Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **438,** 768–778 (2014) doi:10.1093/mnras/stt2244
Advance Access publication 2013 December 14

Probabilistic image reconstruction for radio interferometers

P. M. Sutter,[1,2,3,4★] Benjamin D. Wandelt,[2,3,4,5] Jason D. McEwen,[6,7] Emory F. Bunn,[8] Ata Karakci,[9] Andrei Korotkov,[9] Peter Timbie,[10] Gregory S. Tucker[9] and Le Zhang[10]

[1]*Center for Cosmology and Astro-Particle Physics, Ohio State University, Columbus, OH 43210, USA*
[2]*UPMC Univ Paris 06, UMR7095, Institut d'Astrophysique de Paris, F-75014 Paris, France*
[3]*CNRS, UMR7095, Institut d'Astrophysique de Paris, F-75014 Paris, France*
[4]*Department of Physics, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA*
[5]*Department of Astronomy, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA*
[6]*Department of Physics and Astronomy, University College London (UCL), London WC1E 6BT, UK*
[7]*Mullard Space Science Laboratory (MSSL), University College London (UCL), Surrey RH5 6NT, UK*
[8]*Physics Department, University of Richmond, Richmond, VA 23173, USA*
[9]*Department of Physics, Brown University, 182 Hope Street, Providence, RI 02912, USA*
[10]*Department of Physics, University of Wisconsin, Madison, WI 53706, USA*

Accepted 2013 November 17. Received 2013 October 27; in original form 2013 August 15

ABSTRACT

We present a novel, general-purpose method for deconvolving and denoizing images from gridded radio interferometric visibilities using Bayesian inference based on a Gaussian process model. The method automatically takes into account incomplete coverage of the *uv*-plane, signal mode coupling due to the primary beam and noise mode coupling due to *uv* sampling. Our method uses Gibbs sampling to efficiently explore the full posterior distribution of the underlying signal image given the data. We use a set of widely diverse mock images with a realistic interferometer set-up and level of noise to assess the method. Compared to results from a proxy for point source-based CLEAN method we find that in terms of rms error and signal-to-noise ratio our approach performs better than traditional deconvolution techniques, regardless of the structure of the source image in our test suite. Our implementation scales as $\mathcal{O}(n_{\rm p} \log n_{\rm p})$ provides full statistical and uncertainty information of the reconstructed image, requires no supervision and provides a robust, consistent framework for incorporating noise and parameter marginalizations and foreground removal.

Key words: instrumentation: interferometers – methods: data analysis – methods: statistical.

1 INTRODUCTION

The next generation of large-scale radio interferometers, such as Australian Square Kilometre Array Pathfinder (ASKAP; Johnston et al. 2008), Murchison Widefield Array (MWA; Lonsdale et al. 2009), Precision Array to Probe the Epoch of Reionisation (PAPER; Parsons et al. 2010) and Square Kilometre Array (SKA; Jarvis 2007), promise incredible scientific reward but also incredible data analysis challenges. The tremendous volume of data, high dynamic range, wide bandwidth, large amounts of radio interference and significant foregrounds present serious instrumentation and analysis difficulties (Bhatnagar 2009; Norris et al. 2013). Image deconvolution – the process of removing the effects of signal mode coupling due to the primary beam, noise mode coupling due to *uv* sampling, incomplete Fourier mode sampling and noise – is an important first step in scientific analysis from these instruments (Thompson, Moran & Swenson 2001). The ideal algorithm for performing image deconvolution suitable for these upcoming surveys would be (1) robust to changes in the input signal and instrument configuration, (2) parametrized by as few tunable inputs as possible, (3) as automated as possible, (4) driven by the *data* rather than user-selected guesses as to the source signal, (5) as informative as possible given the difficulty of processing the data even a single time, (6) scalable to the extreme sizes of future data sets, (7) as fast as possible and (8) able to easily incorporate modelling of foregrounds or other systematics.

Unfortunately, the most commonly used algorithm for interferometric image reconstruction, CLEAN (Högbom 1974), satisfies few – if any – of these criteria. Originally developed to remove point source foregrounds from a smooth background source image, CLEAN works by iteratively removing the effects of high-intensity point sources in the Fourier-transformed *uv*-plane. While the original algorithm excels in a limited number of cases, it is not appropriate for general reconstruction problems. Modern implementations of CLEAN that appear in standard radio astronomy packages (e.g. CASA; Jaeger 2008) incorporate several advanced features such as simultaneous deconvolution at multiple scales (Cornwell 2008; Rau & Cornwell

★E-mail: sutter@iap.fr

© 2013 The Authors
Published by Oxford University Press on behalf of the Royal Astronomical Society

Downloaded from https://academic.oup.com/mnras/article/438/1/768/1041817 by guest on 22 April 2022

2011) and time-varying analysis (Stewart, Fenech & Muxlow 2011; Rau 2012).

However, even these more advanced versions of CLEAN require significant fine-tuning and supervision during the deconvolution process. Users must select thresholds for deciding which pixels contain 'point sources' and the amount by which to remove the point source effects in uv-space. The optimal choices for these thresholds are not known in advance for a given observation. Users may also select specific regions of the image to apply more or fewer CLEANing iterations. CLEAN has no final end state: users must decide when an image has been deconvolved 'enough' without introducing unwanted artefacts. Finally, CLEAN gives no information on the uncertainties in the reconstruction.

Realizing the shortcomings of traditional and more sophisticated CLEAN-based algorithms, methods based on a regularized likelihood, such as the maximum entropy method (MEM; Ables 1974; Gull & Daniell 1978; Cornwell & Evans 1985) generated significant interest. MEM produces an image estimate which minimizes the difference between the estimate and the data given the level of noise and some chosen metric. While MEM requires less fine-tuning and supervision than CLEAN, the user must still choose the metric (usually expressed as an entropy functional), and the optimal choice of the metric is not known in advance (Starck et al. 2001). The final reconstructed image is thus only optimal in regards to that metric. Similarly to CLEAN, MEM gives no uncertainty information and in general tends to underestimate the source image intensity (Starck et al. 2001; Sutton & Wandelt 2006).

