

Robust sparse image reconstruction of radio interferometric observations with PURIFY

Luke Pratley,¹★ Jason D. McEwen,¹ Mayeul d’Avezac,² Rafael E. Carrillo,³
Alexandru Onose⁴ and Yves Wiaux⁴

¹Mullard Space Science Laboratory (MSSL), University College London (UCL), Holmbury St Mary, Surrey RH5 6NT, UK

²Research Software Development Group, Research IT Services, University College London (UCL), London WC1E 6BT, UK

³Signal Processing Laboratory (LTS5), Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne CH-1015, Switzerland

⁴Institute of Sensors, Signals, and Systems, Heriot-Watt University, Edinburgh EH14 4AS, UK

Accepted 2017 August 29. Received 2017 August 29; in original form 2016 November 18

ABSTRACT

Next-generation radio interferometers, such as the Square Kilometre Array, will revolutionize our understanding of the Universe through their unprecedented sensitivity and resolution. However, to realize these goals significant challenges in image and data processing need to be overcome. The standard methods in radio interferometry for reconstructing images, such as CLEAN, have served the community well over the last few decades and have survived largely because they are pragmatic. However, they produce reconstructed interferometric images that are limited in quality and scalability for big data. In this work, we apply and evaluate alternative interferometric reconstruction methods that make use of state-of-the-art sparse image reconstruction algorithms motivated by compressive sensing, which have been implemented in the PURIFY software package. In particular, we implement and apply the proximal alternating direction method of multipliers algorithm presented in a recent article. First, we assess the impact of the interpolation kernel used to perform gridding and degriding on sparse image reconstruction. We find that the Kaiser-Bessel interpolation kernel performs as well as prolate spheroidal wave functions while providing a computational saving and an analytic form. Secondly, we apply PURIFY to real interferometric observations from the Very Large Array and the Australia Telescope Compact Array and find that images recovered by PURIFY are of higher quality than those recovered by CLEAN. Thirdly, we discuss how PURIFY reconstructions exhibit additional advantages over those recovered by CLEAN. The latest version of PURIFY, with developments presented in this work, is made publicly available.

Key words: techniques: image processing – techniques: interferometric.

1 INTRODUCTION

Radio interferometry has been critical for imaging the radio Universe at higher resolution and sensitivity than possible with a single radio telescope. However, radio interferometers are limited by the number of possible pairs of antennas in an array, which limits the number of possible measurements made during an observation. Consequently, image reconstruction methods are needed to reconstruct the true sky brightness distribution from the raw data acquired by the telescope, which amounts to solving an ill-posed inverse problem. Traditional methods, which are mostly variations of the Högbom CLEAN algorithm (Högbom 1974), do not exploit modern state-of-the-art image reconstruction techniques.

Next-generation radio interferometers, such as the LOw Frequency ARray (LOFAR; van Haarlem et al. 2013), the Murchison Widefield Array (MWA; Tingay et al. 2013), the Australian Square Kilometre Array Pathfinder (ASKAP; Hotan et al. 2014) and the Square Kilometre Array (SKA; Dewdney et al. 2013), must meet the challenge of processing and imaging extremely large volumes of data. These experiments have ambitious, high-profile science goals, including detecting the Epoch of Re-ionization (EoR) (Koopmans et al. 2015), mapping large-scale structure (Maartens et al. 2015) and investigating cosmic magnetism (Johnston-Hollitt et al. 2015). If these science goals are to be realized, state-of-the-art methods in image reconstruction are needed to process big data and to reconstruct images with high fidelity.

Compressive sensing is a robust framework for signal reconstruction. The theoretical framework of compressive sensing motivates sparse regularization approaches for solving inverse problems, like

* E-mail: Luke.Pratley@gmail.com

those encountered in radio interferometry. The framework of compressive sensing was first applied to radio interferometry in the study of Wiaux et al. (2009a), in the synthesis framework, where it was shown that compressive sensing approaches can produce higher quality reconstructed images than standard interferometric imaging methods. In Carrillo, McEwen & Wiaux (2012), the analysis framework was considered and the sparsity averaging reweighted analysis (SARA) algorithm was developed and applied to radio interferometric imaging, demonstrating excellent performance (see also Carrillo et al. 2013). It has also been shown that the compressive sensing framework can be applied to wide-field-of-view observations (McEwen & Wiaux 2011) and can correct for directional dependent effects, such as non-coplanar baselines (Wiaux et al. 2009b; Wolz et al. 2013). In Carrillo, McEwen & Wiaux (2014), state-of-the-art convex optimization algorithms that scale to very large data sets were developed to solve sparse regularization problems, such as the SARA problem. These algorithms were implemented in the first release of the PURIFY software package (Carrillo et al. 2014) for solving radio interferometric imaging problems by sparse regularization. Recently, new algorithms for solving these problems were developed by Onose et al. (2016), including proximal alternating direction method of multipliers (P-ADMM) and primal dual algorithms, paving the way to image the large radio interferometric data sets that will characterize the SKA era. Alternative compressive sensing approaches have also been applied to aperture synthesis (Li, Cornwell & de Hoog 2011b; Dabbech et al. 2015; Garsden et al. 2015) and rotation measure synthesis (Li et al. 2011a; Sun et al. 2015).

In this work, we implement the P-ADMM algorithm developed by Onose et al. (2016) in the PURIFY software package, which has been entirely redesigned and re-implemented in C++, and apply it to observational data from the Very Large Array (VLA) and the Australia Telescope Compact Array (ATCA). In addition, we discuss conceptual differences between the restored CLEAN image and the reconstructed PURIFY model. The previous version of PURIFY supported only simple models of the measurement operator modelling the telescope. PURIFY now supports a wider range of more accurate measurement operator models, including a number of different convolutional interpolation kernels (for gridding and de-gridding). Moreover, we study how the choice of kernel can affect the quality of sparse image reconstruction.

The remaining sections of the paper are structured as follows. Section 2 reviews the basics of aperture synthesis and radio interferometry. Section 3 discusses radio interferometric imaging in the context of compressive sensing and sparse image reconstruction. Section 4 discusses convolutional interpolation and the different kernels considered. These interpolation kernels are then tested and compared using simulations in Section 5. Section 6 discusses the similarities and differences between images recovered by CLEAN and PURIFY and also considerations in applying PURIFY to real observational data. The reconstruction of images from observations made by the VLA and ATCA are presented in Section 7. Section 8 states the final conclusions.

2 APERTURE SYNTHESIS AND RADIO INTERFEROMETRY

The principles of aperture synthesis date back as far as the work of McCready, Pawsey & Payne-Scott (1947). However, Ryle & Hewish (1960) first described how aperture synthesis could be used to construct a large-scale radio interferometric telescope. Thus, the limit in resolution of single dish radio telescopes could be overcome

by radio interferometric telescopes, improving our ability to observe and therefore understand the radio sky.

In aperture synthesis, an array of antennas are collectively used to image the sky at higher resolution than possible with a single dish, hence synthesizing a larger aperture. Each pair of antennas measures a phase and amplitude of a Fourier component of the brightness distribution across the sky. It is through the measurement of these Fourier components that the sky is effectively imaged. However, due to a limited number of antennas, not all Fourier components can be measured in an observation. An ill-posed inverse problem must be solved to reconstruct the true sky brightness distribution. How this ill-posed inverse problem is solved has a significant impact on the fidelity of the reconstructed image.

Each antenna in an array measures an incoming electric field across its field of view. The electric fields are then cross-correlated between pairs of antennas, using a correlator, in order to calculate the visibility

$$\mathcal{V}(\mathbf{b} = \mathbf{a}_2 - \mathbf{a}_1) = \langle \mathcal{E}(\mathbf{a}_1, t) \mathcal{E}^*(\mathbf{a}_2, t) \rangle_{\Delta t}, \quad (1)$$

where \mathcal{E} is the electric field, \mathbf{a}_1 and \mathbf{a}_2 are the spatial positions of the two antenna, t is the time, and Δt is the time interval over which the expected value, denoted by $\langle \cdot \rangle$, is taken, which is longer than the time-scale of the radio wave observed (Thompson 1999; Thompson, Moran & Swenson 2008). The difference between the positions of the antennas $\mathbf{b} = \mathbf{a}_2 - \mathbf{a}_1$ is called the baseline.

It is well known that a visibility contains spatial information about the brightness distribution across the sky. While there have been more general measurement equations developed for radio interferometry (McEwen & Scaife 2008; Carozzi & Woan 2009; Smirnov 2011; Price & Smirnov 2015), the van Cittert-Zernike theorem (Zernike 1938) states that the visibility \mathcal{V} is related to the sky brightness distribution \mathcal{I}_λ , at wavelength λ , by

$$\mathcal{V}(\mathbf{b}) = \int_{S^2} \mathcal{A}(\boldsymbol{\sigma}) \mathcal{I}_\lambda(\boldsymbol{\sigma}) e^{-2\pi i \lambda \mathbf{b} \cdot \boldsymbol{\sigma}} d\Omega, \quad (2)$$

where \mathcal{A} is the primary beam of the telescope, \mathbf{b} is the baseline separating the two antennas, and $\boldsymbol{\sigma}$ denotes a location on the celestial sphere S^2 with area element $d\Omega$. When the baselines in an array are co-planar and the field of view is narrow, equation (2) reduces to a Fourier relation:

$$\mathcal{V}(u, v) = \int_{\mathbb{R}^2} \mathcal{A}(l, m) \mathcal{I}_\lambda(l, m) e^{-2\pi i (ul + vm)} dl dm, \quad (3)$$

where (l, m) are the coordinates of the plane of the sky, centred on the pointing direction of the telescope and $\mathbf{u} = (u, v)$ are the corresponding Fourier coordinates defined by the baseline: $\mathbf{u} = \mathbf{b}/\lambda$. In this context, a visibility measures a Fourier component of the sky brightness distribution in the plane of the sky, centred on the pointing direction of the telescope (Thompson 1999; Thompson et al. 2008).

The Fourier transform relation of equation (3) cannot be inverted directly to obtain an accurate estimate of $\mathcal{I}_\lambda(l, m)$, since $\mathcal{V}(u, v)$ cannot be measured for all Fourier coordinates. The missing samples of $\mathcal{V}(u, v)$ leave equation (3) as an ill-posed inverse problem, which has an infinite number of possible solutions. To recover a suitable, unique solution, regularization is used to inject prior information regarding the underlying signal.

The most common techniques used to solve for the true sky brightness distribution are CLEAN (e.g. Högbom 1974) and the maximum entropy method (MEM) (e.g. Cornwell & Evans 1985). The basic CLEAN algorithm was developed in the 1970s (Högbom 1974). CLEAN implicitly imposes a sparse prior in a point source

(Dirac) basis (Marsh & Richardson 1987) and is essentially a matching pursuit algorithm (Mallat & Zhang 1993). Variations of CLEAN have also been developed for resolved and extended structures, multifrequency synthesis and polarized sources (Clark 1980; Schwab 1984b; Steer, Dewdney & Ito 1984; Sault, Staveley-Smith & Brouw 1996; Cornwell 2008; Offringa et al. 2014; Pratley & Johnston-Hollitt 2016). The MEM algorithm regularizes the ill-posed radio interferometric inverse problem through an entropic prior, maximizing an objective function comprised of an entropy term and a data fidelity term (in practice an additional flux constraint is typically imposed in radio interferometric applications of MEM; Cornwell & Evans 1985). In practice, CLEAN often struggles to image diffuse structure, while MEM struggles to resolve point sources. CLEAN, and its variants, are of widespread use in radio interferometric imaging today, while MEM has not experienced such widespread adoption.

3 COMPRESSIVE SENSING FOR RADIO INTERFEROMETRIC IMAGING

In its fundamental form, compressive sensing provides a framework for recovering signals from small numbers of measurements and considers the efficient design of the signal measurement process (Candès, Romberg & Tao 2006a,b; Donoho 2006; Candès & Wakin 2008). In radio interferometry, there is little control over the measurement process since the baseline configurations are typically limited by the interferometer (nevertheless, there may be scope for telescope optimization; Wiaux et al. 2009b; Wolz et al. 2013). The compressive sensing framework, however, motivates a robust method of reconstructing images from the visibilities measured by a telescope through sparse regularization. Sparse regularization exploits the fact that many natural signals – such as astronomical images – are sparse or compressible; that is, for a suitable representation (e.g. wavelet basis) most of the coefficients for the ground truth image are zero or close to zero, respectively. In this section, we review sparse regularization and how it is applied to radio interferometric imaging.

3.1 Sparse regularization

Consider the ill-posed inverse problem of estimating the image $\mathbf{x} \in \mathbb{R}^N$ from measurements $\mathbf{y} \in \mathbb{C}^M$, where the measurements are acquired by the process $\mathbf{y} = \Phi \mathbf{x} + \mathbf{n}$, where the operator $\Phi \in \mathbb{C}^{M \times N}$ models the acquisition system and $\mathbf{n} \in \mathbb{C}^M$ represents noise. This problem accurately models interferometric imaging, as discussed in more detail in the subsequent sections. For now, we consider sparse regularization approaches to solve this general problem.

Sparse regularization techniques promote sparse solutions when solving ill-posed inverse problems. Typically, natural signals are sparse in a suitable basis (e.g. a Dirac, Fourier or wavelet basis) or, more generally, in a sparsifying dictionary. The atoms (cf. basis functions) of the dictionary (Rubinstein, Bruckstein & Elad 2010) can be represented by columns of the operator $\Psi \in \mathbb{C}^{N \times D}$, where N is the number of pixels in the image and D is the number of coefficients of the sparse representation, i.e. $\alpha \in \mathbb{C}^D$. The image can then be decomposed into its sparse representation by $\mathbf{x} = \Psi \alpha$.

A sparse solution to the inverse problem described above can be promoted by imposing a penalty on the number of non-zero coefficients of the sparse representation α through the ℓ_0 -norm, where the ℓ_0 -norm $\|\alpha\|_{\ell_0}$ is defined as the number of non-zero coefficients of α . In principle, the inverse problem can then be

solved by minimizing the ℓ_0 -norm of the sparse coefficients, subject to a data fidelity constraint:

$$\min_{\alpha \in \mathbb{C}^D} \|\alpha\|_{\ell_0} \quad \text{subject to} \quad \|\mathbf{y} - \Phi \Psi \alpha\|_{\ell_2} \leq \epsilon. \quad (4)$$

Given the solution to this problem, denoted α^* , a recovered image can be synthesized by $\mathbf{x}^* = \Psi \alpha^*$. The solution to this minimization problem is given by a model that matches the measurements, within error $\epsilon \in \mathbb{R}^+$, while being constructed from a minimal number of coefficients in the sparse representation. However, this problem cannot be solved in a high dimensional setting because the ℓ_0 -norm is non-differentiable and the minimization problem is non-convex: it is considered an NP hard problem.

The closest convex relaxation of the ℓ_0 problem is the ℓ_1 problem:

$$\min_{\alpha \in \mathbb{C}^D} \|\alpha\|_{\ell_1} \quad \text{subject to} \quad \|\mathbf{y} - \Phi \Psi \alpha\|_{\ell_2} \leq \epsilon, \quad (5)$$

where the ℓ_p -norm is defined by $\|\mathbf{r}\|_{\ell_p} = (\sum_i |r_i|^p)^{1/p}$ (hence the ℓ_1 -norm is the sum of the absolute value of the components of a vector and the ℓ_2 -norm is the usual Euclidean norm). This ℓ_1 minimization problem also promotes sparsity and in some cases exhibits the same solution as the ℓ_0 problem (Candès et al. 2006a; Donoho 2006). Furthermore, since the ℓ_1 minimization problem is a convex problem it can be solved using efficient convex optimization algorithms (e.g. Combettes & Pesquet 2011).

