, ed. U. von Toussaint, 184
- Enßlin, T. A., & Frommert, M. 2011, *Phys. Rev. D*, 83, 105014
- Enßlin, T. A., & Weig, C. 2010, *Phys. Rev. E*, 82, 051112
- Enßlin, T. A., Frommert, M., & Kitaura, F. S. 2009, *Phys. Rev. D*, 80, 105005
- Finley, D. G., & Goss, W. M. 2000, *Radio interferometry: the saga and the science* (NRAO)
- Garrett, M. A. 2012, in *From Antikythera to the Square Kilometre Array: Lessons from the Ancients*, PoS(Antikythera & SKA)041
- Geman, S., & Geman, D. 1984, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 6, 721
- Greiner, M. 2013, Master Thesis, Ludwig-Maximilians-University Munich, Germany
- Greiner, M., & Enßlin, T. A. 2015, *A&A*, 574, A86
- Greisen, E. W. 1990, in *Acquisition, Processing and Archiving of Astronomical Images*, eds. G. Longo, & G. Sedmak, 125
- Gull, S. F., & Daniell, G. J. 1979, in *IAU Colloq. 49: Image Formation from Coherence Functions in Astronomy*, ed. C. van Schooneveld, *Astrophys. Space Sci. Lib.*, 76, 219
- Hastings, W. K. 1970, *Biometrika*, 57, 97
- Högbom, J. A. 1974, *A&AS*, 15, 417
- Huang, K. 1963, *Statistical Mechanics* (New York: John Wiley)
- Jasche, J., Kitaura, F. S., Wandelt, B. D., & Enßlin, T. A. 2010, *MNRAS*, 406, 60
- Jaynes, E. T. 2003, *Probability Theory: The Logic of Science* (Cambridge University Press)
- Junklewitz, H., & Enßlin, T. A. 2011, *A&A*, 530, A88
- Karakci, A., Sutter, P. M., Zhang, L., et al. 2013, *ApJS*, 204, 10
- Lazarian, A., & Pogosyan, D. 2012, *ApJ*, 747, 5
- Mood, A. M., Graybill, F. A., & Duane, C. B. 1974, *Introduction to the theory of statistics* (McGraw Hill)
- Neal, R. M. 1993, Technical Report CRG-TR-93-1, Dept. of Computer Science, University of Toronto
- Oppermann, N., Junklewitz, H., Robbers, G., & Enßlin, T. A. 2011a, *A&A*, 530, A89
- Oppermann, N., Robbers, G., & Enßlin, T. A. 2011b, *Phys. Rev. E*, 84, 041118
- Oppermann, N., Selig, M., Bell, M. R., & Enßlin, T. A. 2013, *Phys. Rev. E*, 87, 032136
- Peskin, M. E., & Schroeder, D. V. 1995, *An Introduction to Quantum Field Theory* (Westview Press)
- Rau, U., & Cornwell, T. J. 2011, *A&A*, 532, A71
- Rau, U., Bhatnagar, S., Voronkov, M. A., & Cornwell, T. J. 2009, *IEEE Proc.*, 97, 1472
- Reid, R. I., & CASA Team 2010, *BAAS*, 42, 568
- Rybicki, G. B., & Lightman, A. P. 1985, *Radiative processes in astrophysics* (Wiley Online Library)
- Ryle, M., & Hewish, A. 1960, *MNRAS*, 120, 220
- Sault, R. J., & Oosterloo, T. A. 2007, ArXiv e-prints [[arXiv:astro-ph/0701171](https://arxiv.org/abs/astro-ph/0701171)]
- Sault, R. J., & Wieringa, M. H. 1994, *A&AS*, 108, 585
- Schwab, F. R. 1984, *AJ*, 89, 1076
- Selig, M., & Enßlin, T. 2015, *A&A*, 574, A74
- Selig, M., Bell, M. R., Junklewitz, H., et al. 2013, *A&A*, 554, A26
- Selig, M., Oppermann, N., & Enßlin, T. A. 2012, *Phys. Rev. E*, 85, 021134
- Skilling, J., Strong, A. W., & Bennett, K. 1979, *MNRAS*, 187, 145
- Smirnov, O. M. 2011a, *A&A*, 527, A106
- Smirnov, O. M. 2011b, *A&A*, 527, A107
- Spangler, S. R. 1982, *ApJ*, 261, 310
- Spangler, S. R. 1983, *ApJ*, 271, L49
- Sutter, P. M., Wandelt, B. D., & Malu, S. S. 2012, *ApJS*, 202, 9
- Sutter, P. M., Wandelt, B. D., McEwen, J. D., et al. 2014, *MNRAS*, 438, 768
- Taylor, G. B., Carilli, C. L., & Perley, R. A. 1999, *Synthesis Imaging in Radio Astronomy II*, *ASP Conf. Ser.*, 180
- Thompson, A. R., Moran, J. M., & Swenson, G. W. 1986, *Interferometry and synthesis in radio astronomy* (Germany: Wiley-VCH Verlag)
- Transtrum, M. K., & Sethna, J. P. 2012, ArXiv e-prints [[arXiv:1201.5885](https://arxiv.org/abs/1201.5885)]
- Waelkens, A. H., Schekochihin, A. A., & Enßlin, T. A. 2009, *MNRAS*, 398, 1790
- Wiaux, Y., Jacques, L., Puy, G., Scaife, A. M. M., & Vanderghynst, P. 2009, *MNRAS*, 395, 1733

Appendix A: Derivation of RESOLVE

For a complete derivation of [RESOLVE](#), we first provide some general remarks, and then divide the section into two parts, where we derive a Maximum a posteriori solution for the signal field, and for its power spectrum.