As Bhatnagar & Cornwell (2004) and Puetter, Gosnell & Yahil (2005) pointed out, optimal reconstruction techniques must be spatially adaptive and operate at multiple scales simultaneously. Many authors have proposed relatively new alternative methods based on compressed sensing techniques (Suksmono 2009; Wiaux et al. 2009, 2010; Wenger et al. 2010; Li et al. 2011; McEwen & Wiaux 2011; Carrillo, McEwen and Wiaux 2012, 2013b; Wolz et al. 2013), Bayesian processes (Ayasso, Rodet & Abergel 2012) and separation of smooth and point-like components (Giovannelli & Coulais 2005).

In this work, we present a general-purpose Bayesian reconstruction algorithm to infer the source image from realistic interferometric radio data. Bayesian analysis starts with building a generative probabilistic forward model of the data. Given this model choice, the posterior probability density function (the *posterior distribution*) quantifies what is known about the source image once the data have been obtained (Gelman et al. 2004). The purpose of this paper is to demonstrate that a surprisingly simple choice of model for the source image, that of an *isotropic Gaussian process*, performs very well in realistically simulated examples for a test suite of widely diverse images.

The key to practical Bayesian image analysis is to be able to navigate efficiently through the very high dimensional parameter space, since every pixel value is an independent parameter. For example, in this paper we will explore posterior distributions in 10^4-dimensional parameter spaces. Gibbs sampling is a powerful technique that has been used successfully for sky maps with more than 10^6 pixels in the context of cosmic microwave background signal reconstruction and power spectrum estimation (Jewell, Levin & Anderson 2004; Wandelt, Larson & Lakshminarayanan 2004; Sutter, Wandelt & Malu 2012). Conceptually, Gibbs sampling iterates between samples of the signal and its power spectrum in a way that respects the joint posterior distribution of signal and power spectrum given the image. This separation allows for significant speedups compared to grid-based evaluations of the posterior: the algorithm scales as $n_{\rm p} \log n_{\rm p}$, where $n_{\rm p}$ is the total number of pixels, in the ideal pre-gridded flat-sky limit discussed here.

Our assumptions for our mock observations allow us to pre-grid the intensities before the iterative solution step. More general curved-sky analysis would necessarily be more expensive, either with gridding–regridding steps during the analysis in an AW-projection method or with the spherical harmonic transform operation in a full sky. This is similar to the computational scalability of traditional CLEAN. However, CLEAN typically completes in order $\sim$10 steps, whereas our method usually requires $\sim$100 iterations. As we will see below, the primary computational cost in each iteration comes from solving a matrix–vector equation. Fortunately, this is a common problem in computational science and there are many fast, scalable solutions available (Press, Flannery & Teukolsky 1986).

While the Gibbs sampling framework itself is independent of prior (see e.g. Sutton & Wandelt 2006 for an implementation based on fluxon models) we choose an isotropic Gaussian process prior. Gaussian processes are surprisingly flexible in describing a variety of images (Mackay 2003). Because Gibbs sampling can be understood as a non-linear generalization of the least-squares optimal signal reconstruction provided by the Wiener filter (see Section 3) without requiring a choice of the signal covariance a priori, successive samples are always constrained by the data in regions of high signal-to-noise ratio (SNR). In regions of low SNR, the Gaussian process is the least informative completion to a full probabilistic model. In this regime the method still maps out the signal likelihood taking into account all modelled signals and imperfections in the data at the two-point correlation level.

In addition to the speed gains mentioned above, the method of Gibbs sampling offers several advantages over traditional deconvolution techniques: (1) it explores the full posterior shape, giving *complete* statistical information on the resulting image; (2) it *automatically* takes into account signal mode coupling from the primary beam; (3) the sampled representation produced by the method makes marginalization *trivial*; (4) it provides optimal reconstruction (i.e. Wiener filtering) *without* assuming a signal covariance; (5) since we have already specified the signal prior, the method has *no* tunable parameters and operates completely unsupervised and (6) it offers a *sparse* reconstruction technique with a full Bayesian motivation.

Section 2 presents our mock observations and interferometer set-up used to assess our new method. Next, in Section 3 we outline the method of Gibbs sampling as applied to radio interferometers. In Section 4 we compare our method to a proxy for traditional CLEAN in terms of reconstructed image fidelity, residuals and statistics such as rms error and residual SNR levels. Finally, we conclude in Section 5 with a discussion of planned extensions and the potential role of this method in future observations.

2 SIMULATED OBSERVATIONS

2.1 Interferometer set-up

We model the visibility data d obtained from an interferometric observation as

$$d = IFAs + In, \tag{1}$$

where s is a vector containing a discretization of the input sky, A is a primary beam pattern, F is a Fourier transform operator that converts from pixel space to the uv-plane, I is an interferometer pattern in the uv-plane and n is a Gaussian realization of the noise. We discretize the signal s, data d and noise n with $n_{\rm p}$ elements.

Downloaded from https://academic.oup.com/mnras/article/438/1/768/1041817 by guest on 22 April 2022

We generate all input visibilities within a 20 deg^2 patch discretized to 128 pixels per side, similar to the approach of Myers et al. (2003) and Sutter et al. (2012). While this size of patch violates the strict flat-sky approximation, it allows us to explore the validity of our technique at higher resolutions and probe the range of scales accessible to realistic interferometers. Because it is easier in interferometry to perform calculations in visibility space than in image space (Baron, Kloppenborg & Monnier 2012), we do not establish an observation wavelength for these mock observations, since our results and conclusions are independent of wavelength.