The problem defined by equation (5) is proposed in the standard synthesis setting, where one recovers the coefficients α and synthesizes the recovered image by $\mathbf{x} = \Psi \alpha$. Alternatively, we can propose the problem in the analysis setting using the adjoint wavelet transform Ψ^\dagger :

$$\min_{\mathbf{x} \in \mathbb{R}^N} \|\Psi^\dagger \mathbf{x}\|_{\ell_1} \quad \text{subject to} \quad \|\mathbf{y} - \Phi \mathbf{x}\|_{\ell_2} \leq \epsilon, \quad (6)$$

where one recovers the image \mathbf{x} directly, while still imposing sparsity in some sparse representation. When the sparsifying operator Ψ is an orthogonal basis the solutions of the synthesis and analysis problems are identical. However, for an overcomplete dictionary the solutions are very different and the analysis setting has been shown to perform very well in practice (e.g. Carrillo et al. 2012; Carrillo et al. 2013). Moreover, reweighted schemes to better approximate the solution of the ℓ_0 problem by solving a sequence of ℓ_1 problems can also be considered (Candès, Wakin & Boyd 2008; Carrillo et al. 2012; Carrillo et al. 2013). While these approaches can further improve the quality of the reconstructed image, we do not consider them further here.

Additionally, sparse regularization problems allow extra constraints to be imposed, such as a real and positive valued image, which is the case for total intensity (Stokes I) radio interferometric observations. However, the positivity and real-valued image constraints may be removed for polarimetric imaging, such as linear polarization or the Stokes parameters. Complex valued linear polarization reconstructions of $P = Q + iU$ can also be performed in principle and will be rotationally invariant for rotations in P (Pratley & Johnston-Hollitt 2016).

3.2 Radio interferometric measurement operator

In solving sparse regularization problems, the measurement operator is required to compare how close the reconstructed model matches the measured data. How close the measurement operator matches the true measurement process will have an impact on reconstruction quality.

In the context of radio astronomy, the measurement process is given by equation (3). We assume co-planar baselines and a small wide-field of view here; we do not consider direction-dependent effects in the measurement operator, although they can nevertheless be modelled in the framework presented (Wiaux et al. 2009b; Wolz et al. 2013). In the compressive sensing setting, the measurements $\mathbf{y} \in \mathbb{C}^M$ denote the visibilities $y_i = \mathcal{V}(u_i, v_i)$ and the image $\mathbf{x} \in \mathbb{R}^N$ denotes the sky brightness distribution $x_p = \mathcal{I}_\lambda(l_p, m_p)$ (for $i = 1, \dots, M$ and $p = 1, \dots, N$). The measurement operator $\Phi \in \mathbb{C}^{M \times N}$ specifies a discrete representation of equation (3). Ideally, Φ would represent a direct Fourier transform from the N pixels of the image to the M non-uniformly spaced visibilities. However, this would require $\mathcal{O}(MN)$ computations. Consequently, a direct Fourier transform of the visibilities is not possible for the settings experienced in practice, where a single observation may be comprised of very large numbers of visibilities and high-resolution reconstructed images are required.

Alternatively, it is possible to approximate a direct Fourier transform. One can first interpolate the visibilities on to a regularly spaced grid, which requires order $\mathcal{O}(M)$ operations. Then, it is possible to take advantage of the fast Fourier transform (FFT), which requires order $\mathcal{O}(N \log N)$ operations. This approach requires considerably fewer computations than the direct Fourier transform (Briggs, Schwab & Sramek 1999), rendering a non-uniform Fourier transform computationally feasible for very large observational data sets, but it is an approximation. This approximation is the standard approach considered in radio astronomy.

The standard radio interferometric measurement operator Φ can be written as a series of linear operators:

$$\Phi = \mathbf{WGFZDB}, \quad (7)$$

where $\mathbf{B} \in \mathbb{C}^{N \times N}$ is the primary beam of telescope, $\mathbf{D} \in \mathbb{C}^{N \times N}$ is a gridding correction operator that scales the image to correct for the interpolation convolution kernel, $\mathbf{Z} \in \mathbb{C}^{\alpha^2 N \times N}$ is a zero-padding operator that provides oversampling by factor α in each dimension of the Fourier domain, $\mathbf{F} \in \mathbb{C}^{\alpha^2 N \times \alpha^2 N}$ is a FFT operator, $\mathbf{G} \in \mathbb{C}^{M \times \alpha^2 N}$ is a convolutional interpolation operator that uses a convolution kernel to interpolate visibilities from Fourier coefficients on a regular grid to Fourier components in the continuous Fourier plane, and $\mathbf{W} \in \mathbb{C}^{M \times M}$ weights the measurements according to their error. Alternatively, it is possible to include the weighting \mathbf{W} by weighting the ℓ_2 -norm directly. A diagram of the process of applying the measurement operator Φ and its adjoint Φ^\dagger is shown in Fig. 1. Since the weights are applied in the measurement operator, it is necessary to also weight the measurements, i.e. $\mathbf{y} \rightarrow \mathbf{W}\mathbf{y}$.

3.3 Radio interferometric imaging with PURIFY

To apply compressive sensing techniques to radio interferometry, one needs to pose the sparse regularization problems in Section 3.1 and then solve them using the measurement operator of Section 3.2. The software package PURIFY has been designed and written for this purpose.

The first public version of PURIFY was written in C and solved the problems described in Carrillo et al. (2014), where it was shown on simulations to produce higher fidelity reconstructed images than standard radio interferometric imaging methods. To solve ℓ_1 minimization problems, PURIFY calls the Sparse OPTimisation (SOPT) software package (Carrillo et al. 2012; Carrillo et al. 2013). This first version of PURIFY used the simultaneous-direction method of multipliers (SDMM) algorithm (Carrillo et al. 2014). Recently, new algorithms have been developed for radio interferometry

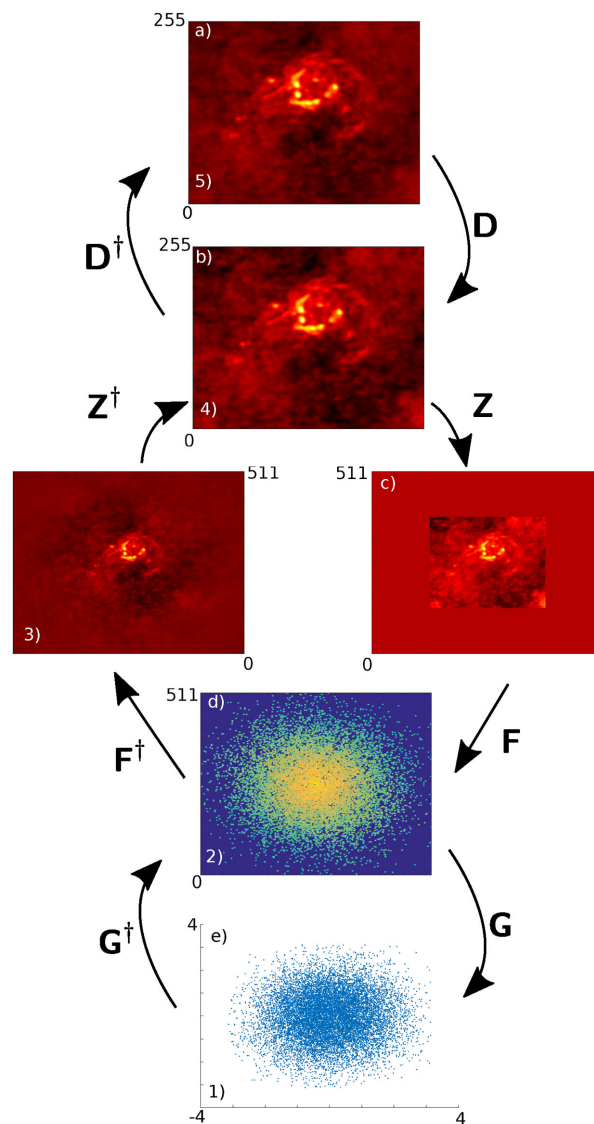


Figure 1. Representation of the application of the forward and adjoint measurement operator. The labels (a)–(e) represent the process of the forward measurement operator, while numbers (1)–(5) represent the process of the adjoint operator. The measurement operator consists of the following steps: (a) observed image; (b) image is corrected for degridding; (c) image is zero-padded to twice the field of view; (d) image is Fourier transformed; (e) Fourier coefficients are convolved to continuous points off of the grid. The adjoint measurement operator consists of the following steps: (1) Fourier coefficients in a continuous plane; (2) Fourier coefficients are gridded on to an oversampled grid; (3) image from the transformed Fourier coefficients; (4) image cut-out; (5) image corrected for the gridding.

imaging by Onose et al. (2016), including the proximal alternating direction method of multipliers (P-ADMM) and primal dual algorithms, which have numerous advantages for the analysis of very large data sets (see Onose et al. 2016 for further discussion).

New versions of PURIFY and SOPT have been released to coincide with the current article. Both PURIFY and SOPT have been completely redesigned and rewritten in C++11 to work on Linux and Mac operating systems. The EIGEN¹ library is used for matrix and array

¹ <http://eigen.tuxfamily.org>

manipulation (Guennebaud et al. 2010), and CASACORE² is used to read observational data in the form of measurement sets (McMullin et al. 2007). SOPT is not only useful for interferometric imaging: it is a general purpose code for solving sparse regularization problems and can be used to solve a variety of problems. The first version of PURIFY was limited to measurement operators based on Gaussian kernels for convolutional gridding. The new version of PURIFY, however, supports numerous kernels, including the state-of-the-art kernels discussed in the literature (e.g. Fessler & Sutton 2003), as described in Section 4. Additionally, the P-ADMM algorithm of Onose et al. (2016) has been implemented in PURIFY and SOPT. Implementation of the primal dual algorithm of Onose et al. (2016) into PURIFY and SOPT is part of future work. The primal dual algorithm achieves greater flexibility, in terms of memory requirements and computational burden per iteration, by using full splitting and randomized updates. All results presented in this article are obtained with the P-ADMM algorithm, solving the analysis problem of equation (6), with an additional positivity constraint (however, it is possible to remove the positivity or reality constraints). While the development of fully distributed implementations of the algorithms supported by PURIFY and SOPT is ongoing, current versions are parallelized with OPENMP, so that the gridding, degriding and FFT calculations can be performed efficiently. The latest versions of PURIFY³ and SOPT⁴ are now publicly available.

4 CONVOLUTIONAL GRIDDING AND DEGRIDDING

The fidelity of reconstructed radio interferometric images depends not only on the technique used to solve the resulting inverse problem but also on the accuracy with which the measurement operator models the measurement process. Ideally, the measurement operator would match the measurement process exactly. However, this is not possible due to the computational time required for a direct Fourier transform. We are forced to use a measurement operator that interpolates the visibilities on to and off of a regular grid through the operator \mathbf{G} , so that we may apply an FFT \mathbf{F} to regularly spaced data. Interpolation is typically performed by convolution with a suitable kernel, which then determines the convolutional degriding operator \mathbf{G} . Several interpolating convolutional kernels have been suggested in the literature; we introduce a subset of these kernels in this section. The choice of convolution kernel affects the quality of the image, through aliasing error, and total computation time, through the support size of the kernel. Ideally, a convolution kernel will have minimal support while maximally suppressing aliasing error, allowing high-quality images to be reconstructed in minimal computation time.

4.1 Degriding

To replicate the measurement process, Fourier coefficients need to be interpolated off of the FFT grid, i.e. they need to be *degridded*. An ideal interpolation that does not change the content of an image is the well known (Shannon) Sinc interpolation (Whittaker 1915; Shannon 1949), where a continuous band-limited image can be exactly reconstructed from the discrete Nyquist sampled signal. Sinc interpolation can also be considered in the context of interpolating

the Fourier domain, which is exact for a space-limited image. In practice, Sinc interpolation in this context can be performed by zero-padding the image domain, which up-samples the Fourier domain via Sinc interpolation.

In the context of degriding, a Sinc interpolation kernel preserves the image and frequency content of the signal when the image has a limited field of view. However, Sinc interpolation is computationally expensive because the Sinc kernel does not have finite support in harmonic space. A computationally inexpensive method, due to its small support, is to interpolate in the Fourier domain using the nearest neighbour grid point. Nearest neighbour interpolation in the Fourier domain corresponds to convolving with a Box kernel, which corresponds to multiplying with a Sinc function in the image domain. Since the Sinc function has infinite support in the image domain, this introduces artefacts known as aliasing error. The Sinc and nearest-neighbour approaches to interpolating visibilities represent the two extreme cases.

We require kernels with small support in harmonic space (so they are computationally efficient) and small support in image space (to suppress aliasing error). However, the uncertainty principle means there is a fundamental limit on how localized a function can be in both harmonic space and image space. In practice, we seek a trade-off between the two extremes, so that the support of the kernel in harmonic space is not so large as to be computationally expensive, while the support in image space is also well-localized to suppress aliasing error.

Since the interpolation is performed by a convolution, it is necessary to correct for this operation, which can be achieved by multiplication in the image domain with an appropriate window. Furthermore, interpolation accuracy can be increased by zero-padding in the image domain to up-sample the Fourier domain. The process of degriding therefore starts by scaling the image by the diagonal operator \mathbf{D} , which preemptively corrects for the interpolation kernel of \mathbf{G} . This correction is calculated from the reciprocal of the inverse Fourier transform of the interpolation kernel. The image is then zero-padded using the zero-padding operator \mathbf{Z} that up-samples harmonic space. An FFT is applied to obtain an up-sampled Fourier grid using the operator \mathbf{F} . The model measurements are then interpolated off of the grid using the circular convolution operator \mathbf{G} . The explicit construction of \mathbf{G} is discussed in Section 4.4.

4.2 Gridding

Most image reconstruction algorithms in radio astronomy require going both backward and forward between the image and measurement domain. Typically, mapping from the measurement domain to the image domain is performed by the adjoint of the measurement operator, since the measurement operator does not have a defined inverse, given by

$$\Phi^\dagger = \mathbf{B}^\dagger \mathbf{D}^\dagger \mathbf{Z}^\dagger \mathbf{F}^\dagger \mathbf{G}^\dagger \mathbf{W}^\dagger. \quad (8)$$

Gridding can be considered the reverse process of degriding. Mathematically, the gridding operator is the adjoint of the degriding operator and is performed by application of \mathbf{G}^\dagger . The full adjoint measurement operator consists of the following operations. First, the weighting $\mathbf{W}^\dagger = \mathbf{W}$ is applied, before the visibilities are interpolated on to an up-sampled Fourier grid using \mathbf{G}^\dagger . Then an inverse FFT is performed by \mathbf{F}^\dagger to produce an image. The image is cropped to the desired field of view using \mathbf{Z}^\dagger , and the convolution is corrected by \mathbf{D}^\dagger . Lastly, the adjoint of the primary beam \mathbf{B}^\dagger is applied.

A consequence of interpolating the visibilities on to a grid is that the signal is now represented via a Fourier series rather than

² <http://casacore.github.io/casacore>

³ <http://basp-group.github.io/purify>

⁴ <http://basp-group.github.io/sopt>

a Fourier transform. This means the imaged region has periodic boundary conditions. In the case of a radio interferometer, the visibilities can contain information over the entire sky, and the signal may not end at the boundaries of the imaged region. In this case, the interpolation kernel is used to apodize aliasing error, where structure from outside the boundaries of the imaged region is folded back in (Briggs et al. 1999).

4.3 Aliasing error

In the case where the convolution kernel does not sufficiently attenuate the image outside the imaged region, the periodicity of the image will cause features from outside the imaged region to fold into the image. Two ways to minimize aliasing error are to either image a wider field of view, so that the primary beam of the telescope naturally attenuates structures outside the field of view, or to choose a convolution kernel that attenuates the aliasing error sufficiently.

An ideal convolution kernel would set the image to zero outside the imaged field of view, which would eliminate aliasing error. This can be done with a Sinc convolution kernel, which is computationally expensive. An inexpensive kernel, like a Box kernel, is highly delocalized in the image domain, so does not suppress structure outside the imaged field of view from being folded back in.