From Sect. 2, we recall the basic premises of the inference problem to be solved. We want to find the statistically optimal reconstruction of the total intensity signal I given a data model,

$$d = RI + n = Re^s + n, \quad (\text{A.1})$$

under the assumptions that

- I follows log-normal statistics, such that $s = \log I$ follows Gaussian statistics;
- the noise n follows Gaussian statistics as well;
- and R models the linear response of a radio interferometer (see Eq. (2.2) in Sect. 2).

Under these assumptions the likelihood $\mathcal{P}(d|s)$ and the signal prior $\mathcal{P}(s)$ take the following form as was shown in (11)

$$\begin{aligned} \mathcal{P}(d|s) &= \mathcal{G}(d - Re^s, N) \\ &= \frac{1}{\det(2\pi N)^{1/2}} e^{-1/2 ((d - Re^s)^\dagger N^{-1} (d - Re^s))}, \end{aligned} \quad (\text{A.2})$$

$$\begin{aligned} \mathcal{P}(s) &= \mathcal{G}(s, S) \\ &= \frac{1}{\det(2\pi S)^{1/2}} e^{-1/2 (s^\dagger S^{-1} s)}. \end{aligned} \quad (\text{A.3})$$

Then, the posterior of s

$$\mathcal{P}(s|d) \propto \mathcal{G}(d - Re^s, N) \mathcal{G}(s, S) \quad (\text{A.4})$$

can become highly non-Gaussian due to the nonlinearity introduced by (A.1).

As a further complication, we have to assume a priori that the signal covariance $S = \langle s^\dagger s \rangle$ is unknown. Assuming statistical homogeneity and isotropy for the signal statistics, we parameterize its power spectrum $P(k)$ as a decomposition into spectral parameters p_i and positive projection operators $S^{(i)}$ onto a number of spectral bands such that the bands fill the complete Fourier domain

$$S = \sum_i p_i S^{(i)}. \quad (\text{A.5})$$

consists of two inference steps to solve the main problem (12) iteratively for s and all p_i . We fully describe both steps individually in the following subsections.

A.1. Reconstruction of the signal field s

For the reconstruction of the signal field s , we assume the power spectrum parameters p_i to be known from a previous inference step. This can formally be expressed by marginalizing over them while assuming a delta distribution for the known parameters p^*

$$\begin{aligned} \mathcal{P}(s|d, p^*) &= \int \mathcal{D}p \mathcal{P}(s|d, p) \mathcal{P}(p|p^*) \\ &= \int \mathcal{D}p \mathcal{P}(s|d, p) \delta(p - p^*). \end{aligned} \quad (\text{A.6})$$

For convenience, we rewrite our notation to work with the Hamiltonian $H(s, d)$ instead of the posterior $\mathcal{P}(s|d)$,

$$\mathcal{P}(s|d) := \frac{e^{-H(d, s)}}{Z} \quad (\text{A.7})$$

with $Z := \mathcal{P}(d)$. This effectively expresses our problem in more familiar terms of statistical physics, while the Hamiltonian $H(s, d) = -\log(\mathcal{P}(d|s)\mathcal{P}(s))$ still comprises all important signal-dependent terms and is usually easier to handle than the posterior.

The Hamiltonian of problem (A.4) reads

$$\begin{aligned} H(s, d) &= -\log(\mathcal{G}(d - Re^s, N) \mathcal{G}(s, S)) \\ &= \frac{1}{2} s^\dagger S_p^{-1} s + \frac{1}{2} (e^s)^\dagger M e^s - j^\dagger e^s + H_0, \end{aligned} \quad (\text{A.8})$$

where $j = R^\dagger N^{-1} d$, $M = R^\dagger N^{-1} R$ and H_0 summarizes all terms that are not dependent on the signal s .

Using the Gibbs free energy ansatz of [EnBlin & Weig \(2010\)](#), [Oppermann et al. \(2013\)](#) have shown that it is possible to rederive the critical filter for this Hamiltonian. However, in practice, it is only solvable under the assumption of a diagonal M in signal space. Otherwise we would be forced to explicitly compute arbitrary components of the very large matrix of size n_s^2 , representing the operator M , which is computational infeasible. Unfortunately, for the response under consideration here (2.2), with an incomplete sampling of the Fourier plane in data space, M is not diagonal in general.