We model the primary beam pattern A as a Gaussian with peak value of unity and standard deviation 1°.5. With these parameters the primary beam decreases to a value of 10^{-3} halfway to the edge of the box. This allows us to include all Fourier modes up to the Nyquist frequency in our analysis and ensures that the periodic boundary conditions inherent in the Fourier transform do not cause unwanted edge effects. We prevent the primary beam from reaching values below 10^{-4}. This value reflects a balance between faithfully representing the suppression of signals far from the point centre (so that Fourier transforms have correct periodic boundaries) and the need to preserve numerical stability in the conjugate-gradient algorithm utilized in the method below.

We assemble the interferometer array in a simple way by randomly placing 12 antennas and selecting all baseline pairs within the uv-plane. We then allow the assembly to rotate uniformly for 6 h while observing the same sky patch at the north celestial pole. This choice of antenna arrangement and integration time roughly corresponds to existing instruments, such as an extended configuration of Atacama Large Millimeter/submillimeter Array (ALMA; Nyman et al. 2010) or Very Large Array (VLA; Perley et al. 2011), although our fiducial set-up uses fewer elements to highlight the performance of our method in less-than-optimal observing regimes. The interferometer pattern is discretized to the same resolution as our input images ($n_\mathrm{p} = 128$ pixels on a side; described below). We construct the interferometer pattern I by placing a value of one wherever a baseline length intersects a pixel during its rotation and zeros elsewhere. We show the resulting uv-plane coverage in Fig. 1. This

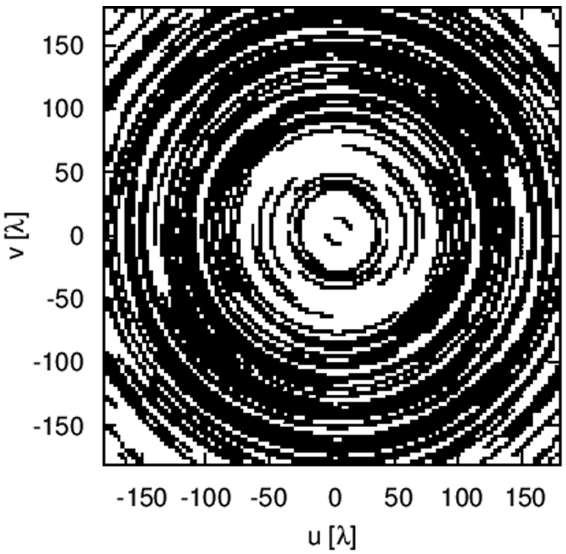


Figure 1. Fiducial interferometer pattern (black) after randomly placing 12 antennas and integrating for 6 h assuming a pointing at the north celestial pole. The interferometer pattern is discretized to the same resolution as our input images ($n_\mathrm{p} = 128$ pixels on a side).

configuration covers roughly 70 per cent of the uv-plane, although the coverage varies significantly for each ℓ-bin, where ℓ is the radius of a given annulus in the uv-plane. Some bins, especially at very low and very high ℓ, have zero coverage due to the lack of baselines at that distance. However, even if these bins had adequate coverage, we expect statistics here to be relatively poor due to the reduced number of modes in these regions. Most bins have at least 60 per cent coverage and several bins have complete coverage.

We determine the noise per pixel by summing the integration time spent in that pixel by all baselines. We do not adopt a noise model for a particular instrument; rather, we set the noise variance to be

$$\sigma_i^2 \propto 1/t_{\mathrm{obs},i}\,, \tag{2}$$

where t_obs is the observation time in pixel i. We then set an overall SNR of 10 by multiplying all noise variances by a constant value to maintain $|IFAs|/|In| = 10$. This provides a scaling of the noise that would normally be caused by instrument effects such as the effective area of the apertures and the system temperature in a realistic observation. We use a Gaussian realization of this noise variance to generate n in equation (1). We then multiply this noise by the interferometer pattern I to maintain consistency with the signal. Fig. 2 shows a particular noise realization for one of our test images and the resulting data d.

2.2 Test images

Radio interferometers are used to study a wide variety of interesting astrophysical and cosmological phenomena, such as supernova remnants (Bhatnagar et al. 2011), molecular gas clouds (Gratier et al. 2010), cluster radio haloes (Cassano et al. 2010), magnetic fields in dwarf galaxies (Heesen et al. 2011), the interstellar medium (Zhang, Hunter & Elmegreen 2012), the galactic centre (McClure-Griffiths et al. 2012) and the cosmic microwave background (Pearson et al. 2005). To assess the ability of our method to cope with this variety of targets we select eight input images drawn from the CASA (Jaeger 2008) user guide,[1] which we present in Fig. 3.

The test images represent a diverse variety of realistic – and a few unrealistic – imaging scenarios, such as a protoplanetary disc, a face-on spiral galaxy, a cluster, an active galactic nucleus (AGN) jet and lobe and the face of *Einstein*. These images were provided in a mix of resolutions and dynamic ranges. To simplify our analysis (but without loss of generality), we remapped all images to a uniform grid $n_\mathrm{p} = 128$ pixels on a side and renormalized all intensities to a peak of unity. Some test images had artificial artefacts built-in, and the remapping procedure introduced some additional glitches in the image. We left these artefacts intact to test the ability of our isotropic Gaussian process prior to recover highly anisotropic portions of the data.