To increase image quality and computational performance, a convolution kernel needs a minimal support in harmonic space while attenuating the image outside the field of view. Any attenuation within the imaged field of view is corrected for by \mathbf{D} , calculated from the Fourier transform of the gridding kernel.

If the gridding kernel apodizes the image domain strongly within the gridded field of view, correcting by \mathbf{D} will induce numerical errors (Schwab 1980). This means that while the suppression due to the gridding kernel can reduce aliasing error, correcting for it has the potential to cause numerical error.

4.4 Interpolation kernels

Next, we introduce the convolution kernels used in this work. The width (support) of the gridding kernel J is given in units of grid cells. The oversampling ratio in each dimension is denoted by α .

The degrading matrix is a circular convolution matrix that interpolates the measurements off of the discrete Fourier grid on to the continuous Fourier plane. The convolution can be seen as a weighted average of the nearest neighbour grid points. The interpolation kernel determines the weighting of each grid point. Weighting is maximum at the location of the measurement and typically decreases in value when the grid points are further from the measurement location.

In 1D Fourier space, the degrading matrix \mathbf{G} is constructed from a kernel $d(u)$ by (Fessler & Sutton 2003)

$$\mathbf{G}_{i, \{k_i+j\}_K} = d(u_i - (k_i + j)), \quad (9)$$

where i is the index of the measurement y_i , k_i is the closest integer to visibility coordinate $u_i - J/2$ (in units of pixels) and $j = 1 \dots J$ are the possible non-zero entries of the kernel. The modulo- K function is denoted by $\{\cdot\}_K$, where $K = \alpha\sqrt{N}$ is the dimension of the Fourier grid in 1D (for notational sake, the 2D Fourier grid is comprised of $N = \sqrt{N} \times \sqrt{N}$ samples).

The diagonal convolution correction operator \mathbf{D} can be calculated in a similar way:

$$\mathbf{D}_{i,i} = s\left(\frac{i}{K} - \frac{1}{2}\right), \quad (10)$$

where $s(x)$ is the reciprocal of the inverse Fourier transform of $d(u)$. In practice, \mathbf{D} can be computed numerically from \mathbf{G} or analytically if the inverse Fourier transform of the convolution kernel is tractable.

4.4.1 Sinc

The Sinc convolution kernel is ideal when its infinite support is considered. This convolution kernel can be written as (Schwab 1978; Greisen 1979)

$$d(u) = \left(\frac{u\pi}{N}\right)^{-1} \sin\left(\frac{u\pi}{N}\right). \quad (11)$$

The convolution correction is

$$s(x) = \begin{cases} \frac{1}{N}, & \text{if } |x| \leq \frac{N}{2} \\ 0, & \text{otherwise} \end{cases}. \quad (12)$$

The advantage of the Sinc convolution kernel is that it corresponds to multiplication by a Box function in the image domain, which bounds the signal at the edges of the imaged region. Consequently, there is close to no aliasing error.

4.4.2 Box

The Box function is fast to compute since it is localized in harmonic space, but it does not suppress aliasing error effectively. This kernel has the form (Schwab 1978; Greisen 1979):

$$d(u) = \begin{cases} \frac{1}{J}, & \text{if } |u| \leq \frac{J}{2} \\ 0, & \text{otherwise} \end{cases}. \quad (13)$$

The Fourier transform of the Box function is the Sinc function, so the convolution correction reads

$$s(x) = \left[\frac{\sin(xJ\pi)}{xJ\pi}\right]^{-1}. \quad (14)$$

The Sinc function is not bounded by the edges of the image, and the sidelobes of the Sinc function can cause large aliasing error. This is why the Box function is far from ideal, even if it is fast to compute.

4.4.3 Gaussian

The Gaussian kernel is moderately well-localized in both image and Fourier space and takes the form:

$$d(u) = e^{-\frac{u^2}{2\sigma^2}}. \quad (15)$$

The gridding correction is calculated by the Fourier transform and also takes the form of a Gaussian:

$$s(x) = \left[\frac{\pi}{2\sigma^2}\right]^{-1/2} e^{2x^2\pi^2\sigma^2}. \quad (16)$$

An optimal choice for σ as a function of the support size J was found in the work of Fessler & Sutton (2003), where it was shown that $\sigma = 0.31J^{0.52}$ works better than using the typical value $\sigma = 1$. In the early years of radio astronomy, in the 1970s, the Gaussian kernel was used for convolutional gridding (Thompson et al. 2008).

4.4.4 Prolate spheroidal wavefunction

Prolate spheroidal wavefunctions (PSWFs) do not have an explicit analytic form, but there are several ways of characterizing them (Stratton 1935; Slepian & Pollak 1961; Landau & Pollak 1961, 1962). The most useful way to characterize PSWFs

is in terms of energy concentration. PSWFs are bandlimited functions that maximize the energy concentration in a given interval, by finding the function f that maximizes the ratio

$$\frac{\int_{-\tau}^{\tau} |f(t)|^2 dt}{\int_{-\infty}^{\infty} |f(t)|^2 dt}, \quad (17)$$

for an interval $[-\tau, \tau]$. For a convolution kernel, this is an ideal property since we want the convolution kernel to have minimal support in the Fourier domain and to have a maximal amount of energy concentrated over the imaged region in the image domain. This allows one to have minimal support in the Fourier domain while maximally suppressing aliasing error in the image domain.

The standard choice of PSWFs in radio astronomy are a modified version, where more energy is weighted towards the centre of the image, since typically this is the scientific region of interest. The standard choice of weighted PSWFs are described in the work of Schwab (1984a, 1980). The convolution kernel is given by

$$d(u) = |1 - \eta^2(u)|^{\kappa} \psi_{\kappa}(\pi J/2, \eta(u)), \quad (18)$$

where $\eta(u) = 2u/J$, κ is a parameter that varies the weighting, and ψ_{κ} is a zero-order PSWF that can be calculated using a rational approximation:

$$\psi_{\kappa}(\pi J/2, \eta) = \frac{\sum_{k=0}^n p_k (\eta^2 - \eta_2^2)^k}{\sum_{k=0}^d q_k (\eta^2 - \eta_2^2)^k}, \quad (19)$$

where the p_k and q_k polynomial coefficients are specified in Schwab (1980, 1984a). The case of $\kappa = 0$ reduces to an unweighted PSWF. In this work, we use the polynomial coefficients for a support of $J = 6$ and $\kappa = 1$, the standard used in the radio interferometric imaging packages MIRIAD⁵ (Sault, Teuben & Wright 1995) and Astronomical Image Processing System (AIPS; Greisen 1998).⁶ The correction is provided by Schwab (1984a):

$$s(x) \approx \frac{1}{\psi_0(\pi J/2, 2x)}. \quad (20)$$

4.4.5 Kaiser-Bessel

Kaiser-Bessel functions are another useful form of convolution kernel. The zeroth-order Kaiser-Bessel function can be expressed as

$$d(u) = \frac{I_0\left(\beta \sqrt{1 - \left(\frac{2u}{J}\right)^2}\right)}{I_0(\beta)}, \quad (21)$$

where J is the support, I_0 is the zeroth-order modified Bessel function of the first kind, and β determines the spread of the Kaiser-Bessel function (Jackson et al. 1991; Fessler & Sutton 2003). The gridding correction is calculated from the Fourier transform, yielding (Jackson et al. 1991; Fessler & Sutton 2003):

$$s(x) = \left[\frac{\sin\left(\sqrt{\pi^2 x^2 J^2 - \beta^2}\right)}{\sqrt{\pi^2 x^2 J^2 - \beta^2}} \right]^{-1}. \quad (22)$$

An optimal choice for β as a function of the support size J was found in the work of Fessler & Sutton (2003), where it was shown that for $\beta = 2.34J$ the Kaiser-Bessel kernel performs similarly to the optimal min-max kernel considered in Fessler & Sutton (2003). In

Greisen (1979), it is suggested that the zeroth-order Kaiser-Bessel functions perform similarly to the zeroth-order PSWFs, which is consistent with the results of Jackson et al. (1991). Kaiser-Bessel functions, however, have the advantage that they have an analytic expression that can be evaluated easily and accurately. Note that Kaiser-Bessel functions are the standard choice of interpolation kernel in the interferometric imaging package WSCLEAN⁷ (Offringa et al. 2014).

5 SIMULATIONS

In the previous section, we described how the measurement operator Φ approximates a direct Fourier transform. If this approximation is inaccurate, it will introduce error when recovering interferometric images. The choice of the interpolation kernel will therefore have an impact on reconstruction quality. In this section, we perform simulations to assess the performance of different convolution kernels, using the P-ADMM algorithm (Onose et al. 2016) implemented in the latest release of PURIFY to recover images in the analysis framework, with an additional positivity constraint.

5.1 Simulations

To assess the impact that the interpolation kernel has on image reconstruction with PURIFY, we perform quality tests using simulated measurements. We compare the signal-to-noise ratio (SNR) of the reconstructed image with the ground truth image, reconstructing with different uv -coverages and different interpolation kernels. Note that we cannot replicate all of the complexities of the real observational setting with simple simulations. For example, our simulated observations do not include contributions from sources outside the field of view. Nevertheless, simulations where the ground truth image is known are useful for a partial assessment of the performance of different convolution kernels.

To ensure the simulated measurements do not limit the reconstruction quality, a high-quality ‘ground truth’ measurement operator is applied to test images of H II emission of M31 and 30 Doradus (30Dor). The Kaiser-Bessel kernel with a support of 8×8 pixels and an oversampling ratio of $\alpha = 2$ is used for the ground truth measurement operator. The Kaiser-Bessel kernel typically requires only a small support, so choosing a support of 8×8 provides an accurate measurement model (Fessler & Sutton 2003).

We calculate the average SNR for reconstructing M31 and 30Dor from M visibilities, in a way that does not depend on a specific uv -coverage. The uv -coverages are randomly generated to follow a Gaussian variable sampling density with a standard deviation of $\pm\pi/3$ in the uv -plane, where the uv -plane has been normalized to a maximum height and width of $\pm\pi$. Ten sample uv -coverages were generated using M visibilities. The average SNR of a reconstruction from M visibilities was calculated using the ten sample uv -coverages. The standard deviation is used to estimate the spread of the SNRs of the reconstructed images. The test images of M31 and 30Dor and a sample uv -coverage are shown in Fig. 2.

Gaussian noise was added to the simulated visibilities. The input SNR (ISNR) of the measurements was chosen to be 30 dB. The

⁵ <http://www.atnf.csiro.au/computing/software/miriad/>

⁶ <http://www.aips.nrao.edu/index.shtml>

⁷ <https://sourceforge.net/projects/wsclean/>

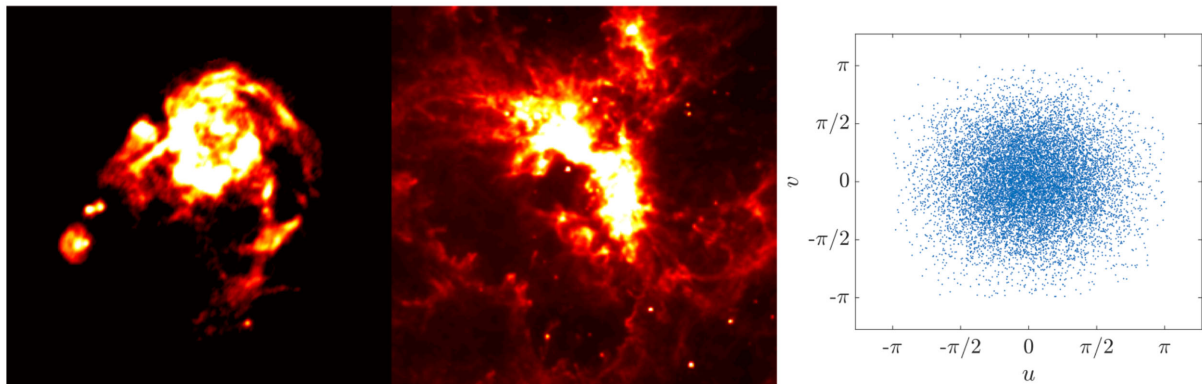


Figure 2. Ground truth images of M31 (left) and 30Dor (middle) used in simulations (of size 256×256). An example of a variable density visibility coverage in the Fourier plane, normalized to a domain of $\pm\pi$ (right). To generate a simulated observation, the measurement operator was applied to a ground truth image. Each simulation has added thermal noise and a random variable density coverage in the Fourier plane. The reconstruction quality was evaluated as a function of the number of Fourier components measured. The SNR was averaged over ten random coverages, with error bar given by the standard deviation (see Fig. 3).

ISNR can be used to calculate the standard deviation of the Gaussian distribution of noise (Carrillo et al. 2014):

$$\sigma_n = \frac{\|\mathbf{y}_0\|_{\ell_2}}{\sqrt{M}} \times 10^{-\frac{\text{ISNR}}{20}}, \quad (23)$$

where \mathbf{y}_0 are the ground truth visibilities, M is the number of visibilities and ISNR is measured in dB.

The noise is assumed to be Gaussian and independently and identically distributed, which allows the use of the χ^2 distribution to estimate the bound ϵ for the ℓ_2 -norm (Carrillo et al. 2014):

$$\epsilon^2 = (2M + 2\sqrt{4M}) \frac{\sigma_n^2}{2}, \quad (24)$$

where for these tests we set ϵ^2 to two standard deviations above the mean of the χ^2 distribution. Following the work of Carrillo et al. (2014), we calculate the SNR from the relation

$$\text{SNR} = 20 \log_{10} \left[\frac{\|\mathbf{x}\|_{\ell_2}}{\|\mathbf{x} - \mathbf{x}^*\|_{\ell_2}} \right], \quad (25)$$

where \mathbf{x} is the ground truth image and \mathbf{x}^* is the reconstructed image.

We solve the ℓ_1 problem in the analysis setting (equation 6), using P-ADMM. For the P-ADMM step size γ , we use the fixed value of

$$\gamma = \beta \|\Psi^\dagger \Phi^\dagger \mathbf{y}_0\|_{\ell_\infty}, \quad (26)$$

with $\beta = 10^{-3}$, as recommended in Carrillo et al. (2014) and Onose et al. (2016), where $\|\Psi^\dagger \Phi^\dagger \mathbf{y}_0\|_{\ell_\infty}$ returns the maximum coefficient of the measurements in the wavelet representation. The reconstructions were solved by assuming sparsity in the SARA wavelet dictionary, which includes a Dirac (i.e. point source) basis and Daubechies wavelets 1–8 (Carrillo et al. 2012; Carrillo et al. 2013). Note that reweighting is not considered. In these simulations, P-ADMM is stopped when the data fidelity constraint is satisfied and the relative difference in the model image between iterations is less than 10^{-3} . Each reconstruction was run for a maximum of 100 iterations.