Thus, we instead use the MAP principle to solve the inference problem for s . Maximizing the posterior readily translates to minimizing the Hamiltonian (A.7). If we take the derivative of the Hamiltonian (A.8) with respect to the signal field s and set it to zero, we get

$$\frac{\delta H(s)}{\delta s} = S_p^{-1} s + e^s M e^s - j e^s = 0. \quad (\text{A.9})$$

This is a high dimensional, nonlinear equation, which can be solved numerically using an iterative optimization algorithm, in our case a steepest descent method. We call the solution of this equation $m = \text{argmax}_s \mathcal{P}(s|d)$.

The solution m is an estimate for the Gaussian field s . To calculate a signal estimate \hat{I} for the original log-normal signal $I = e^s$, we just take the exponential of m

$$\hat{I} = e^m. \quad (\text{A.10})$$

A.2. Uncertainty of the signal reconstruction

A full statistical analysis involves accounting for the uncertainty of the signal estimate. For this, we use the information encoded in the second posterior moment (or covariance) $D = \langle (s - m)(s - m)^\dagger \rangle$ as a measure of the expected uncertainty of the signal reconstruction. Within the MAP approach, we approximate the inverse posterior covariance D^{-1} with the second derivative of the Hamiltonian

$$\begin{aligned} D^{-1} &\approx -\frac{\delta^2 H(s)}{\delta s_x \delta s_y} \Big|_{s=m} = S_p^{-1}{}_{xy} + e^{s_x} M_{xy} e^{s_y} \\ &\quad + e^{s_y} \int dz M_{xz} e^{s_z} - j_x \cdot e^{s_x} \delta_{xy}, \end{aligned} \quad (\text{A.11})$$

which needs to be inverted numerically in practice. In this way, we effectively assume that the real signal posterior is approximated with a Gaussian $\mathcal{G}(m, D)$. Unfortunately, D only approximates the posterior covariance of the Gaussian field m . We need to translate this into a posterior covariance for the full estimate $\hat{I} = e^m$.

If the signal posterior *were* exactly Gaussian, we could just assume our posterior estimate to be of exact log-normal statistics, solve for the mean and variance analytically and thus write

$$\langle e^{s_x} \rangle_{\mathcal{G}(m,D)} = e^{m_x + \frac{1}{2} D_{xx}} \quad (\text{A.12})$$

$$\langle (e^{s_x})^2 \rangle_{\mathcal{G}(m,D)} - \langle e^{s_x} \rangle_{\mathcal{G}(m,D)}^2 = e^{2m_x + D_{xx}} [e^{D_{xx}} - 1], \quad (\text{A.13})$$

using the definitions for the mean and variance of a log-normal distribution (see ,e.g., [Mood et al. 1974](#)). But since the posterior is *not* Gaussian in general, we cannot solve Eqs. ((A.12), (A.13)) analytically. This was, in the first place, the reason why uses the MAP approach (see Sect. 2.3). Nevertheless, since we effectively approximate the full posterior with a Gaussian $\mathcal{G}(m, D)$ when using Eq. (A.11) as the posterior covariance, one might be tempted to just use Eqs. ((A.12), (A.13)) anyhow.

However, in practice, it turns out that within the MAP approach this procedure is prone to overestimating signal estimate and its uncertainty. This is because usually the maximum of a log-normal distribution lies *above* its mean (for details see [Greiner 2013](#)). We thus drop the extra terms of D in the argument of the exponentials in Eqs. ((A.12), (A.13)), keep (A.10), and write

$$\hat{I}_x = e^{m_x} \pm \sqrt{e^{2m_x} [e^{D_{xx}} - 1]} \quad (\text{A.14})$$

if we want to account for the uncertainty in the reconstruction.

A.3. Reconstruction of the power spectrum parameters p

In the second step of [A.2](#), we assume to have a solution for m and D from the last iteration and estimate the unknown spectral parameters p from the signal-marginalized probability of data and power spectrum $\mathcal{P}(p, d)$:

$$\begin{aligned} \mathcal{P}(p, d) &= \int \mathcal{D}s \mathcal{P}(s, d|p) \mathcal{P}(p) \\ &= \int \mathcal{D}s \mathcal{G}(d - \text{Re}^s, N) \mathcal{G}(s, S_p) \mathcal{P}(p). \end{aligned} \quad (\text{A.15})$$

This approach was first derived in [Oppermann et al. \(2013\)](#) for Gaussian signal fields. We closely follow their argument and also show its approximate validity for log-normal fields.