3 METHOD OF GIBBS SAMPLING

Previous works have extensively discussed Gibbs samples (e.g. Wandelt et al. 2004; Sutter et al. 2012), so we only briefly introduce the relevant equations as applied to interferometric observations here. We begin with some initial guess of the angular power spectrum C_ℓ^0 and progressively iterate samples from the conditional distributions:

$$s^{i+1} \leftarrow P\left(s|C_\ell^i, m\right), \tag{3}$$

[1] http://casaguides.nrao.edu/index.php?title=Sim_Inputs

Downloaded from https://academic.oup.com/mnras/article/438/1/768/1041817 by guest on 22 April 2022

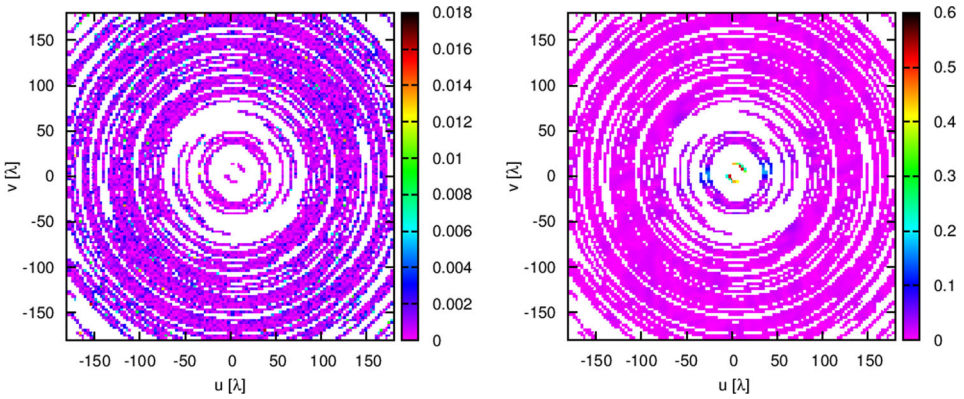


Figure 2. Left: noise realization used in all test cases. The noise is randomly selected assuming a variance given by equation (2) and an overall SNR of 10. Right: example data d (equation 1) in the uv-plane for the *Einstein* test image.

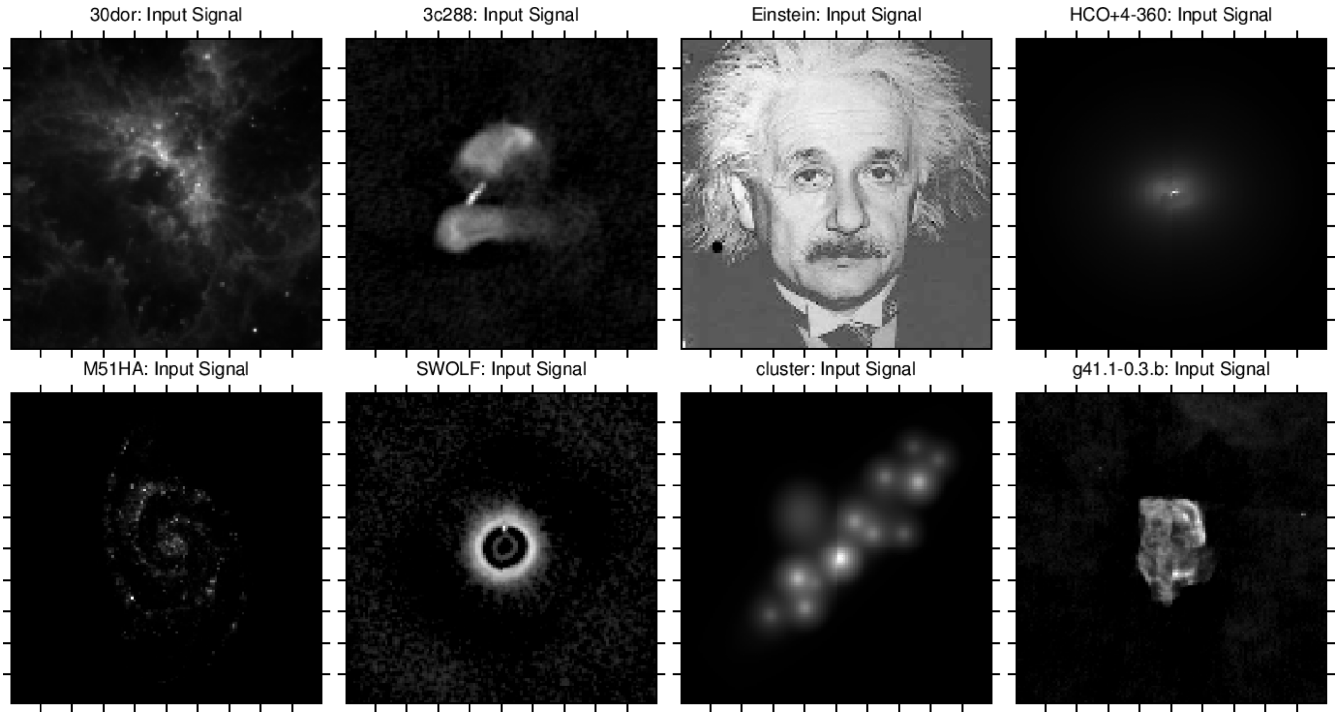


Figure 3. Suite of test input images. All images have been rescaled to 128 × 128 pixels and have had their intensities remapped to a maximum of unity. The repixelization process introduces some artefacts, which we leave in place to assess the ability of our isotropic prior to handle anisotropic data. Each image is 20° across. The colour scale ranges from 0.0 (black) to 1.0 (white).

$$C_\ell^{i+1} \leftarrow P(C_\ell|s^{i+1}), \tag{4}$$

where m is the least squares estimate of the signal s given the data d (i.e. $B^{\rm T}N^{-1}Bm = B^{\rm T}N^{-1}d$, where B represents a full signal-to-data operator). The samples (C_ℓ^i, s^i) converge to samples from the joint distribution $P(C_\ell, s, m) = P(m|s)P(s|C_\ell)P(C_\ell)$ after a sufficient number of iterations.