5.2 Results

The SNR of the reconstructed images as a function of number of visibilities M/N is shown in Fig. 3 for both M31 and 30Dor. Simulations were performed using five of the different interpolation kernels described in Section 4, including Kaiser-Bessel ($J = 4$, $\beta = 2.34J$),

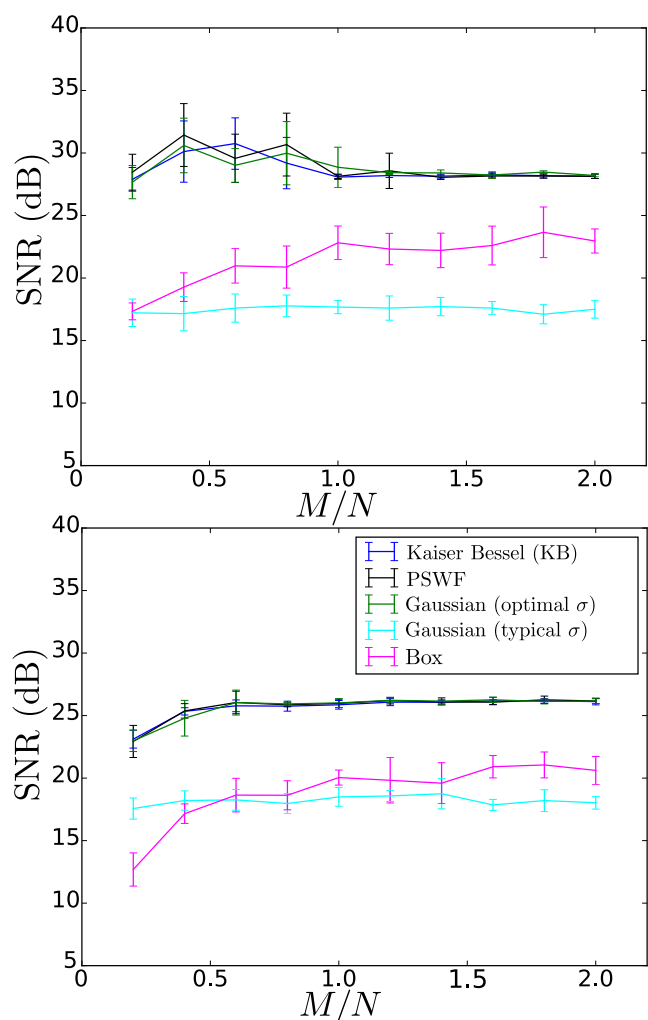


Figure 3. The top and bottom plots of the SNR of the reconstructions of M31 and 30Dor, respectively, with an input SNR of 30 dB. M/N is the ratio of measurements to pixels. Kaiser-Bessel and optimized Gaussian kernels can perform as well as the PSWF. Furthermore, choosing a bad choice of kernel, like a Box function or a Gaussian kernel with a typical σ , limits the possible quality of the reconstruction.

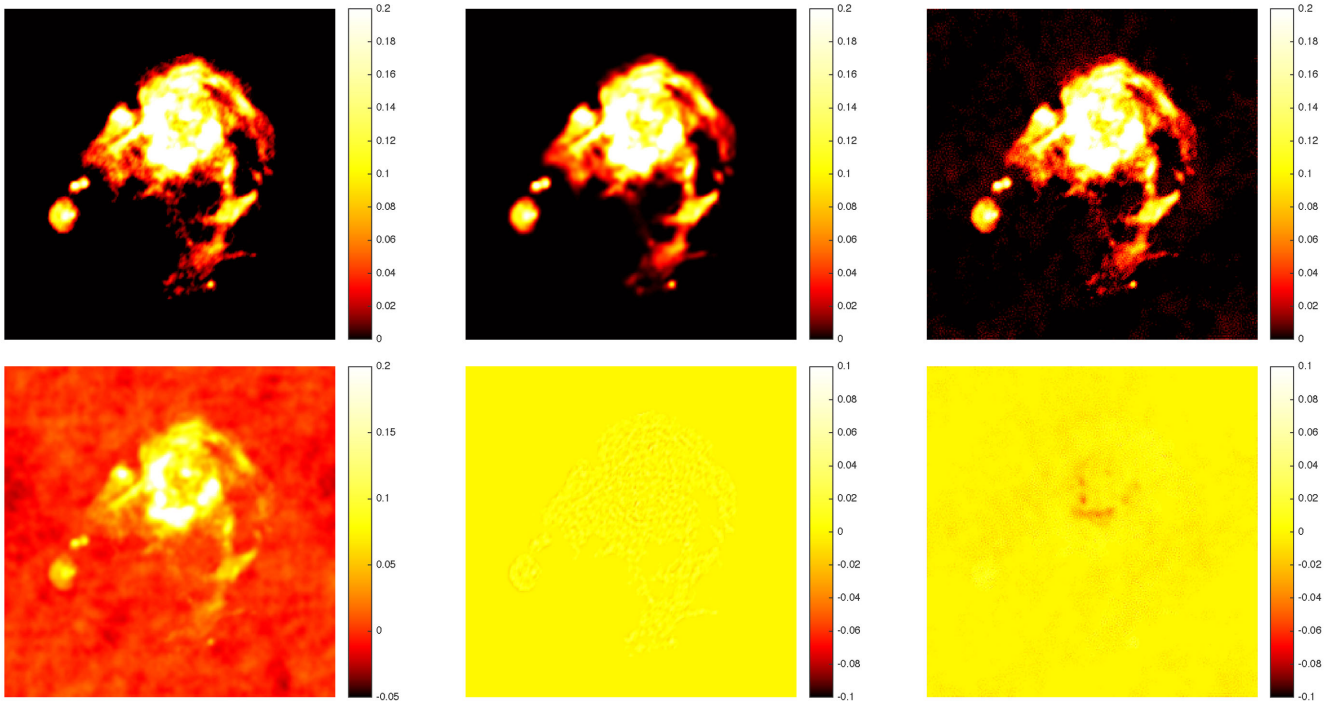


Figure 4. (M31) Left column shows ground truth (top) and dirty image (bottom). Middle column shows reconstructed image (top) and error (bottom) with Kaiser-Bessel kernel. Right column shows reconstructed image (top) and error (bottom) with Box kernel. For these simulations, $M = 2N$ visibilities were used, with an input SNR of 30 dB. The error image shows that the Box kernel reconstruction has artefacts, which explains why the SNR is lower than the Kaiser-Bessel reconstruction. The Box kernel reconstruction did not converge within 100 iterations (based on the convergence criteria described in the text), while the Kaiser-Bessel kernel reconstruction did.

PSWF ($J = 6, \kappa = 1$), Box function ($J = 1$), Gaussian with a typical σ ($J = 4, \sigma = 1$) and optimized σ ($J = 4, \sigma = 0.31J^{0.52}$). An oversampling ratio of $\alpha = 2$ was used for all cases.

Similar SNR results were found for reconstructions using the SARA dictionary for both the M31 and 30Dor images. The Kaiser-Bessel, PSWF and Gaussian kernels with an optimized σ were found to provide reconstructions of the same level of quality. The tests for these kernels converged within 100 iterations.

However, the Gaussian kernel with a typical σ and the Box function provide reconstructions that have an SNR that is 5–10 dB below the other kernels in these tests. Furthermore, for the Box kernel, the reconstructions had often not converged within 100 iterations, while for the Gaussian with a typical σ the majority of tests converged.

To illustrate the difference between reconstructions using the Kaiser-Bessel and Box interpolation kernels, Figs 4 and 5 show example reconstructions for $M = 2N$. Error images are also shown, defined as the difference between the reconstructed and ground truth image. The structure in the Kaiser-Bessel kernel error images looks close to Gaussian error. The structure in the Box kernel error images shows artefacts, which spread throughout the reconstructed image, explaining the lower SNR.

Tests were also performed using only a Dirac basis as the sparsifying dictionary, which provides a proxy for the CLEAN algorithm. The results obtained were consistent with those found with the SARA wavelet dictionary. This suggests that these results found here are likely to apply also to CLEAN and other similar algorithms.

Additional tests were performed at an ISNR of 10 dB, where it was found that there was minimal difference between the reconstructed SNR with different interpolation kernels. This suggests that the choice of interpolation kernel will limit the reconstruction SNR

when the level of artefacts is comparable or greater than the noise level. Consequently, for high dynamic range imaging the choice of kernel is important.

5.3 Discussion

Many calibration and imaging techniques depend on gridding and degridting methods to approximate the Fourier transform. While it has been understood that gridding methods in radio astronomy can impact image quality (Schwab 1978; Greisen 1979; Schwab 1980; Briggs et al. 1999), the current study confirms that gridding with poor kernels reduces the quality of images that can be recovered by sparse regularization approaches, such as those implemented in PURIFY, and also those that can be recovered by CLEAN. The magnitude of the impact depends on the quality of the measurements. For high-quality measurements with high ISNR, the use of poor interpolation kernels will limit the SNR of the reconstruction. At low measurement ISNR, noise dominates the limit imposed by the interpolation kernel.

In particular, we have found that the Gaussian kernel with an optimal σ and the Kaiser-Bessel kernel can both perform as well as the PSWF, while using a smaller support. Moreover, both of the former have analytic forms that can be easily evaluated, which is not the case for the PSWF, where approximations are typically made and look-up-tables used. This suggests that the Kaiser-Bessel kernel is just as good as the PSWF for sparse image reconstruction, and computationally less expensive with a smaller support. These findings are consistent with previous works, suggesting that the Kaiser-Bessel kernel is on par with optimal kernels (Greisen 1979; Jackson et al. 1991; Fessler & Sutton 2003).

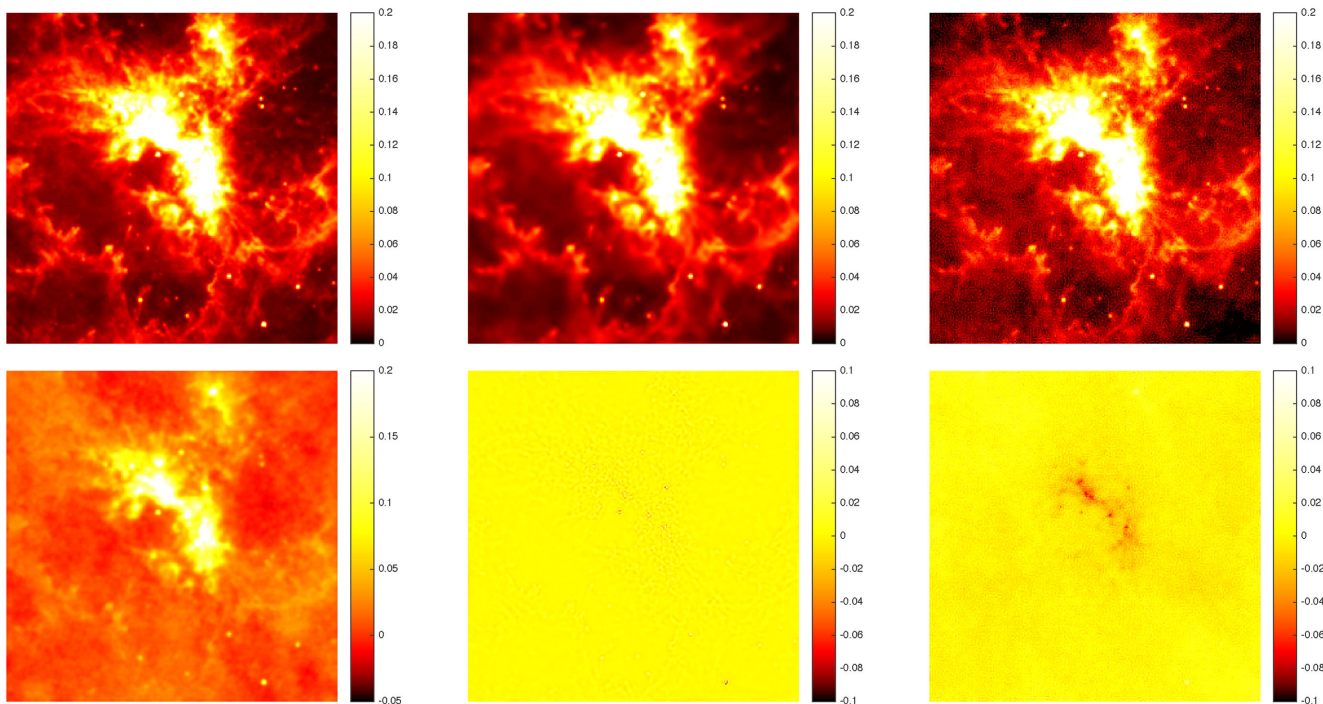


Figure 5. (30Dor) Left column shows ground truth (top) and dirty image (bottom). Middle column shows reconstructed image (top) and error (bottom) with Kaiser-Bessel kernel. Right column shows reconstructed image (top) and error (bottom) with Box kernel. For these simulations, $M = 2N$ visibilities were used, with an input SNR of 30 dB. The error image shows that the Box kernel reconstruction has artefacts, which explains why the SNR is lower than the Kaiser-Bessel reconstruction. The Box kernel reconstruction did not converge within 100 iterations (based on the convergence criteria described in the text), while the Kaiser-Bessel kernel reconstruction did.

6 APPLYING PURIFY TO OBSERVATIONS

The application of compressive sensing to radio interferometry is a relatively new development and to date most of the exploration of compressive sensing has been via simulated observations. Simulations are useful for testing the performance of reconstructions because the ground truth and noise level is known, and appropriate algorithm parameters can be estimated accurately. However, this is not the case when reconstructing images from real observations.

In the next section (Section 7), we demonstrate that PURIFY can perform high-quality image reconstruction on real observations and compare reconstructed images with those recovered by the CLEAN algorithm. However, to compare PURIFY and CLEAN reconstructions, we need to make clear the fundamental differences between the final outputs produced by each approach. In this section, we discuss CLEAN in the context of sparse image reconstruction and clarify where the differences lie. In addition, we describe how to apply PURIFY to real observations, including how to set the pixel size, weighting and other parameters of the algorithm.

6.1 CLEAN comparison

Variations of CLEAN, such as Clark and Cotton-Schwab CLEAN (Clark 1980; Schwab 1984b), work by iteratively building a model of the sky in major and minor cycles. This can be expressed in terms of iterations (Onose et al. 2016)

$$\mathbf{x}^{(t)} = \mathbf{x}^{(t-1)} + \mathcal{T}(\Phi^\dagger(\mathbf{y} - \Phi\mathbf{x}^{(t-1)})), \quad (27)$$

where $\mathbf{x}^{(t)}$ represents the solution after t iterations and \mathcal{T} represents the process of deconvolving the brightest sources in the residuals $\Phi^\dagger(\mathbf{y} - \Phi\mathbf{x}^{(t-1)})$.

CLEAN operates in minor and major cycles, and the minor cycles \mathcal{T} are performed after the calculation of a major cycle $\Phi^\dagger(\mathbf{y} - \Phi\mathbf{x}^{(t-1)})$. The minor cycles iteratively subtract the brightest sources from the image using an approximate PSF, which allows the location of the peaks of multiple sources to be found quickly. The major cycle performs an accurate subtraction of sources located in the minor cycle to generate the residuals for the next round of minor cycles.

CLEAN is essentially a matching pursuit algorithm (Marsh & Richardson 1987), with a threshold constraint as suggested by Högbom (1974), where the algorithm stops when the peak pixel of the residual image is below $\epsilon_{\text{threshold}}$, $\|\Phi^\dagger(\mathbf{y} - \Phi\mathbf{x})\|_{\ell_\infty} \leq \epsilon_{\text{threshold}}$. Most variations of CLEAN impose the prior that the sky is sparse in a Dirac representation (CLEAN components/point sources), while multiscale and adaptive scale pixel decomposition CLEAN consider atoms with wider support to better model a sky containing extended sources (Bhatnagar & Cornwell 2004; Cornwell 2008; Zhang et al. 2016). The solution obtained by the CLEAN algorithm \mathbf{x} is typically called a CLEAN component image.

6.1.1 CLEAN restoration

In the case that the CLEAN components \mathbf{x} could accurately model the entire sky, there would be nothing but noise remaining in the residuals. However, often it is not possible for CLEAN components to model diffuse structures that cannot be represented efficiently by

point sources. For this reason, a final *restored* image is constructed to include structures not deconvolved by CLEAN. The final restored image is found by convolving the CLEAN components with a Gaussian and then adding the residual image:

$$\mathbf{x}^{\text{restored}} = \mathbf{P}\mathbf{x} + \Phi^\dagger(\mathbf{y} - \Phi\mathbf{x}), \quad (28)$$

where \mathbf{P} is a post-processing operator that convolves \mathbf{x} with a Gaussian of the same full width at half-maximum as the dirty beam. The final restored image is expressed in units of Jy Beam⁻¹. These modifications mean the process of constructing a final restored image is not consistent with finding a solution that best fits the data for a given prior, even if the motivations are pragmatic.