To do this, we first need to define a prior for the power spectrum parameters p . In this, we follow [Enßlin & Frommert \(2011\)](#), [Enßlin & Weig \(2010\)](#), and [Oppermann et al. \(2013\)](#), and choose independent inverse-gamma distributions for each spectral parameter p_i ,

$$\begin{aligned} \mathcal{P}(p) &= \prod_i \mathcal{P}_{\text{IG}}(p_i) \\ &= \prod_i \frac{1}{q_i \Gamma(\alpha_i - 1)} \left(\frac{p_i}{q_i} \right)^{-\alpha_i} \exp\left(-\frac{q_i}{p_i}\right), \end{aligned} \quad (\text{A.16})$$

where $\Gamma(\cdot)$ denotes the gamma function, q_i defines an exponential cutoff in the prior for low values of p_i , and α_i is the slope of the power-law decay for large values of p_i . In principle, by tuning these parameters, the prior can be adapted according to the a priori knowledge about the power spectrum. Usually, we use the limits of $q_k \rightarrow 0$ and $\alpha_k \rightarrow 1$ for all k . This turns the inverse-gamma prior into Jeffreys prior ([Jaynes 2003](#)), which is flat on a logarithmic scale. In some tests though, we have allowed for nonunity α_k parameters to suppress unmeasured Fourier modes.

During the reconstruction of the power spectrum, we additionally introduce a smoothness prior as developed by [Oppermann et al. \(2013\)](#) to punish most probably unphysical and numerically unwanted random fluctuations in the power spectrum. In that prescription, the inverse-gamma prior (A.16) is augmented with a probability distribution that enforces smoothness of the power spectrum

$$\mathcal{P}(p) = \mathcal{P}_{\text{sm}}(p) \prod_k \mathcal{P}_{\text{IG}}(p_k). \quad (\text{A.17})$$

The spectral smoothness prior can be written as a Gaussian distribution in $\tau = \log p$:

$$\begin{aligned} \mathcal{P}_{\text{sm}}(p) &\propto \exp\left(-\frac{1}{2\sigma_p^2} \int d(\log k) \left(\frac{\partial^2 \log p_k}{\partial (\log k)^2} \right)^2\right) \\ &\propto \exp\left(-\frac{1}{2} \tau^\dagger T \tau\right), \end{aligned} \quad (\text{A.18})$$

where the differential operator T includes the second derivative of $\tau = \log p$ and a scaling constant σ_p^2 that determines how strict the smoothness should be enforced. This particular form of the prior favors smooth power-law spectra. For all details we refer to ([Oppermann et al. 2013](#)).

As was shown there, the corresponding inverse-gamma prior for the τ parameters can easily be derived from the conservation of probability under transformations

$$\begin{aligned} \mathcal{P}(\tau) &= \mathcal{P}(p) \left| \frac{dp}{d\tau} \right| \\ &= \prod_i \frac{q_i^{\alpha_i - 1}}{\Gamma(\alpha_i - 1)} e^{-(\alpha_i - 1)\tau_i + q_i e^{-\tau_i}}. \end{aligned} \quad (\text{A.19})$$

With this prior, we can calculate the signal-marginalized joint probability (A.15) if we apply one crucial approximation. Since $\mathcal{P}(s, d|\tau)$ in (A.15) is non-Gaussian because of the high nonlinearity of the $e^{(d - \text{Re}^s)}$ – terms, we cannot just move on analytically. We instead use a saddle point method and approximate the argument of the exponential occurring in $\mathcal{P}(s, d|\tau)$, which can be written as $e^{-H(s,d)}$ using (A.7). To perform the saddle point approximation, we replace $H(s, d)$ with its Taylor expansion up to second order around the maximum of the Posterior m , derived in the previous iteration of the signal reconstruction, i.e.,

$$\begin{aligned} e^{-H(s,d)} &\propto e^{(-\frac{1}{2}(d - \text{Re}^s)^\dagger N^{-1} (d - \text{Re}^s) - \frac{1}{2} s^\dagger S_\tau^{-1} s)} \\ &\approx e^{(H(m) + \frac{1}{2}(s-m)^\dagger D(m)^{-1} (s-m))} \end{aligned} \quad (\text{A.20})$$

This effectively approximates the non-Gaussian signal posterior $\mathcal{P}(s, d|\tau)$ with a Gaussian with mean m and covariance D . We note that this procedure is similar to a mean field approximation in statistical physics ([Huang 1963](#)).

With this approximation, we can solve the marginalization integral in (A.15) and calculate $\mathcal{P}(\tau, d)$, or alternatively the Hamiltonian,

$$\begin{aligned} H(d, \tau) &= -\log \mathcal{P}(d, \tau) \\ &= -\log \int \mathcal{D}s \mathcal{G}(d - \text{Re}^s, N) \mathcal{G}(s, S) \mathcal{P}(p) \\ &\approx \frac{1}{2} \text{tr}(\log S_\tau) - \frac{1}{2} \text{tr}(\log D_\tau) + H(m, \tau) \\ &\quad + \sum_i ((\alpha_i - 1) \tau_i + q_i e^{-\tau_i}) + \frac{1}{2} \tau^\dagger T \tau + H_0, \end{aligned} \quad (\text{A.21})$$

where we have used the matrix theorem $\log |S| = \text{tr}(\log S)$, and have collected all terms not depending on τ into a constant H_0 .