Given an angular power spectrum sample C_ℓ^i, we generate a new signal sample by drawing from a multivariate Gaussian with mean $S^i(S^i + N)^{-1}m$ and variance $((S^i)^{-1} + N^{-1})^{-1}$. Here S and N are the signal and noise covariance, respectively. We do this by solving the set of equations

$$\begin{aligned} Ms^{i+1} &= A^{\rm T}F^{-1}I(INI)^{-1}d + F^{-1}S^{-1/2}F\xi_1 \\ &\quad + A^{\rm T}F^{-1}I(INI)^{-1/2}F\xi_2, \end{aligned} \tag{5}$$

where we define the matrix operator M as

$$M \equiv F^{-1}S^{-1}F + A^{\rm T}F^{-1}I(INI)^{-1}IFA. \tag{6}$$

In the above equations $A^{\rm T}$ is the primary beam transpose and F^{-1} is the inverse Fourier transform. The first term in the right-hand side of the above equation provides the solution for the Wiener-filtered map, while the second and third terms of equation (5) provide random fluctuations with the required variance. The vectors ξ_1 and ξ_2 are of length $n_{\rm p}$ with elements drawn from a standard normal distribution. As an illustration, Fig. 4 shows an example Wiener-filtered map (i.e. from just solving the first term in the right-hand side) and full signal sample s^i at a single iteration for the *Einstein* test case. We see that the fluctuation terms in equation (5) fill in regions weakly constrained by the data with a guess that mimics the known portions of the signal given the level of noise.

Downloaded from https://academic.oup.com/mnras/article/438/1/768/1041817 by guest on 22 April 2022

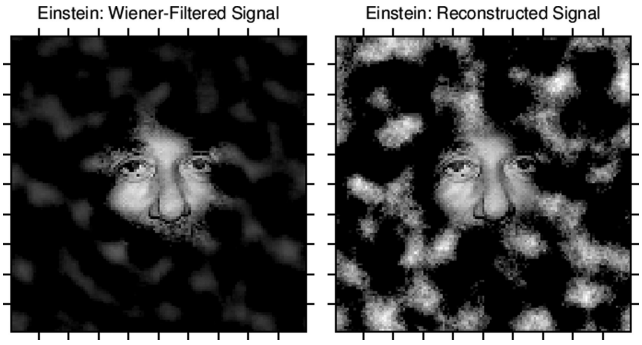


Figure 4. Sample iteration of the Gibbs sampling algorithm for the *Einstein* test image, showing the Wiener-filtered signal (left) and full sample reconstruction with mock fluctuations added (right). Each image is 20° across. The colour scale ranges from 0.0 (black) to 1.0 (white).

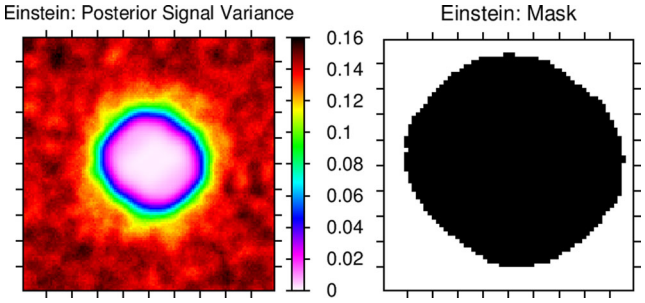


Figure 5. Posterior signal variance in image space (left) and acceptance mask derived from the variance (right) for the *Einstein* test image. The variance map is 20° across, and to show the detailed shape of the variance mask, we have zoomed in on the inner 10°. The acceptance region of the mask is shown in black.

The signal covariance matrix S is diagonal in the uv-plane for isotropic signals (which we assume as part of our Gaussian process prior), so $S_{\ell,\ell'} = C_\ell \delta_{\ell,\ell'}$, where $\ell = 2\pi u$, with u being the radial distance in the uv-plane. Here and throughout we assume the flat-sky approximation that makes this identity valid. By construction, Gibbs sampling explores the exact posterior and therefore treats the couplings introduced by partial sky coverage optimally (Wandelt et al. 2004). The algorithm 'knows' about the couplings since they are contained in the data model (equation 1) which underlies the analysis. However, to a very good approximation the noise covariance matrix N is diagonal, and thus we will assume this for simplicity. We assign to the matrix N entries equal to $N_{i,j} = \sigma_i^2 \delta_{i,j}$, where σ_i is the noise variance for the ith pixel in the uv-plane. The construction $I(INI)^{-1}$ provides a pseudo-inverse of N, so that any locations in the uv-plane with no antenna coverage do not yield infinities when taking the inverse.

We solve numerically the above matrix–vector equation using a preconditioned conjugate-gradient scheme (Press et al. 1986). The preconditioner approximates the diagonal components of M and is

$$P^{-1} = F^{-1} I (INI)^{-1} IF(F^{-1}\tilde{A}^2), \tag{7}$$

where $\tilde{A}$ is the Fourier transform of the primary beam pattern. We implemented the code to solve the above equations with the open-source PETSc library (Balay et al. 1997, 2010, 2011) and the MPI-parallelized version of FFTW (Frigo & Johnson 2005).

Given the latest signal sample, s^i, we generate a new angular power spectrum sample from equation (4) by computing the variance π_l^2 in annuli of constant ℓ on the Fourier-transformed signal. We then use this variance to draw from the probability density $P(C_\ell|s^i)$, which follows an inverse Gamma distribution, by creating a vector p_ℓ of length n_ℓ (assuming a Jeffreys' ignorance prior) and unit Gaussian random elements. Here, n_ℓ is the number of pixels in the bin ℓ. The next power spectrum sample is then simply

$$C_\ell^{i+1} = \frac{\pi_\ell}{|p_\ell|^2}. \tag{8}$$

In the above, we assume $\ell(\ell+1)C_\ell$ to be constant across the width of each annulus in the uv-plane. The width of each annulus can be set as desired. For the test cases which we work with in this paper we chose the width to be $8\pi/L$, where L is the longest baseline of the interferometer. This is four times the uv-space resolution. This choice limits correlations between angular power spectrum bins which develop as a consequence of partial sky coverage. All ℓ-bins have uniform width except for the first, which we restrict to cover only the central zone where we enforce $C_0^i = 0$, since our analysis cannot constrain the DC mode. We wish to capture as much power spectrum information as possible, so we correspondingly widen the width of the second bin to close this gap.