The CLEAN residuals are therefore not a true representation of how well the restored image models the true sky. Rather, the residuals $\Phi^\dagger(\mathbf{y} - \Phi\mathbf{x})$ of a reconstructed CLEAN image are due to the CLEAN components \mathbf{x} , not the final restored image $\mathbf{x}^{\text{restored}}$.

An additional systematic that can occur with the CLEAN method is that the dirty beam may not be well approximated by a Gaussian, which is assumed in constructing the restored image (Oberoi, Attridge & Doeleman 2003). This could impact studies that require accurate characterization of point sources, such as weak lensing (Patel et al. 2015). Additionally, in low-frequency imaging the ionospheric distortion on short time-scales can produce a non-Gaussian dirty beam. For low-frequency radio astronomy this is a serious issue, as discussed in Hurley-Walker et al. (2017).

6.2 PURIFY

PURIFY adopts the prior that the sky has a sparse representation. This can include a representation as a collection of point sources and/or single or multiple wavelet dictionaries. This allows more flexibility when modelling both point sources and extended sources simultaneously, providing more accurate deconvolution of complex structure. As a result, diffuse structures are not expected in the residual image; hence, there is no need to combine the model with the residuals as is done with the CLEAN algorithm. PURIFY provides a final image that is completely deconvolved, eliminating the need to convolve the model with a Gaussian beam.

PURIFY therefore provides several advantages over CLEAN (in addition to improved image quality and the ability to scale to big data). First, it means the residuals correspond to the final image used for scientific analysis, such that the final image is the model that minimizes the error (this is not true for the CLEAN restored image). Secondly, the final model image recovered has units of Jy Pixel⁻¹, rather than Jy Beam⁻¹. This provides an advantage when computing statistics on an image and for general scientific interpretation, because there is no need to include Gaussian and dirty beam dependence.

6.3 Choice of pixel size

The final image recovered by PURIFY is sampled at discrete pixel values; hence, there is a choice in the size of a pixel of the discrete image representing the sky brightness. The size and number of pixels can be determined by the resolution and field of view of the telescope. The size of the pixel can be estimated from the resolving power of the longest baseline and number of pixels determined by the field-of-view imaged (by the Nyquist relation).

However, radio astronomy packages such as Common Astronomy Software Applications (CASA) or MIRIAD typically assume between 3 and 5 pixels across the full width at half-maximum (FWHM) of the synthesized beam, found by least-squares fitting a

Gaussian to the main lobe of the synthesized beam (Sault et al. 1995; McMullin et al. 2007; Offringa et al. 2014).

Ideally, the size of the image should include all of the bright sources within the telescope's field of view. When bright sources are outside the imaged field of view they cannot be modelled but may be aliased into the imaged region, which can limit image fidelity.

PURIFY is flexible with regard to the pixel sampling rate and size and these parameters can be set by the user. However, the default approach to setting the pixel size is to adopt Nyquist sampling since the resolution of the model is fundamentally limited by the uv -sampling pattern.

6.4 Weighting

In radio interferometry, it is standard practice to weight the measurements according to natural, uniform or robust weighting schemes, which are described in detail in Briggs (1995). The visibilities are weighted by the natural weighting scheme to optimize the sensitivity of an observation. However, for observations containing extended emission, the sidelobes in the image domain due to natural weighting can dominate the synthesized beam. In this case, CLEAN can perform badly, so the visibilities are uniformly weighted to minimize sidelobes. We concisely review different weighting schemes, including the standard natural, uniform and robust weighting schemes used in radio interferometry. PURIFY supports all of these schemes.

6.4.1 Natural

Natural weighting maximizes the sensitivity of the observation, with weights set to $\mathbf{W}_{i,i}^{\text{natural}} = \sigma_i^{-1}$, where σ_i is the standard deviation of the error for visibility y_i . Note that here we consider the weighting operator as a component of the measurement operator following equation (7); hence, its entries are given by σ_i^{-1} , rather than a scaling of the visibilities only, in which case the weights are given by σ_i^{-2} . Natural weighting is also known as whitening: each measurement has the same (unit) variance after weighting (Carrillo et al. 2014). Whitening is a standard weighting approach in statistical data analysis and image processing. Using natural weighting for interferometric imaging allows one to use a χ^2 distribution when comparing how well the model visibilities fit the data, which can be used for a statistical interpretation of the bound on the ℓ_2 -norm.

6.4.2 Uniform

Uniform weighting minimizes the amplitude of sidelobes over a given field of view, which is achieved by calculating an average weighting from the nearest neighbours of a visibility. Explicitly, an average weight is calculated by

$$\mathbf{W}_{i,i}^{\text{gridded}} = \sqrt{\frac{1}{|\mathcal{S}_i|} \sum_{k \in \mathcal{S}_i} (\mathbf{W}_{k,k}^{\text{natural}})^2}, \quad (29)$$

where \mathcal{S}_i denotes the set of visibility indices that are included in the grid cell corresponding to visibility i , and $|\mathcal{S}_i|$ denotes the number of elements in \mathcal{S}_i . The uniform weights are then calculated by normalizing the natural weights:

$$\mathbf{W}_{i,i}^{\text{uniform}} = \frac{\mathbf{W}_{i,i}^{\text{natural}}}{\mathbf{W}_{i,i}^{\text{gridded}}}. \quad (30)$$

It is possible to control the field of view at which the synthesized beam sidelobe suppression due to weighting occurs by changing the resolution of the grid cells. As the grid resolution increases, the field

of view for dirty beam sidelobe suppression increases, although the suppression level is reduced. As the field of view for suppression increases, the weighting tends to natural weighting.

6.4.3 Robust

Robust weighting allows one to vary a robustness parameter R to continuously move between natural and uniform weighting:

$$\mathbf{W}_{i,i}^{\text{robust}} = \frac{\mathbf{W}_{i,i}^{\text{natural}}}{\sqrt{1 + \rho \left(\mathbf{W}_{i,i}^{\text{gridded}} \right)^2}}, \quad (31)$$

where

$$\rho = \frac{\sum_k \left(\mathbf{W}_{k,k}^{\text{natural}} \right)^2}{\sum_k \left(\mathbf{W}_{k,k}^{\text{gridded}} \right)^4} \times 10^{-2R + \log_{10}(25)}. \quad (32)$$

6.5 Parameter choice

The parameters of PURIFY are set automatically, following the recommendations of Carrillo et al. (2014) and Onose et al. (2016). We also consider some minor modifications of these schemes that can be useful when analysing real observations, where, for example, the errors on the visibilities that are provided (i.e. weights) may not be accurate. Two parameters that need to be set carefully are the bound on the data fidelity error bound ϵ and the step size of the algorithm γ . We suggest a method to estimate ϵ using the Stokes V visibilities and to adaptively estimate the step size γ during the first steps of the algorithm.

6.5.1 Choosing the error bound ϵ

The parameter ϵ determines the error on how closely the model visibilities are required to match the measured visibilities. If ϵ is too small the model will start to fit to noise and if ϵ is too large the model will not model structures accurately.

In the case of natural weighting, ϵ can be estimated by (Carrillo et al. 2014)

$$\epsilon^2 = (2M + q\sqrt{4M}) \frac{\sigma_n^2}{2}, \quad (33)$$

where ϵ^2 is set to q standard deviations above the mean of the χ^2 distribution. However, for typical observations $2M \gg \sqrt{4M}$, so this interpretation is less useful (due to the concentration of measure in high dimensions). For real observations with large M , we simply estimate ϵ from the mean of the χ^2 distribution and allow a scaling:

$$\epsilon_\eta = \eta\sqrt{M}\sigma_n, \quad (34)$$

where η allows one to vary ϵ to include non-thermal noise contributions, such as instrumental errors and radio frequency interference (RFI). When using this latter approach to set ϵ we explicitly denote the η dependence by ϵ_η .

In principle, standard calibration and self-calibration methods can be applied with PURIFY, but to date these have not yet been tested. Such an approach may be considered by choosing a high error bound for ϵ to generate a sky model of the brightest sources, applying a calibration algorithm to recover calibration parameters, before iterating.

In the case that the source of noise in the visibilities is thermal, the weights should be accurate. However, if the weights are not accurate

it is possible to use Stokes V to estimate the noise level and thus ϵ . This is because Stokes V rarely contains astrophysical sources and so is dominated by thermal noise. To estimate the noise on a measurement, we use the median absolute deviation (MAD) method (Rousseeuw & Croux 1993; Hoaglin, Mosteller & Tukey 2000)

$$\sigma_n = \sqrt{\left[\frac{\text{Median}(\text{Real}(\mathbf{W}\mathbf{y}_V))}{0.67449} \right]^2 + \left[\frac{\text{Median}(\text{Imag}(\mathbf{W}\mathbf{y}_V))}{0.67449} \right]^2}, \quad (35)$$

where $\mathbf{W}\mathbf{y}_V$ is the weighted Stokes V visibilities. The MAD method provides a robust way to estimate σ_n given Gaussian noise, and should be reliable when Stokes V is dominated by thermal noise.

Furthermore, if the weights are only proportional to the standard deviation of noise, they will be incorrect by a scaling factor. The MAD method can be used to determine the standard deviation of the noise from a sample distribution. While using the MAD method to estimate σ_n is intended to work for thermal noise contributions, it might not be accurate when there are polarimetric, amplitude and phase calibration errors or RFI.

6.5.2 Adapting the step size γ

In Carrillo et al. (2014), it is suggested that the algorithm step size γ can be set by

$$\gamma = \beta \|\Psi^\dagger \mathbf{x}^{(0)}\|_{\ell_\infty}, \quad (36)$$

where $\mathbf{x}^{(0)}$ is an initial estimate of the image. Typically, the initial estimate is chosen as $\mathbf{x}^{(0)} = \Phi^\dagger \mathbf{y}$ (i.e. the dirty image). While the choice of γ should not affect the final result of the algorithm, it does affect the rate of convergence.

We adapt this approach and allow γ to be re-estimated as the algorithm progresses, before settling on a fixed value of γ to guarantee convergence. In this case, a candidate adaptive step size for the i th iteration can be calculated $\tilde{\gamma}_i = \beta \|\Psi^\dagger \mathbf{x}^{(i)}\|_{\ell_\infty}$. If the current candidate for the step size changes by a small amount only, there is no need to change the step size used. In this case, a general rule for adapting the step size can be set:

$$\gamma_i = \begin{cases} \tilde{\gamma}_i, & \text{if } \frac{\tilde{\gamma}_i - \gamma_{i-1}}{\gamma_{i-1}} > \delta_{\text{adapt}} \\ \gamma_{i-1}, & \text{if } \frac{\tilde{\gamma}_i - \gamma_{i-1}}{\gamma_{i-1}} \leq \delta_{\text{adapt}} \\ \gamma_{i-1}, & \text{if } i \geq i_{\text{adapt}}, \end{cases} \quad (37)$$

where δ_{adapt} is the minimum relative difference needed for adapting the step size and i_{adapt} is the number of iterations after which the step size will not be adapted and will remain fixed.

6.6 Input parameters of PURIFY

As described already, the parameters of PURIFY are set automatically and so PURIFY can be run simply by providing the filename of an input measurement set and the output filename of the image to be recovered. The user does not need to set any parameters. However, the default settings can be overridden.

The main parameters of interest that a user may want to overwrite are specified in Table 1. These include the η value in setting ϵ_η , the β parameter in setting γ , the δ_{adapt} and i_{adapt} parameters that control adapting γ , the relative variation of the solution criteria δ , the residual norm convergence criteria ξ , and the maximum number of iterations i_{max} .

In analysing the observations considered in the next section, the value of η varies from 1.4 to 7 and depends on the quality of the data

Table 1. Description of main user parameters for using PURIFY to reconstruct an observation. All parameters are set automatically but can be overwritten.

Parameter	PURIFY option	Description	Value
η	--l2_bound	Parametrization of the fidelity constraint: $\epsilon_\eta = \eta\sqrt{M}\sigma_n$	$\eta = 1.4$ (default); $\eta \in [1, 10]$ (typical)
β	--beta	Parametrization of the step size of the algorithm: $\tilde{\gamma}_i = \beta\ \Psi^\dagger \mathbf{x}^{(i)}\ _{\ell_\infty}$ (default). One can also fix $\gamma = \beta\ \Psi^\dagger \mathbf{x}^{(0)}\ _{\ell_\infty}$	$\beta = 10^{-3}$ (default)
δ_{adapt}	--relative_gamma_adapt	Relative difference criteria for adapting γ_i	$\delta_{\text{adapt}} = 0.01$ (default)
i_{adapt}	--adapt_iter	Number of iterations to consider adapting the step size γ_i (should be before convergence)	$i_{\text{adapt}} = 100$ (default)
δ	--relative_variation	Relative difference convergence criteria on the ℓ_2 -norm of the solution: $\frac{\ \mathbf{x}^{(i)} - \mathbf{x}^{(i-1)}\ _{\ell_2}}{\ \mathbf{x}^{(i)}\ _{\ell_2}} \leq \delta$	$\delta = 5 \times 10^{-3}$ (default)
ξ	--residual_convergence	Convergence criteria on the ℓ_2 residual norm: $\ \mathbf{y} - \Phi\mathbf{x}\ _{\ell_2} \leq \xi\epsilon_\eta$	$\xi = 1$ (default); require $\xi \geq 1$
i_{max}	--niters	Maximum number of iterations	$i_{\text{max}} = \infty$ (default)

set, such as how free it is from calibration error and RFI. The i_{adapt} parameter is set to a fraction of the maximum number of iterations. It is important to set i_{adapt} such that the step size γ stops adapting before convergence. The relative variation criteria of the objective function was chosen to be $\delta = 5 \times 10^{-3}$. The choice of residual norm convergence criteria ξ also depends on the quality of the data set.

7 PURIFY RECONSTRUCTION OF OBSERVATIONS

In this section, we compare the use of PURIFY and Cotton-Schwab CLEAN for reconstructing total intensity (Stokes I) observations made by the VLA and the ATCA. In particular, we consider observations of the radio galaxies 3C129, Cygnus A, PKS J0334-39 and PKS J0116-473. To perform the Cotton-Schwab CLEAN algorithm, we use WSCLEAN (Offringa et al. 2014). WSCLEAN is a standard choice for imaging in several MWA (Tingay et al. 2013) science pipelines including continuum, transients, EoR and polarization modes (Murphy et al. 2015; Wayth et al. 2015; Jacobs et al. 2016; Lenc et al. 2016; Offringa et al. 2016). For PURIFY, we present results using the P-ADMM algorithm (Onose et al. 2016), in the analysis setting, with a positivity constraint and the SARA wavelet dictionary (Carrillo et al. 2012), without reweighting. Results with alternative algorithms that are being implemented in PURIFY (e.g. the primal dual algorithm; Onose et al. 2016) will be presented in future work.

7.1 Observations

In this section, we discuss the details of the observations considered. The sampling patterns in the uv -plane for each observation are shown in Fig. 6.

7.1.1 3C129

The observation of the bent tailed radio galaxy 3C129 has a phase centre of RA = 04^h45^m31^s.695, Dec. = +44°55'19".95 (J2000), and was obtained from the NRAO archive. It was performed using the VLA with the project code AT0166, with two 50 MHz channels centred at 4.59 and 4.89 GHz. The observations were performed on 1994 July 25 in configuration B and on 1994 November 3 in configuration C, respectively. The total integration time was 79.7 min in configuration B and 15.8 min in configuration C. The calibration and flagging of RFI was performed using CASA, following the

standard procedure found in the CASA manual. The gains were calibrated using sources 0420+417, 0518+165 and 0134+329 to solve for the instrumental and source polarization. Source 0420+417 was observed alternately to solve the polarimetric calibration solutions with parallactic angle coverage.