Taking the derivative of (A.21) with respect to one parameter τ_i and replacing $p_i = e^{\tau_i}$, we find

$$p_i = \frac{q_i + \frac{1}{2} \text{tr}((mm^\dagger + D)S^{(i)})}{\alpha_i - 1 + \frac{q_i}{2} + (T \log p)_i}. \quad (\text{A.22})$$

With this equation we can update the power spectrum parameters for each iteration using the current m and D .

This is in perfect accordance with previous findings (Enßlin & Frommert 2011; Enßlin & Weig 2010; Oppermann et al. 2013) and shows effectively that we can rediscover the critical filter for a pure MAP approach if we accept the approximation (A.20) as valid.

A.4. RESOLVE and image weighting

In aperture synthesis, imaging is usually combined with a weighting scheme that is included in the Fourier inversion of the visibilities. Essentially, the term W in (3), defining the dirty image I^D , can be expanded to hold more factors than the mere sampling function

$$I^D = \mathcal{F}^{-1}(T \cdot B \cdot w \cdot S \cdot \mathcal{F}I), \quad (\text{A.23})$$

with $W = T \cdot B \cdot w \cdot S$, where T is a possible tapering of outer visibilities, B is a user-chosen baseline weighting, w are the statistical noise weights obtained from an analysis of the thermal noise, and S is the sampling function. In this section, we prove that w implicitly converges to a meaningful set of weights.

Historically, mainly two weighting schemes have been employed. *Natural* weighting just multiplies every visibility point with the inverse thermal noise variance for the particular baseline and is therefore a simple, noise-dependent down-weighting mechanism. *Uniform* weighting ensures that the weight per gridded visibility cell is constant and, hence, effectively gives higher weight to outer baselines, where usually less visibility points are found in a grid cell.

In a seminal work (Briggs 1995b), Briggs has shown that natural weighting can be obtained under the constraint that the sample variance of the image should be minimized. In contrast, uniform weighting can be shown to reduce sidelobe levels, but actually downgrades sensitivity at the same time.

In the same work, a new weighting scheme was devised that interpolates between these two extremes, called *robust* weighting. The robust weights are determined as

$$W(k) \propto \frac{1}{1 + \sigma^2(k)/s_p^2(k)}, \quad (\text{A.24})$$

where σ^2 is the thermal noise variance, and s_p^2 is some parameter that originally was derived having in mind some measure of the source power at the given visibility (Briggs 1995b). In practice, s_p^2 is usually adjusted by hand to meet the needs of the astronomer for having a tradeoff between sensitivity and resolution.

This form of weighting can be explained within the presented Bayesian framework, and, furthermore, we show that an algorithm like w naturally converges to optimal robust-weighting-like parameters according to the ratio of estimated noise and signal power.

For this, we consider the negative logarithm of the posterior (15), i.e., the Hamiltonian of our inference problem (see Eq. (A.4) in Appendix A for details),

$$H(s, d) = \frac{1}{2} s^\dagger S^{-1} s + \frac{1}{2} (e^s)^\dagger M e^s - j^\dagger e^s + H_0. \quad (\text{A.25})$$

We can expand the exponents in a Taylor series and separate the quadratic from the higher orders in s as we have done in (16)

$$H(s, d) = \frac{1}{2} s^\dagger (S^{-1} + M) s - s^\dagger j + H_0 + \sum_{k=3}^{\infty} \frac{1}{k!} \Delta(M, j)_{x_1 \dots x_k}^k s_{x_1} \dots s_{x_k}. \quad (\text{A.26})$$

If we now apply the MAP principle and set the derivative with respect to s to zero, we find

$$(S^{-1} + M) s - j + \Delta(M, j, s) = 0 \quad (\text{A.27})$$

where we have defined $\Delta(M, j, s) = \frac{\delta}{\delta s} \sum_{k=3}^{\infty} \frac{1}{k!} \Delta(M, j)_{x_1 \dots x_k}^k s_{x_1} \dots s_{x_k}$.

We can partly solve this equation for s :

$$s = (S^{-1} + M)^{-1} j - (S^{-1} + M)^{-1} \Delta(M, j, s). \quad (\text{A.28})$$

The first term is the analytic solution to the quadratic part of the full log-normal Hamiltonian. It was shown to be equivalent to a Wiener Filter applied to the data d (Enßlin et al. 2009), which would be the optimal solution for a purely Gaussian signal field.

Using (20) for the covariance matrices S and N and $j = R^\dagger N^{-1} d$, we can write the Wiener Filter operator in (A.28), $F = (S^{-1} + M)^{-1} R^\dagger N^{-1}$, in Fourier space:

$$F(k) = \frac{1}{1 + P_n^g(k)/P_s(k)}, \quad (\text{A.29})$$

where $P_n^g = G_{ku} P_n(u)$ is the noise power spectrum on the regular grid, defined by the gridding operator G .

This has the exact same form as the definition of the robust weights (A.24), and even the original premise is fulfilled that the factor s_p^2 in (A.24) should be connected to the source power. The great difference is that the Wiener Filter naturally weights each mode in Fourier space differently, given that the signal power spectrum $P_s(k)$ is known.