To determine convergence so that our iterative samples from the conditional distributions (equation 4) are indeed samples from the joint distribution, we employ multiple chains with different random number seeds. Our convergence criterion is the Gelman–Rubin (G–R) statistic, which compares the variance among chains to the variance within each chain. The G–R statistic asymptotes to unity, so convergence is said to be achieved when this statistic is below a given tolerance level for each ℓ-bin (Gelman & Rubin 1992). For our test images we stopped iterating when the G–R statistic reached less than 1.1, which took from 500 to 1500 steps, depending on the image.

After convergence, we take the mean of the signal samples as our reconstructed image and simply calculate the variance for each pixel to assess the uncertainty in the reconstruction. Fig. 5 shows the variance for the *Einstein* test image. We naturally see insignificant variance in the centre of the primary beam where the data support is strongest, with steadily increasing variance away from the pointing centre.

We may use the variance to build a mask of the final reconstructed image in which we only accept pixels below a given variance threshold. While this certainly is not necessary for the method to work, it provides an easy way to combine the posterior mean and variance information in a single plot. First, we compute the asymptotic variance $\sigma_{\rm asmp}$, or the mean of the variance along the outer edge of image plane. We then reject any pixel whose variance σ is greater than $\sigma_{\rm asmp}/2$ *and* whose SNR $s/\sqrt{\sigma} < 1$. This last criterion prevents us from masking low-intensity, but still well-constrained, pixels. The values chosen to create the mask are arbitrary, and we choose them solely on aesthetic merit. We repeat: this mask is not necessary to perform the deconvolution and produce a reconstructed signal. We only use it to neatly incorporate the measured uncertainty information in the plotted images.

As we see in Fig. 5, the acceptance mask generally follows the primary beam pattern. The detailed shape near the mask edge is influenced by the relative SNR levels and highlights the non-obvious nature of the regions of reliable reconstruction. We also notice an asymmetry in the acceptance mask: this is in response to the asymmetric nature of the fiducial interferometer set-up (Fig. 1) and provides strong evidence that the symmetric nature of our Gaussian process prior does not greatly influence our inference.

Downloaded from https://academic.oup.com/mnras/article/438/1/768/1041817 by guest on 22 April 2022

4 COMPARISON TO CLEAN

We compare images recovered by Gibbs sampling, as described in the previous section, to those recovered by a proxy for the point source-based CLEAN algorithm. CLEAN is implemented in various radio interferometric imaging packages (such as CASA). However, it is not straightforward to use these packages to reconstruct images from simulated visibilities already defined on gridded coordinates; these packages are instead tailored to analyse observations made by real interferometric telescopes, with data in a specific format. Consequently, we compare to reconstructions made with a proxy for the point source-based CLEAN algorithm. While there are more sophisticated implementations of CLEAN that would undoubtedly perform better with our test images, this gives us a simple standard of comparison allowing us to demonstrate the viability of our method. We will include further comparisons in future work.

It was shown by Wiaux et al. (2009) that ℓ_1 reconstruction with the Dirac basis (i.e. pixel basis) results in very similar reconstruction quality to CLEAN (see Wiaux et al. 2009, fig. 1). This is to be expected since it is known already that CLEAN is closely related to ℓ_1 reconstruction with the Dirac basis (Marsh & Richardson 1987). We thus take a similar approach to that taken by recent studies (McEwen & Wiaux 2011; Carrillo et al. 2012, 2013b; Wolz et al. 2013) and use ℓ_1 reconstruction with a Dirac basis as a proxy for the CLEAN algorithm. The reconstructed image that serves as a CLEAN proxy is therefore given by the solution of the optimization problem:

$$\min \|s\|_1 \text{ such that } \|d - IFAs\|_2 \leq \epsilon, \tag{9}$$

where $\|\cdot\|_1$ denotes the ℓ_1 norm and ϵ is related to a residual noise level estimator (see e.g. Wiaux et al. 2009; Carrillo et al. 2012). We solve this problem using the Sparse Optimisation (SOPT)[2] package (Carrillo et al. 2012, 2013a), using the Douglas–Rachford splitting algorithm (Combettes & Pesquet 2007). Note that the SOPT package is a versatile code, capable of solving much more sophisticated optimization problems than the straightforward ℓ_1 minimization performed here (for examples of more extensive use see Carrillo et al. 2012, 2013b; Wolz et al. 2013).

Figs 6 and 7 are galleries of the input images (without noise), our posterior mean reconstruction using Gibbs sampling and a standard reconstruction using our implementation of CLEAN. For all these images we have zoomed in to the inner 10° where the primary beam selects the most prominent signal. We have applied our variance-based acceptance mask for our Gibbs sampling reconstructions. We see that Gibbs sampling is able to recover the complete range of source images to very high fidelity. We even faithfully recover glitches in the input image, such as the asymmetric protoplanetary disc in the *SWOLF* image and discontinuous lobe structure in the *3c288* image. This emphasizes the power of the Wiener filter: in regions of strong data support, our samples are driven to the data regardless of choice of prior. We also recover low-intensity portions of the image both in the centre of the image and towards the edge of the primary beam, such as the regions between the spiral arms in the *M51HA* image. However, there are other low-intensity regions, such as the lower portion of the moustache in the *Einstein* image, where the signal is too low to distinguish it from the noise, and our mask rejects those pixels.

A comparison to the CLEAN images especially highlights the abilities of our Gibbs sampling reconstruction technique. In all images, we recover a broader range of fluxes over a wider extent than CLEAN (as in e.g. the *cluster* reconstruction). Furthermore, Gibbs sampling is able to recover portions of the image further into the edges of the primary beam, as can be easily seen in the *3c288* and *HCO+4-360* images. Finally, in images with glitches, such as *SWOLF*, CLEAN tends to exaggerate the asymmetries.