7.1.2 Cygnus A

The VLA observation and reduction of Cygnus A in the X band (central frequency of 8.953 GHz, and 92 MHz bandwidth) was performed by Rick Perley⁸ (PI: Perley, project code 14B-336 (legacy: AP658)). Cygnus A was observed in 2014 between November 3 (18:39:44.0 UTC) to November 4 (04:28:12.0 UTC), using configuration C. The pointing centre was located at RA = 19^h59^m28^s.356, Dec. = +40°44'02".075 (J2000). The data were reduced and calibrated using AIPS, following standard procedure that can be found in the AIPS Cookbook.⁹

7.1.3 PKS J0334-39

The observation of PKS J0334-39 was first presented in the work of Pratley et al. (2013), where the tailed radio galaxy's polarimetric structure was used to probe the environment of the galaxy cluster Abell 3135. The observation was also reprocessed using self-calibration in Pratley & Johnston-Hollitt (2016), where it was used as an example of applying Generalized Complex CLEAN (Pratley & Johnston-Hollitt 2016) to an observation. The observation was performed using the ATCA (with the pre-CABB correlator) in 2001 is centred on RA = 03^h34^m07".18, Dec. = -39°00'03".19 (J2000), at a central frequency of 1.384 GHz. There are 12 channels, each with a width of 8 MHz. The observation was performed in configuration 6A for 59 min, 1.5A for 76 min, 750A for 79.7 min, 375 for 75.4 min. A detailed description of the calibration procedure, performed using MIRIAD, can be found in Pratley et al. (2013).

7.1.4 PKS J0116-473

The observation of PKS J0116-473 used in this work was first presented in Saripalli, Subrahmanyam & Shankar (2002). The total intensity, polarimetric structure and morphology of PKS J0116-473

⁸ Private communication.

⁹ <http://www.aips.nrao.edu/cook.html>

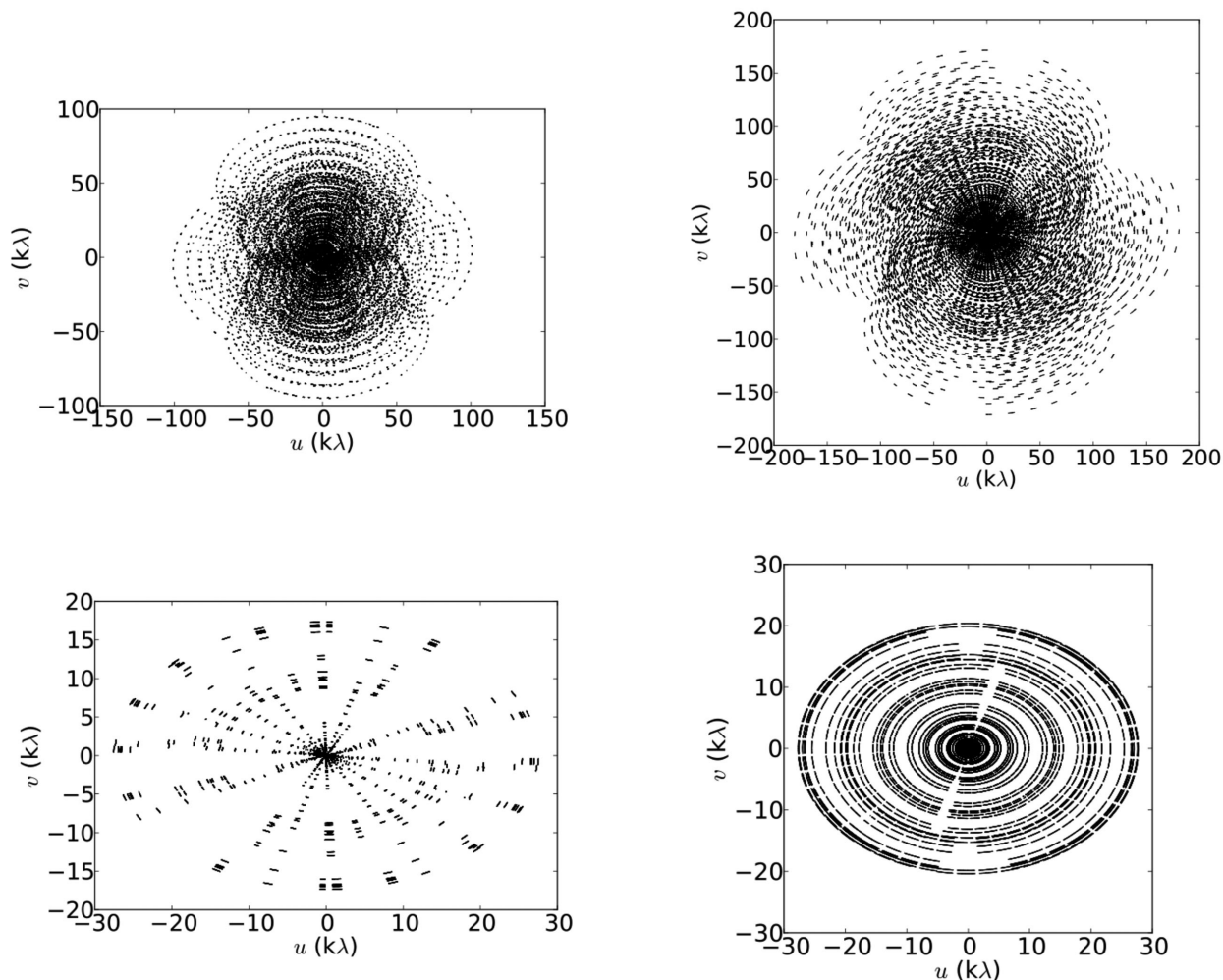


Figure 6. Plots showing the uv -coverage of the observations of 3C129 (top left), Cygnus A (top right), PKS J0334-39 (bottom left) and PKS J0116-473 (bottom right). Units of u and v are kilo-wavelengths (kilo- λ).

have been studied in detail at 12 and 22 cm emission. The ATCA observations of PKS J0116-473 used in this work were extracted from the archive (PI: Shankar, project code C770), then calibrated and flagged following a standard ATCA data reduction procedure found in the *MIRIAD* manual.¹⁰ The phase centre is located at RA = $14^{\text{h}}59^{\text{m}}15^{\text{s}}.75$, Dec. = $-36^{\circ}55'47''.87$ (J2000) and the central observation frequency is 1.384 GHz. After flagging and removing channels due to cross-channel interference, there are 12 channels each with 8 MHz channel width. The observations were performed in 1999 on January 10 and 12 (configuration 375, 1115 min integration), on February 7 (750C, 1088.3 min), on February 20 (6C, 1109.3 min), and on April 24 and 25 (1.5C, 1112 min). Sources PKS B1934-638 and PKS B0823-500 were used to set the flux density scale at 1.384 GHz. The time variations in complex antenna gains and bandpass were calibrated using alternating observations of the unresolved source PKS B0153-410.

¹⁰ <http://www.atnf.csiro.au/computing/software/miriad/userguide/userhtml.html>

7.2 Reconstructions

In this section, we present the reconstructions from real observations. We show the reconstructed model image, alongside the residuals. For the *CLEAN* reconstructions, we show the post-processed restored image (see Section 6.1.1), while for *PURIFY*, there is no need for post-processing so there is no restored image but only a reconstructed model image (see Section 6.1.1). For *PURIFY* reconstructions, we use natural weighting, and for *CLEAN*, we use both natural and uniform weightings.¹¹

The *CLEAN* thresholds and FWHM of the restoring beams can be found in Table 2. The *CLEAN* components are restricted to be positive valued. *CLEAN* has not been restricted to regions around the source. *CLEAN* was run until the residual peak reached the cut-off flux value. We are careful to make the distinction between the *restored* image and the *reconstructed* image for *CLEAN* (see Section 6.1.1), since the restored image is not used to generate the residuals. When we refer

¹¹ Rather than using measurement sets for the ATCA data sets, the tables were read with *PURIFY* from uvfits files. In all other cases, the observations were read from measurement sets.

Table 2. Table listing details of settings used to recover CLEAN images.

Observation	Weighting	Beam size	Cut-off
3C129	Natural	$2.07'' \times 1.88''$, 158°	0.0025 Jy
	Uniform	$1.30'' \times 1.06''$, 33°	
Cygnus A	Natural	$3.48'' \times 2.81''$, 105°	0.1 Jy
	Uniform	$2.25'' \times 1.99''$, 97.4°	
PKS J0334-39	Natural	$45.6'' \times 36.8''$, 171°	0.001 Jy
	Uniform	$8.6'' \times 4.3''$, 17°	
PKS J0116-473	Natural	$40.0'' \times 24.6''$, 158°	0.001 Jy
	Uniform	$6.33'' \times 4.72''$, 3°	

to the reconstructed image, we are referring to the CLEAN component image.

For PURIFY, the error constraint in the model is set using ϵ_η . The P-ADMM step size was set adaptively as described in Section 6.5.2. PURIFY images have a resolution set by the longest baseline in the observation.

Images recovered by CLEAN and PURIFY, and auxiliary plots, are shown in Figs 7–10. Reconstructions of the source 3C129 are shown in Fig. 7 for a pixel width and height of 0.4 arcsec. The PURIFY reconstruction was performed using a value of $\eta = 0.9$ and $\xi = 1$, and ran for 75 iterations. The step size was adapted for the first $i_{\text{adapt}} = 20$ iterations. Fig. 8 contains the reconstructions of Cygnus A for a pixel width and height of 0.5 arcsec. The PURIFY reconstruction was performed using $\eta = 2.14$ and $\xi = 7.07$, and ran for 183 iterations. The step size was adapted for the first $i_{\text{adapt}} = 100$ iterations. Reconstructions of the source PKS J0334-39 are shown in Fig. 9 for a pixel width and height of 2 arcsec. The PURIFY reconstruction was performed using $\eta = 1$ and $\xi = 2$, and ran for 372 iterations. The step size was adapted for the first $i_{\text{adapt}} = 200$ iterations. Reconstructions of the source PKS J0116-473 are shown in Fig. 10 for a pixel width and height of 2.4 arcsec. The PURIFY reconstruction was performed using $\eta = 1$ and $\xi = 2.3$, and ran for 707 iterations. The step size was adapted for the first $i_{\text{adapt}} = 500$ iterations.

The run times for these reconstructions range from an hour to several hours using a high-performance desktop computer to produce images of sizes 1024×1024 and 2048×2048 pixels. Currently, a large factor in the computational cost and run time for PURIFY is computing wavelet transforms for a number of dictionaries. In the case that only a Dirac basis is used and no wavelet transforms are performed, the run time is reduced considerably for large image sizes. However, this greatly reduces the quality of the reconstructed image, because a Dirac basis is not an efficient representation of extended structures. As discussed, highly distributed and parallelized algorithms will be implemented in future work to reduce the runtime significantly (Onose et al. 2016). While CLEAN methods appear computationally efficient, this comes at a significant cost to reconstruction quality and with additional restrictions on the ability for distribution.

In all cases, PURIFY provides more complete reconstructions than CLEAN. When comparing with the CLEAN component images, the CLEAN component images are not smooth and do not reconstruct the diffuse emission well (due to the point source model of CLEAN), while the PURIFY recovered images model diffuse emission. After post-processing the CLEAN component image to yield the CLEAN restored image and comparing with PURIFY, it is also clear that PURIFY provides higher quality reconstructions.

The dirty and residual images of PURIFY are shown in Jy Beam⁻¹ for comparison. To convert from units of Jy Pixel⁻¹ to Jy Beam⁻¹, the image is divided by the peak of the point spread function (PSF) $\Phi^\dagger \mathbf{W} \mathbf{1}$, where $\mathbf{1}$ denotes a vector of ones. This allows direct com-

parisons of the residual images between CLEAN and PURIFY, since they will have the same units without arbitrary scaling. To compare the residuals the scale of the colour axis has been set to a common scale, using three times the median root-mean-square (rms) values between the residual images in Table 4. The histograms show the full range of pixel values in the residuals, determined by the peak of the absolute residuals, to allow one to inspect outliers.

For all observations, PURIFY can model faint extended structure while also modelling the bright compact sources. Additionally, the PURIFY model has left little structure in the residuals. This is also clear from the histogram of the residual pixel brightness, which shows the residuals are dominated by Gaussian noise. The CLEAN reconstruction leaves visible diffuse structure in the residuals. The histogram of the residual images show large peaks below the CLEAN cut-off.

The primary difference that natural and uniform weightings have on CLEAN is that uniform weighting suppresses the synthesized beam sidelobes. While this lowers the sensitivity of the observation, CLEAN then performs better at modelling fine structure with CLEAN components, with diffuse structure left in the residuals, which are then added back in the CLEAN restored image.

The dynamic range is used to assess the quality of reconstructions quantitatively and is calculated by

$$\text{DR} = \frac{\sqrt{N} \|\Phi\|^2}{\|\Phi^\dagger (\mathbf{y} - \Phi \mathbf{x})\|_{\ell_2}} \max\{\mathbf{x}_k\}, \quad (38)$$

i.e. the ratio of the peak of the recovered image to the rms of the residuals (for a normalized measurement operator). The weights are assumed to be in the measurement operator. The norm of the measurement operator is included so that the dynamic range does not scale arbitrarily under the choice of the normalization of the measurement operator. For CLEAN, we follow the standard approach and use the peak of the *restored* image, divided by the rms of the residual image. The dynamic ranges of the images recovered by CLEAN and PURIFY can be found in Table 3, where PURIFY consistently recovers images with higher dynamic range. The rms of the residuals around the scientific region of interest (see Table 4) show that PURIFY consistently fits the measurements better than CLEAN.

Table 4 compares the rms of the residual images with in the regions shown in Figs 7–10. Other than 3C129, PURIFY shows a consistent order of magnitude improvement in the rms of the residuals.

7.3 Discussion

From a scientific standpoint, the PURIFY models show more structure than those recovered by CLEAN. This is clear when looking at the surface brightness variation of the jets of 3C129 and Cygnus A. For 3C129 and Cygnus A, unlike the CLEAN restored images, the surface brightness structure is well resolved in the images recovered by PURIFY.