Since w reconstructs this power spectrum, it is capable of doing this type of weighting, as is every algorithm that leads to an equation like (A.28) and simultaneously gets information about the signal power spectrum. Of course, w solves (A.28) iteratively, and only the converged solution gives optimal weights for the log-normal inference problem. No simple and direct equivalence can be given between these effective weights and robust weighting. It is not even meaningful to write them down explicitly since the sum in $\Delta(M, j, s)$ in principle extends infinitely.

We conclude that the classical robust weighting can theoretically be understood as the optimal solution to a signal reconstruction problem of a Gaussian signal field, equivalent to a Wiener Filter operation, and that algorithms invoking higher orders of signal statistics than the Wiener Filter, like w , do an extended weighting operation modified by the higher statistical moments. In fact, this similarity between the robust weights and Wiener Filtering was already mentioned by Briggs himself (Briggs 1995b), although in that work, no clear explanation of the connection was given.

Appendix B: Technical supplements

B.1. Mathematical notation

All physical quantities and functions common to radio astronomical imaging are represented in a basis-free notation as vectors and operators, defined on an arbitrary-dimensional functional vector space. Operations on vectors are represented by inner products, appropriately defined for discrete and continuous spaces as:

$$\begin{aligned}
 - \text{discrete space: } a^\dagger b &:= \sum_x \overline{a_x} b_x, \\
 - \text{continuous space: } a^\dagger b &:= \int dx \overline{a(x)} b(x) dx, \quad (\text{B.1})
 \end{aligned}$$

where the \dagger symbol stands for a transposing operation (and a possible complex conjugation in case of a complex vector). In contrast, where needed explicitly, the \cdot symbol will denote component-wise multiplication, so that $(a \cdot b)_x = a(x) b(x)$.

We now can effortlessly combine discrete and continuous quantities in our notation. This is important, since, in real observations, the visibility V_k is always a function defined over a discrete, complex Fourier space, spanned by n_d measurements, whereas the sky brightness I_x is in principle a continuous function, defined over an infinitely large, real space. Of course, on the computer, a continuous space needs to be discretized again, but the assumption is that, within needed accuracy, discretization still allows for a theoretical description of quantities like I_x as continuous fields. This assumption is frequently made in computational field theory as well (see, e.g., [Peskin & Schroeder 1995](#)). For a more thorough discussion on the framework of signal inference, see [Enßlin et al. \(2009\)](#).

If the inner product actually is just a discretized version of a continuous one, a volume factor needs to be included in the sum, which, for clarity, is explicitly omitted in this study for all equations. In computational practice though, this is unavoidable, since all quantities effectively become discrete when finally calculated on a computer (for details see [Selig et al. 2013](#)).

B.2. RESOLVE and standard imaging procedures

operates under some different concepts from standard imaging procedures in radio astronomy. The greatest difference comes about because of the nature of the image reconstruction as a Bayesian statistical estimator. It is only completely interpretable when considered together with its uncertainty. Because of the unavailability of a proper image uncertainty in virtually all standard methods, most notably in CLEAN, images are usually interpreted differently from how they should be with . Furthermore, some features that usually are entirely set by the user are implicitly achieved in . What follows is a brief list of important points and issues to be considered when interpreting in terms of standard notions. It should hopefully serve to help reconnect to well-known procedures in imaging when using .

- *deconvolution* of the dirty beam is the most obvious problem in radio imaging (see Eq. (3)). An illustration as to how achieves this deconvolution needs to take different points into account. For one, the multiplicative term e^m in the fix-point Eq. (23) acts effectively as a convolution kernel in uv -space. This enforces some amount of smoothness in the visibility structures. This smoothness is exploited by

for extrapolating the measured visibilities into the regions of uv -space without direct measurements. A more complex explanation of course also needs to encapsulate the effect of the reconstructed power spectrum on this smoothing scale. A more general explanation draws from the fact that the proposed Bayesian log-normal reconstruction is by design data driven. This means that the algorithm much more comfortably adjusts side-lobe structures to strong emission regions roughly present in the data, as opposed to the effectively suppressed case of leaving those structures as fainter emission regions in the final signal estimate.