Fig. 8 shows the residuals (i.e. difference maps) between the mean posterior signal generated by Gibbs sampling and the input test image. In all cases the variance-based acceptance mask has been applied. We see that the maximum residual occurs in the centre of the *Einstein* test image, where the Gibbs method slightly underestimates the source intensity. This is probably due to the method attempting to average the very high contrast features adjacent to the centre. The *Einstein* image turns out to be the most difficult: the residuals for all other images are typically an order of magnitude smaller. While the residuals are somewhat correlated with the distributions of the source image, like the *Einstein* test image it appears that our method performs poorest in regions of high intensity contrast. However, these differences are very mild: typically on the order of 1 per cent of the input source intensity.

We may further quantify the differences between the input signals and their reconstructions with CLEAN and our Gibbs sampling method by binning the intensities, as we do in Fig. 9. Bounding the histograms of the posterior mean are 2σ error bars measured from the variance in the generated samples. This is another example of the kind of information unavailable in traditional reconstruction techniques. For almost all test images and intensity bins, our posterior mean reconstruction is within two standard deviations of the input signal. This is unsurprising: our Bayesian method automatically discovers the local variance because that variance is directly related to the relative level of signal and noise in a given bin. In a few bins, such as the high-intensity bin of the *Einstein* test image, our method tends to underestimate the true intensity. This is due to the high pixel-by-pixel contrast in the central image region discussed above.

While the CLEAN image intensities largely fall within the uncertainty ranges of the Gibbs reconstruction, there is a systematic steepening of the distributions: CLEAN tends to have too many low-intensity pixels and correspondingly too many high-intensity pixels. This validates the discussion above which noted that CLEAN does not fully reproduce the observed range of input fluxes.

Finally, we may further simplify the comparison by reducing our measurement error to a single scalar. Two error metrics are commonly used: the rms of the residual map, and the SNR. This last quantity itself has several variations; we take that of Carrillo et al. (2012):

$$\text{SNR} \equiv 20 \log_{10} \frac{\sigma_{\text{s}}}{\sigma_{\text{s}-\hat{\text{s}}}}, \tag{10}$$

where σ_{x} is the standard deviation of image x, with s denoting the original image and $\hat{\text{s}}$ denoting the reconstructed image. For both measures we first apply our variance-based acceptance mask before calculating the error metrics.

Fig. 10 shows the rms and SNR for each of our test images for both Gibbs sampling and traditional CLEAN. As expected from our residuals, Gibbs sampling performed most poorly in terms of rms with the *Einstein* test image due to its highly complex structure. With the exception of that image and *g41.1-0.3.b*, all rms errors are below 4×10^{-3}. The SNR for all reconstructions with Gibbs sampling fall between 10 and 25. In both rms and SNR measures Gibbs sampling

[2] http://basp-group.github.io/sopt/

Downloaded from https://academic.oup.com/mnras/article/438/1/768/1041817 by guest on 22 April 2022