The CLEAN restored images of PKS J0334-39 and PKS J0116-473 with uniform weighting show an improvement over natural weighting for deconvolving the fine structure, as well as containing diffuse structure. However, uniform weighting is known to suppress large-scale structure and lowers the sensitivity of the observation (as discussed in Hindson et al. 2014). However, PURIFY has the ability to reconstruct the fine details of PKS J0334-39 and PKS J0116-473 without uniform weighting. This demonstrates that PURIFY has the potential to reconstruct observations that can be used to perform a more detailed analysis of morphology and structure of diffuse sources. The reconstruction of Cygnus A shows that it is possible

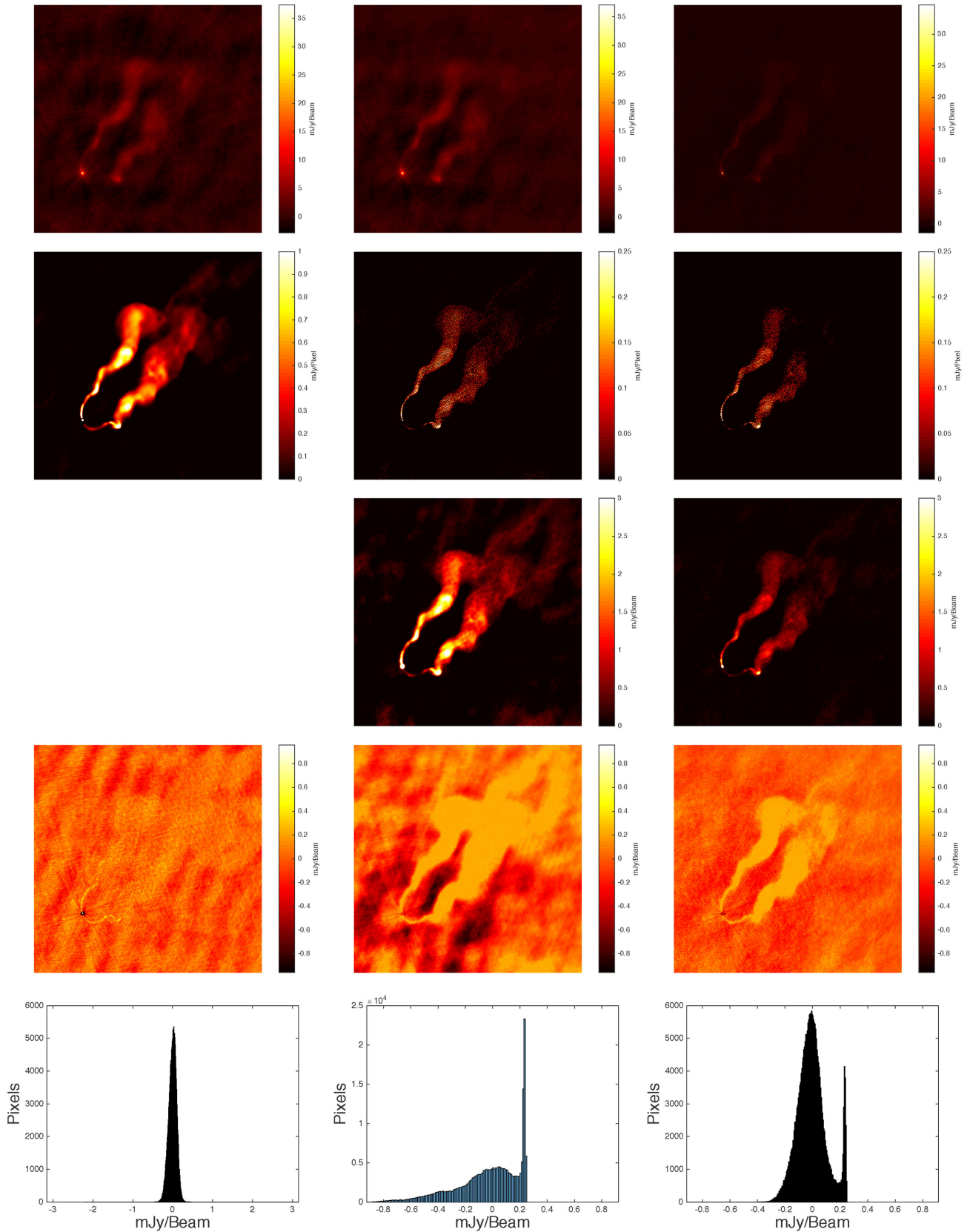


Figure 7. PURIFY and CLEAN reconstructions of 3C129. Each pixel is 0.4 arcsec, and the images are 1024×1024 pixels. The pixels within $[400, 900] \times [400, 900]$ are shown in the images and histogram of this figure. Left column shows a PURIFY reconstruction with natural weighting. Middle and right columns show CLEAN reconstructions with natural and uniform weightings, respectively. From the top to bottom row: synthesized (i.e. dirty) image, model image, restored image, residual image and a histogram of residual image. PURIFY does not require any post-processing and so does not produce a restored image.

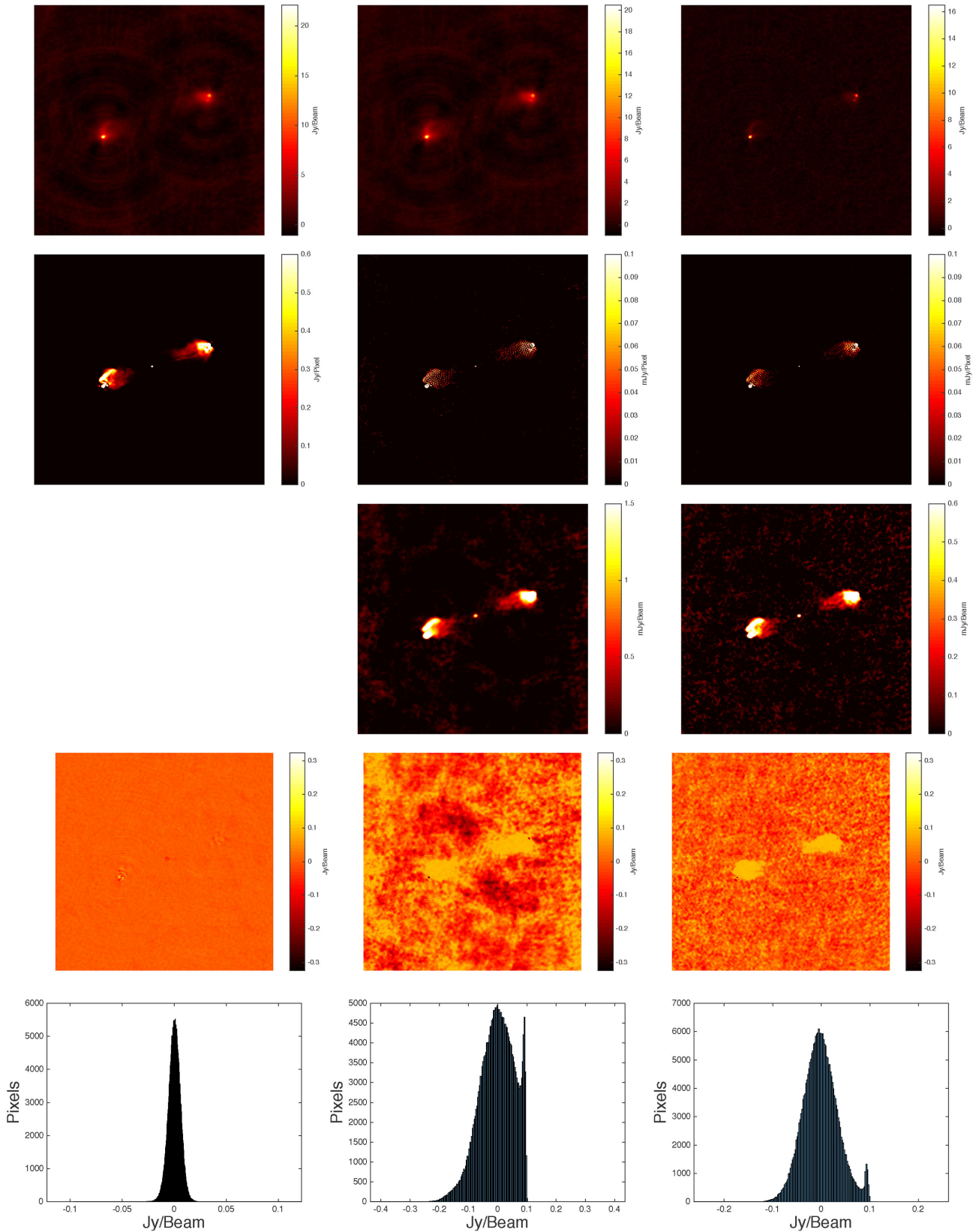


Figure 8. PURIFY and CLEAN reconstructions of Cygnus A. Each pixel is 0.5 arcsec, and the images are 1024×1024 pixels. The pixels within $[256, 756] \times [256, 756]$ are shown in the images and histogram of this figure. Left column shows a PURIFY reconstruction with natural weighting. Middle and right columns show CLEAN reconstructions with natural and uniform weightings, respectively. From the top to bottom row: synthesized (i.e. dirty) image, model image, restored image, residual image and a histogram of residual image. PURIFY does not require any post-processing and so does not produce a restored image.

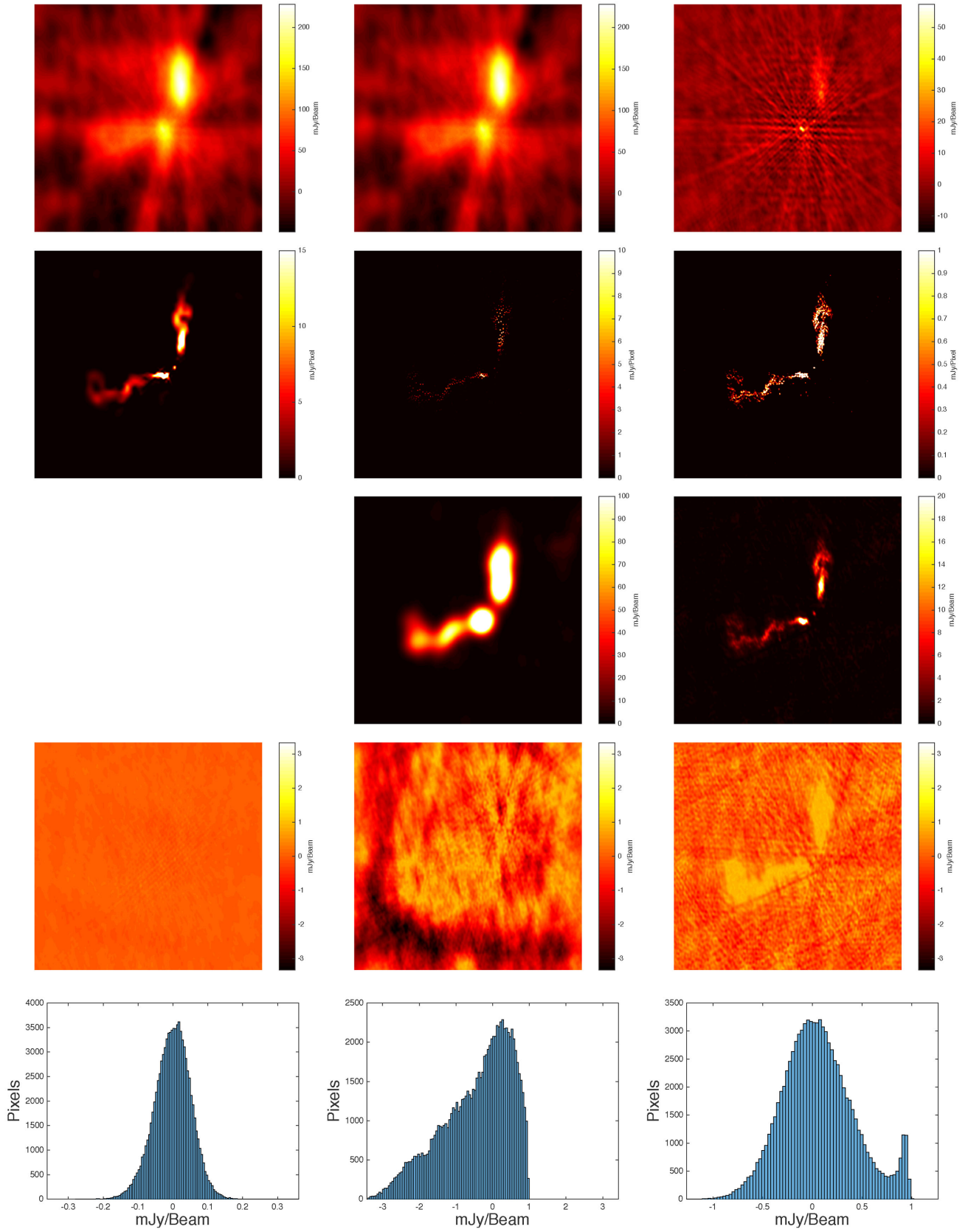


Figure 9. PURIFY and CLEAN reconstructions of PKS J0334-39. Each pixel is 2 arcsec, and the images are 2048×2048 pixels. The pixels within $[862, 1162] \times [862, 1162]$ are shown in the images and histogram of this figure. Left column shows a PURIFY reconstruction with natural weighting. Middle and right columns show CLEAN reconstructions with natural and uniform weightings, respectively. From the top to bottom row: synthesized (i.e. dirty) image, model image, restored image, residual image and a histogram of residual image. PURIFY does not require any post-processing and so does not produce a restored image.

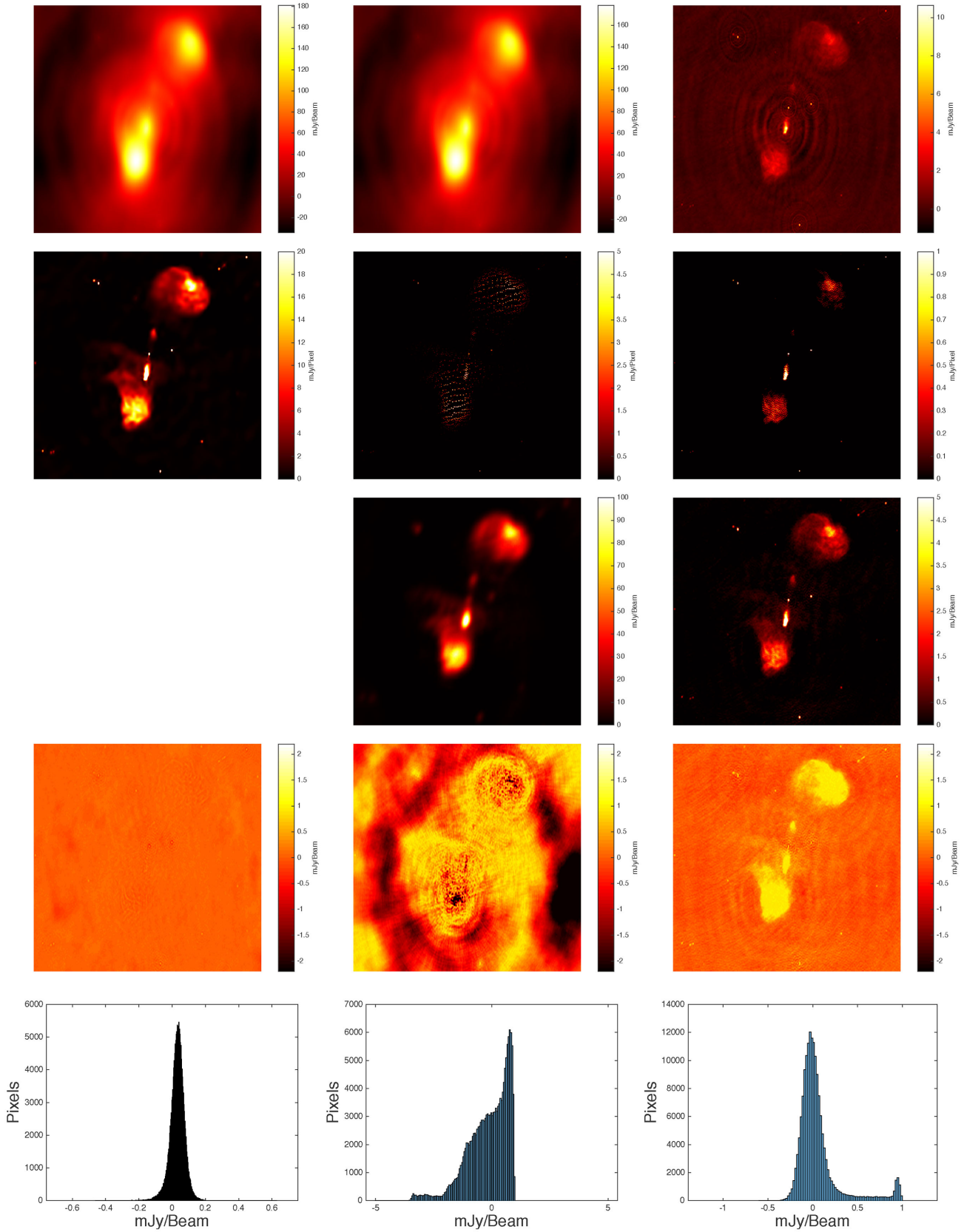


Figure 10. PURIFY and CLEAN reconstructions of PKS J0116-473. Each pixel is 2.4 arcsec, and the images are 2048×2048 pixels. The pixels within $[800, 1200] \times [800, 1200]$ are shown in the images and histogram of this figure. Left column shows a PURIFY reconstruction with natural weighting. Middle and right columns show CLEAN reconstructions with natural and uniform weightings, respectively. From the top to bottom row: synthesized (i.e. dirty) image, model image, restored image, residual image and a histogram of residual image. PURIFY does not require any post-processing and so does not produce a restored image.