- *image resolution* is handled by in a slightly different way. Because of the inherent robust-like image weighting of (see A.4) and its capabilities of extrapolating in unmeasured regions of uv -space, the algorithm automatically sets an optimal resolution scale, which is even capable of achieving some degree of superresolution. This resolution scale cannot be simply predicted beforehand. After the algorithm has been applied though, the achieved resolution can be estimated. A comparison of the raw reconstructed signal power mm^\dagger with its uncertainty D in Fourier space reveals the Fourier modes for which the reconstruction becomes uncertain and is smoothed because of the implicit robust weighting. For a strictly conservative approach, it is of course always possible to smooth the final image with an instrument resolution kernel, should the superresolution be in doubt. In contrast, classical image weights or tapering on top of should not simply be applied without further consideration. They represent an additional filter operation, which may (or may not) let to diverge. In any case, all additional weighting terms should be implemented into the response operator.
- *residual images* are usually defined as the inverse Fourier transform of the difference between the visibility data and the reconstructed image $I_{\text{res}} := \mathcal{F}^{-1}(d - \mathcal{F}m)$. They are very often used as a diagnostic for the quality of image reconstruction and to check for known patterns of reconstruction errors. Usually, algorithms are required to get as close to a pure noise-like residual as possible (see ,e.g., [Bhatnagar & Cornwell 2004](#)). Although meets these criteria for the presented simulated data (see Fig. 3), some caution is advised for. In general, a residual image alone is not necessarily the best quality measure for , nor should it necessarily be expected to be close to a pure noise background image in every case. The algorithm was designed to find a MAP estimate for the signal as an approximation to a posterior mean, and there is in general no reason that this estimate alone needs to be equivalent to an image with a residual that resembles pure background noise. It should be more reliable to consider the residual together with the reconstructed uncertainty, where the range of residual images consistent with the image and its uncertainty should encompass a pure noise image outcome.
- *image rms-noise and dynamic range* are typical quantifiers of radio astronomical image properties. For a meaningful application to images, again, the image uncertainty needs to be taken into account. As explained above for the achieved resolution, smoothes small scale noisy features in low signal-to-noise regions, while increasing uncertainty. Thus, calculating an rms value from these regions simply from the reconstructed image might be strongly biased. Instead, using the range of values, consistent with the uncertainty, $I = e^m + \sqrt{e^{2m} [e^D - 1]}$ should give correct results, where a conservative approach would

use lower bounds for the peak value and upper bounds for rms-estimations.

B.3. Bayesian derivation of MEM

As outlined in Sect. 3, MEM aims to maximize the image entropy S_{im} ,

$$S_{\text{im}} = - \int dx s(x) \log(s(x)/m(x)), \quad (\text{B.2})$$

where $m(x)$ is a model image of the observed signal, thus allowing us to introduce some kind of prior information into the problem. The data enter this formalism as a constraint for the maximization problem. Usually, one adds a term to (B.2) that measures the closeness of the entropic signal reconstruction to the data in the form of a $\chi^2(d, Rs)$ distribution, which is nothing else but the log-likelihood of (11):

$$\begin{aligned} \frac{1}{2}\chi^2(d, Rs) &= \frac{1}{2}(d - Rs)^\dagger N^{-1}(d - Rs) \\ &= -\log(P(d|s)) + \text{const.} \end{aligned} \quad (\text{B.3})$$

With (27) and (B.3), MEM achieves a solution by extremizing

$$J(d, s) = -\log P(d|s) - \mu S_{\text{im}} \quad (\text{B.4})$$

for s . The multiplier μ is usually adjusted during the extremization to meet numerical constraints (see Cornwell & Evans 1985, for details).

We now repeat a short section from (Enßlin & Weig 2010), analyzing the assumptions of this approach from the viewpoint of Bayesian signal inference.

As we have identified (11) as the log-likelihood, it is also possible to re-identify the prior distribution. If we interpret $J(d, s)$ as a Hamiltonian $H(d, s)$, than the entropy term can be understood as a log-prior

$$\mu S_{\text{im}}(s) = \log \mathcal{P}(s). \quad (\text{B.5})$$

With this, we can read off the underlying prior distribution implicitly assumed in MEM,

$$\begin{aligned} \mathcal{P}(s) &= \exp \left[-\mu \int dx s(x) \log \left(\frac{s(x)}{m(x)} \right) \right] \\ &= \prod_x \left(\frac{s(x)}{m(x)} \right)^{-\mu s(x)}. \end{aligned} \quad (\text{B.6})$$

This prior is very specific. It extremely suppresses strong pixel values and thereby always favors to smooth out emission over all pixels in the image, while sharp peaks are heavily down-weighted. In the case of the model $m(x)$ being a close approximation to the real signal, the prior becomes effectively flat and MEM turns basically into a maximum likelihood fit.

In comparison to [Enßlin & Weig \(2010\)](#), it implicitly assumes no correlation between pixels, and a generally more than exponentially falling brightness distribution. Both features are significant deviations from the assumptions behind [Enßlin & Weig \(2010\)](#). We argue that the presented MEM prior is effectively not well suited for extended emission, but rather attempts a model-dependent, “blind” image smoothing operation as opposed to the data-driven usage of reconstructed spatial correlations in [Enßlin & Weig \(2010\)](#). In this, [Enßlin & Weig \(2010\)](#) is much better suited for structured and not only maximally smooth emission. In addition, tests indicate that [Enßlin & Weig \(2010\)](#) can also handle more compact emission to some degree, something that MEM usually has significant problem with.

Appendix C: Implementation of RESOLVE

C.1. General implementation

We have implemented [Enßlin & Weig \(2010\)](#) in P [Enßlin & Weig \(2010\)](#), where crucial parts have been translated into more efficient [Enßlin & Weig \(2010\)](#) code using C [Enßlin & Weig \(2010\)](#)⁸. The actual implementation of the algorithm makes heavy use of the versatile inference library NIFT [Enßlin & Weig \(2010\)](#) (Selig et al. 2013).