Table 3. Table listing the dynamic range of each reconstruction. When computing the dynamic range for PURIFY reconstructions the calculation includes the norm of the measurement operator, so the dynamic range does not scale arbitrarily under the choice of the norm of the measurement operator. For CLEAN, we follow the standard approach and use the peak of the *restored* image, divided by the rms of the residual image.

Observation	PURIFY	CLEAN (natural)	CLEAN (uniform)
3C129	72 444	220	495
Cygnus A	312 928	372	472
PKS J0334-39	1 701 050	208	263
PKS J0116-473	1 185 700	153	361

Table 4. Table listing the rms of each reconstruction (units are in mJy Beam⁻¹).

Observation	PURIFY	CLEAN (natural)	CLEAN (uniform)
3C129	0.10	0.23	0.11
Cygnus A	6.1	59	36
PKS J0334-39	0.052	1.00	0.37
PKS J0116-473	0.054	0.88	0.24

to accurately reconstruct diffuse bright structures in the presence of compact bright sources.

Modelling extended structure accurately is particularly important for understanding the underlying physics of radio sources and their environment. Bent tailed radio galaxies, such as 3C129, are an example of where this is important (Miley et al. 1972). The morphology of bent tailed radio galaxies can be used as a probe of their local cluster environment (Gunn & Gott 1972; Freeland, Cardoso & Wilcots 2008; Douglass et al. 2011; Pfrommer & Jones 2011; Pratley et al. 2013, 2015).

Additionally, an important class of diffuse, low surface brightness radio sources are cluster relics and haloes (e.g. Brunetti et al. 2008; Hindson et al. 2014; Martinez Aviles et al. 2016; Shakouri, Johnston-Hollitt & Pratt 2016), which are believed to be caused by shocks and turbulence in the outskirts of galaxy clusters (Cassano et al. 2013, 2015). Radio haloes and relics are not well understood, and they are prime examples of sources with diffuse low surface brightness structure that relates to the physics within the intra-cluster medium and merging galaxy clusters. However, galaxy clusters often contain bright compact sources, providing a challenge in deconvolving low surface brightness sources. PURIFY's ability to accurately model extended structure and reconstruct images with high dynamic range has the potential to improve scientific interpretations of many radio interferometric observations.

8 CONCLUSIONS

In this work, we have further developed the PURIFY software package so that it can be easily applied to observational data from radio interferometric telescopes. PURIFY has been completely redesigned and reimplemented in C++ and now supports the P-ADMM algorithm developed recently by Onose et al. (2016). Furthermore, the capabilities of convolutional degridging in the measurement operator have been expanded.

Using simulations, we studied the impact of a number of different interpolation kernels on the quality of images recovered by sparse reconstruction approaches to interferometric imaging. The Kaiser-Bessel kernel was found to perform very well – as well as other

optimal kernels – while requiring a smaller support size, thereby reducing computation cost, and having an analytic expression that can be evaluated easily and efficiently.

PURIFY was applied to observational data from the VLA and ATCA telescopes, recovering high-quality interferometric images superior to those recovered by CLEAN. First, the PURIFY residuals contain less extended structure and are more Gaussian with a lower rms. Secondly, the model images recovered by PURIFY are of sufficient quality that there is no need to perform any post-processing as is done for CLEAN (such as restoring the image). Thirdly, all images recovered by PURIFY show an increase in dynamic range when compared to those recovered by CLEAN, in some cases in excess of an order of magnitude. On visual inspection, the images recovered by PURIFY reveal extended structure in greater detail. For example, in reconstructed images of 3C129, the internal structure of the radio jets is much more apparent (Fig. 7). Such an improvement in reconstruction quality can be important in facilitating a better scientific understanding of astrophysical processes.

While the current version of PURIFY can be readily used to recover high-fidelity images from observations made by radio interferometric telescopes, numerous extensions and improvements are planned for future releases. In future, we will implement the primal dual algorithm of Onose et al. (2016), highly distribute and parallelize the algorithms supported following the strategies outlined in Carrillo et al. (2014) and Onose et al. (2016), and add support for direction-dependent effects following the approach outlined in Wolz et al. (2013), for example.

ACKNOWLEDGEMENTS

LP thanks Melanie Johnston-Hollitt for discussions on the practical constraints of current interferometric imaging and useful comments on the manuscript, Xiaohao Cai for discussions on optimization and useful comments on the manuscript, and Arwa Dabbech for general discussions. We thank Rick Perley and Oleg Smirnov for making the VLA observation of Cygnus A available and thank André Ofriinga for assistance with WSCLEAN. This work was supported by the UK Engineering and Physical Sciences Research Council (EPSRC, grants EP/M011089/1 and EP/M008843/1) and the UK Science and Technology Facilities Council (STFC, grant ST/M00113X/1).

REFERENCES

- Bhatnagar S., Cornwell T. J., 2004, *A&A*, 426, 747
 Briggs D., 1995, Ph.D. thesis, The New Mexico Institute of Mining and Technology, New Mexico
 Briggs D. S., Schwab F. R., Sramek R. A., 1999, in Taylor G. B., Carrilli C. L., Perley R. A., eds, *ASP Conf. Ser. Vol. 180, Synthesis Imaging in Radio Astronomy II*. Astron. Soc. Pac., San Francisco, p. 127
 Brunetti G. et al., 2008, *Nature*, 455, 7215, 944
 Candès E. J., Wakin M. B., 2008, *IEEE Signal Process. Mag.*, 25, 21
 Candès E., Romberg J., Tao T., 2006a, *IEEE Trans. Inf. Theory*, 52, 489
 Candès E., Romberg J., Tao T., 2006b, *Comm. Pure and Appl. Math.*, 59, 1207
 Candès E. J., Wakin M. B., Boyd S. P., 2008, *J. Fourier Anal. Appl.*, 14, 877
 Carozzi T. D., Woan G., 2009, *MNRAS*, 395, 1558
 Carrillo R. E., McEwen J. D., Wiaux Y., 2012, *MNRAS*, 426, 1223
 Carrillo R. E., McEwen J. D., Van De Ville D., Thiran J. P., Wiaux Y., 2013, *IEEE Signal Process. Lett.*, 20, 591
 Carrillo R. E., McEwen J. D., Wiaux Y., 2014, *MNRAS*, 439, 3591
 Cassano R. et al., 2013, *ApJ*, 777, 141
 Cassano R. et al., 2015, *Proceedings of Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, 2014 June 9–13. Giardini Naxos, Italy, id. 73

- Clark B. G., 1980, *A&A*, 89, 377
- Combettes P., Pesquet J. C., 2011, *Proximal Splitting Methods in Signal Processing*. Springer, New York
- Cornwell T. J., 2008, *IEEE J. Sel. Top. Signal Process.*, 2, 793
- Cornwell T. J., Evans K. F., 1985, *A&A*, 143, 77
- Dabbech A., Ferrari C., Mary D., Slezak E., Smirnov O., Kenyon J. S., 2015, *A&A*, 576, A7
- Dewdney P., Turner W., Millenaar R., McCool R., Lazio J., Cornwell T., 2013, Document number SKA-TEL-SKO-DD-001 Revision, 1, 1
- Donoho D. L., 2006, *IEEE Trans. Inf. Theory*, 52, 1289
- Douglass E. M., Blanton E. L., Clarke T. E., Randall S. W., Wing J. D., 2011, *ApJ*, 743, 199
- Fessler J. A., Sutton B. P., 2003, *IEEE Trans. Signal Process.*, 51, 560
- Freeland E., Cardoso R. F., Wilcots E., 2008, *ApJ*, 685, 858
- Garsden H. et al., 2015, *A&A*, 575, A90
- Greisen E. W., 1979, *The Effects of Various Convolution Functions on Aliasing and Relative Signal-to-Noise Ratios*, VLA Scientific Memorandum 131, National Radio Astronomy Observatory, Charlottesville, Virginia
- Greisen E. W., 1998, *The Creation of AIPS*, AIPS Memorandum 100, National Radio Astronomy Observatory, Charlottesville, Virginia
- Guennebaud G. et al., 2010, *Eigen v3*, <http://eigen.tuxfamily.org>
- Gunn J. E., Gott J. R., III, 1972, *ApJ*, 176, 1
- Hindson L. et al., 2014, *MNRAS*, 445, 330
- Hoaglin D. C., Mosteller F., Tukey J. W., 2000, *Understanding Robust and Exploratory Data Analysis*. Wiley Classics Library, Wiley-Interscience, New York
- Högbom J. A., 1974, *A&AS*, 15, 417
- Hotan A. W. et al., 2014, *PASA*, 31, e041
- Hurley-Walker N. et al., 2017, *MNRAS*, 464, 1146
- Jackson J. I., Meyer C. H., Nishimura D. G., Macovski A., 1991, *IEEE Trans. Med. Imaging*, 10, 473
- Jacobs D. C. et al., 2016, *ApJ*, 825, 114
- Johnston-Hollitt M. et al., 2015, *Proceedings of Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, 2014 June 9–13, Giardini Naxos, Italy, id. 92
- Koopmans L. et al., 2015, *Proceedings of Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, 2014 June 9–13, Giardini Naxos, Italy, id. 1
- Landau H. J., Pollak H. O., 1961, *Bell Syst. Tech. J.*, 40, 65
- Landau H. J., Pollak H. O., 1962, *Bell Syst. Tech. J.*, 41, 1295
- Lenc E. et al., 2016, *ApJ*, 830, 38
- Li F., Brown S., Cornwell T. J., de Hoog F., 2011a, *A&A*, 531, A126
- Li F., Cornwell T. J., de Hoog F., 2011b, *A&A*, 528, A31
- Maartens R., Abdalla F. B., Jarvis M., Santos M. G., 2015, *Proceedings of Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, 2014 June 9–13, Giardini Naxos, Italy, id. 16
- Mallat S. G., Zhang Z., 1993, *IEEE Trans. Signal Process.*, 41, 3397
- Marsh K. A., Richardson J. M., 1987, *A&A*, 182, 174
- Martinez Aviles G. et al., 2016, *A&A*, 595, A116
- McCready L. L., Pawsey J. L., Payne-Scott R., 1947, *Proc. R. Soc. London Ser. A*, 190, 357
- McEwen J. D., Scaife A. M. M., 2008, *MNRAS*, 389, 1163
- McEwen J. D., Wiaux Y., 2011, *MNRAS*, 413, 1318
- McMullin J. P., Waters B., Schiebel D., Young W., Golap K., 2007, in Shaw R. A., Hill F., Bell D. J., eds, *ASP Conf. Ser. Vol. 376, Astronomical Data Analysis Software and Systems XVI*. Astron. Soc. Pac., San Francisco, p. 127
- Miley G. K., Perola G. C., van der Kruit P. C., van der Laan H., 1972, *Nature*, 237, 269
- Murphy T. et al., 2015, *MNRAS*, 446, 2560
- Oberoi D., Attridge J., Doeleman S., 2003, *PSFs and Best Fits Beams in AIPS and MIRIAD*, LOFAR Memorandum 6, Haystack Observatory, Massachusetts Institute of Technology, Westford, Massachusetts
- Offringa A. R. et al., 2014, *MNRAS*, 444, 606
- Offringa A. R. et al., 2016, *MNRAS*, 458, 1057
- Onose A., Carrillo R. E., Repetti A., McEwen J. D., Thiran J. P., Pesquet J. C., Wiaux Y., 2016, *MNRAS*, 462, 4314
- Patel P. et al., 2015, *Proceedings of Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, 2014 June 9–13, Giardini Naxos, Italy, id. 30
- Pfrommer C., Jones T. W., 2011, *ApJ*, 730, 22
- Pratley L., Johnston-Hollitt M., 2016, *MNRAS*, 462, 3483
- Pratley L., Johnston-Hollitt M., Dehghan S., Sun M., 2013, *MNRAS*, 432, 243
- Pratley L., Johnston-Hollitt M., Dehghan S., Sun M., 2015, in Massaro F., Cheung C. C., Lopez E., Siemiginowska A., eds, *Proc. IAU Symp. 313, Extragalactic Jets from Every Angle*. Cambridge Univ. Press, p. 301
- Price D. C., Smirnov O. M., 2015, *MNRAS*, 449, 107
- Rousseeuw P. J., Croux C., 1993, *J. Am. Stat. Assoc.*, 88, 1273
- Rubinstein R., Bruckstein A. M., Elad M., 2010, *Proc. IEEE*, 98, 1045
- Ryle M., Hewish A., 1960, *MNRAS*, 120, 220
- Saripalli L., Subrahmanyan R., Shankar N. U., 2002, *ApJ*, 565, 256
- Sault R. J., Teuben P. J., Wright M. C. H., 1995, in Shaw R. A., Payne H. E., Hayes J. J. E., eds, *ASP Conf. Ser. Vol. 77, Astronomical Data Analysis Software and Systems IV*. Astron. Soc. Pac., San Francisco, p. 433
- Sault R. J., Staveley-Smith L., Brouw W. N., 1996, *A&AS*, 120, 375
- Schwab F. R., 1978, *Suppression of Aliasing by Convolutional Gridding Schemes*, VLA Scientific Memorandum 129, National Radio Astronomy Observatory, Charlottesville, Virginia
- Schwab F. R., 1980, *Optimal Gridding*, VLA Scientific Memorandum 132, National Radio Astronomy Observatory, Charlottesville, Virginia
- Schwab F. R., 1984a, in Roberts J. A., ed., *Indirect Imaging. Measurement and Processing for Indirect Imaging*. Cambridge Univ. Press, Sydney, p. 333
- Schwab F. R., 1984b, *AJ*, 89, 1076
- Shakouri S., Johnston-Hollitt M., Pratt G. W., 2016, *MNRAS*, 459, 2525
- Shannon C. E., 1949, *Proc. IRE*, 37, 10
- Slepian D., Pollak H. O., 1961, *Bell Syst. Tech. J.*, 40, 1, 43
- Smirnov O. M., 2011, *A&A*, 531, A159
- Steer D. G., Dewdney P. E., Ito M. R., 1984, *A&A*, 137, 159
- Stratton J. A., 1935, *Proc. Natl. Acad. Sci.*, 21, 51
- Sun X. H. et al., 2015, *AJ*, 149, 60
- Thompson A. R., 1999, in Taylor G. B., Carilli C. L., Perley R. A., eds, *ASP Conf. Ser. Vol. 180, Synthesis Imaging in Radio Astronomy II*. Astron. Soc. Pac., San Francisco, p. 11
- Thompson A. R., Moran J. M., Swenson G. W., Jr., 2008, *Interferometry and Synthesis in Radio Astronomy*. John Wiley & Sons
- Tingay S. J. et al., 2013, *PASA*, 30, e007
- van Haarlem M. P. et al., 2013, *A&A*, 556, A2
- Wayth R. B. et al., 2015, *PASA*, 32, e025
- Whittaker E. T., 1915, *Proc. R. Soc. Edinburgh*, 35, 181
- Wiaux Y., Jacques L., Puy G., Scaife A. M. M., Vanderghenst P., 2009a, *MNRAS*, 395, 1733
- Wiaux Y., Puy G., Boursier Y., Vanderghenst P., 2009b, *MNRAS*, 400, 1029
- Wolz L., McEwen J. D., Abdalla F. B., Carrillo R. E., Wiaux Y., 2013, *MNRAS*, 436, 1993
- Zernike F., 1938, *Physica*, 5, 785
- Zhang L., Bhatnagar S., Rau U., Zhang M., 2016, *A&A*, 592, A128

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