To perform the gridding and degriding operations needed in radio astronomical applications, we use the generalized fast Fourier transformations package [Enßlin & Weig \(2010\)](#)⁹. The grid convolution is performed using a Kaiser-Bessel kernel following Beatty et al. (2005).

For numerical optimization, we use a self-written steepest descent solver and in some cases the conjugate gradient routine provided by the S P [Enßlin & Weig \(2010\)](#) package.

The algorithm is controlled by a number of numerical procedures and parameters, governing the grade of convergence and the degree of accuracy. Apart from standard parameters, such as the maximum number of iterations or the accuracy of the steepest descent. The most important are:

- Different starting guesses for s and p might have a strong impact on the performance or the solution of [Enßlin & Weig \(2010\)](#). In nonlinear optimization, there is, for instance, always the danger of only converging to a local minimum. Experience showed that in most cases, it is optimal to use constant fields and simple generic power spectra as starting guesses to prevent any biases. However, other options are available, e.g., a CLEAN or a dirty map, and/or their respective empirical power spectra, in some cases allowing for an improvement in computation time.
- To calculate D for (A.22), we have to numerically invert D^{-1} and statistically probe the needed matrix entries (Selig et al. 2012) using an implicit representation of the operator as a coded function. For this, we employ a conjugate gradient routine whose convergence and accuracy parameters must be set. This numerical inversion is usually the most serious bottleneck in computation time (see Sect. C.2). Especially calculating D for an estimate of the signal uncertainty can be a time consuming task, depending on the accuracy needed.
- For observations with rather poor uv -coverage, problems might occur with the inversion of the operator D , which sometimes tends to be numerically nonpositive definite during early iterations. In that case, we have implemented a solution where a diagonal matrix with a user-defined positive constant M_0 gets added to D^{-1} to ensure positive-definiteness. While the solution is slowly converging over the global iterations, M_0 is constantly decreased. This is a standard approach in numerical optimization, see, for instance Transtrum & Sethna (2012).
- For large data sets, it is sometimes of high advantage to bin the power spectrum instead of mapping it over all possibly allowed modes set by the user defined image size. Otherwise, the calculation might take prohibitively long.

C.2. Analysis of algorithmic efficiency

As visualized in Fig. 1, [Enßlin & Weig \(2010\)](#) mainly consists of two parts, a signal estimator, and a power spectrum estimator. They are iterated N_{global} times, until convergence is achieved, while both the

⁸ See <http://docs.cython.org/>

⁹ See <https://github.com/mrbell/gfft>

maximum number of iterations and the exact convergence criteria can be set by the user. The signal estimator utilizes a steepest descent algorithm to solve Eq. (23), which needs N_{sd} internal iterations. The power spectrum is estimated with Eq. (25), where the trace of the inverse operator given by Eq. (24) needs to be calculated. Since the operator is only given implicitly, its diagonal entries need to be probed N_{pr} – times using random vectors (Selig et al. 2012), where, for each probe, the operator Eq. (24) has to be inverted using a conjugate gradient algorithm.

The steepest descent iterations are dominated by the operations needed to calculate M (see Eq. (23)), which involves the response operator R with a FFT and a subsequent Gridding operation. Therefore, its computational cost goes roughly with $N_{sd}(O(n_d) + O(n_s \log(n_s)))$, where n_d is the total number of visibilities, and n_s the number of pixels in image space.

The conjugate gradient is dominated by the need to compute the same operation, only, at least some fraction of n_s times, and for each probe individually. Usually a maximum of $\sqrt{n_s}$ iterations of the conjugate gradient are performed. This leads to a total computational cost of roughly $N_{pr}(O(\sqrt{n_s}n_d) + O(\sqrt{n_s}n_s \log(n_s)))$.

A realistic assessment of the asymptotic overall algorithmic efficiency is complicated because all of the iteration numbers, N_{global} , N_{sd} , and N_{pr} , can in principle vary strongly from case to case. Although N_{sd} usually will be larger than N_{pr} ¹⁰, the conjugate gradient term will likely dominate the algorithmic costs. In realistic applications, n_d will usually be larger than n_s , because, for modern instrument data sets, the number of visibilities can reach the millions. In that case, the algorithmic efficiency probably tends to $N_{global}N_{pr}O(\sqrt{n_s}n_d)$.

In addition, this analysis shows that calculating an estimate for the uncertainty of the signal reconstruction is very costly. To accurately compute the diagonal of D , a large number of probes is needed so that N_{pr} can easily exceed the thousands.

On our development machine, with up to eight used CPUs and a maximum of 64 GB working memory, the nonoptimized code produced the results presented in Sect. 3.1 in roughly a couple of hours for the low-noise case, and a couple of days for the high-noise case. For the relatively small size of the simulated VLA snapshot data sets, we never used more than a few percent of the memory but this would most likely change for larger data sets.

¹⁰ At least empirically taken from the simulations, the number of probes can be kept well below a couple of hundreds.