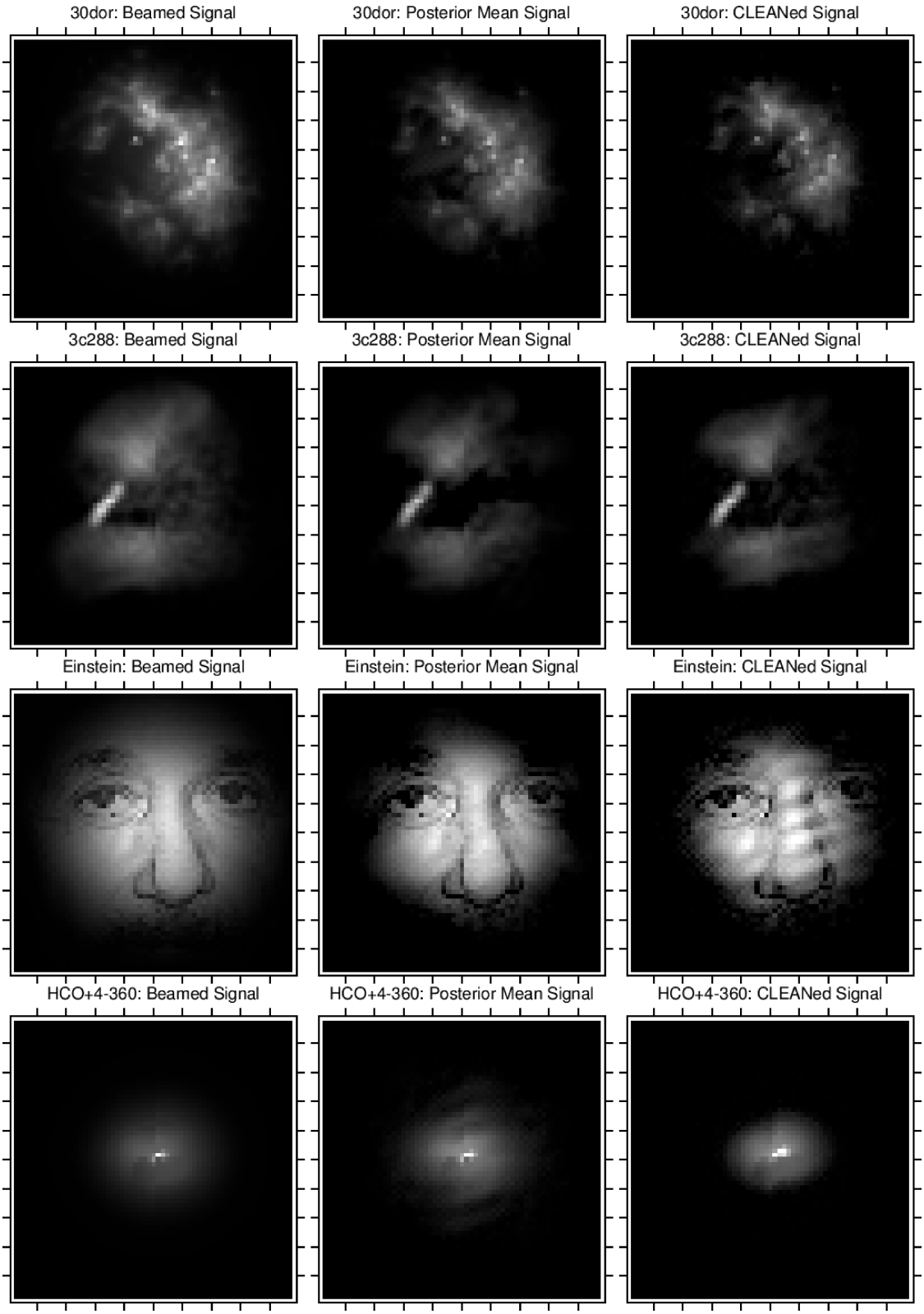


Figure 6. Beamed input images without noise added (*As*; left-hand column), posterior mean signal after 500 Gibbs sampling iterations (middle column) and CLEAN reconstruction (right-hand column). Axes are marked in degree. For clarity we have zoomed in on the innermost 10° where the signal is most prominent. The colour scale ranges from 0.0 (black) to 0.6 (white).

Downloaded from https://academic.oup.com/mnras/article/438/1/768/1041817 by guest on 22 April 2022

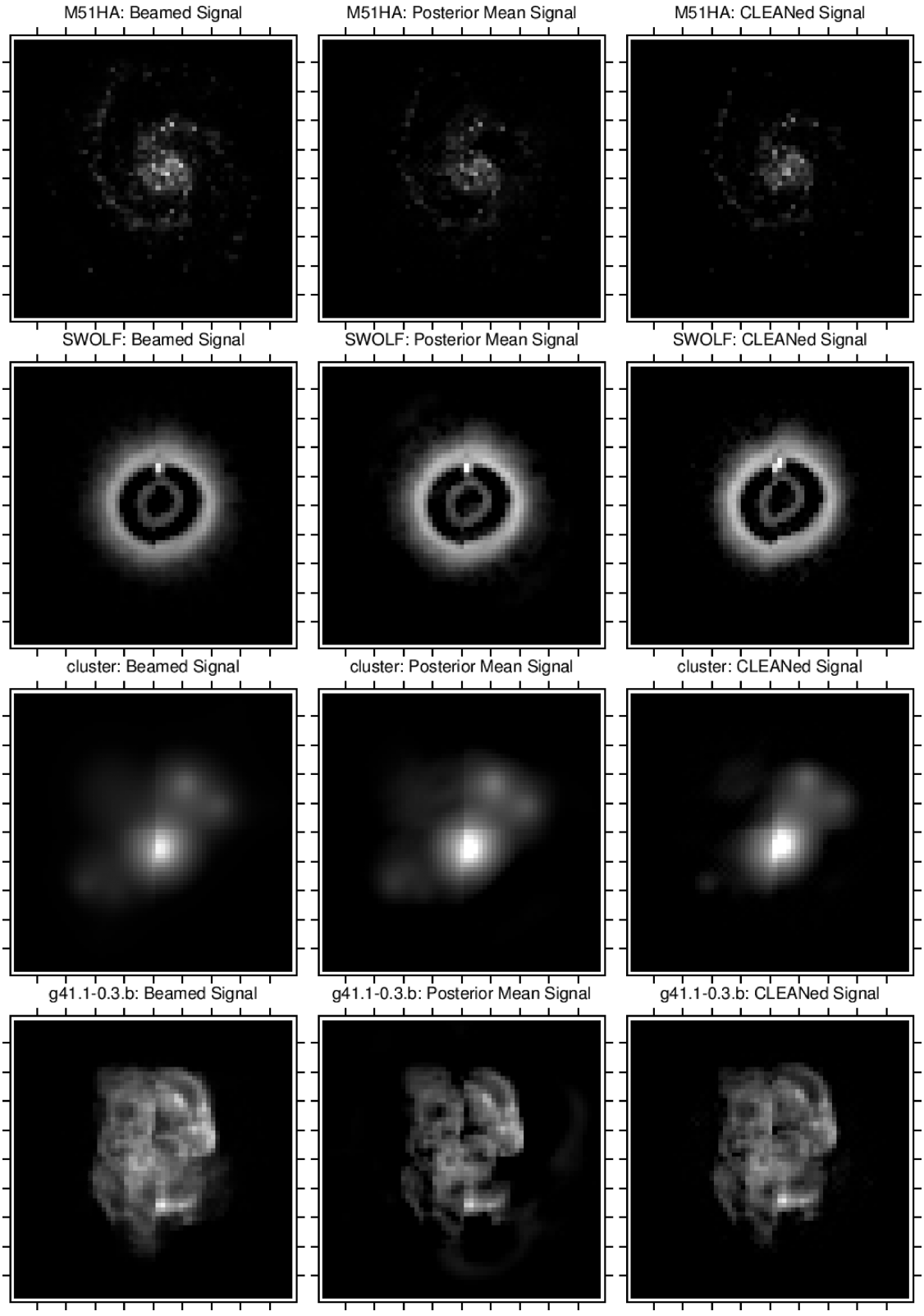


Figure 7. Same as Fig. 6 but for the remaining four test images.

outperforms CLEAN for all test images. The largest differences occur in images with high degrees of symmetry (e.g. *M51HA*, *SWOLF*) since in these cases our isotropic Gaussian process prior performs best in marginal SNR portions of the image.

5 CONCLUSIONS

We have presented an innovative method for deconvolving radio interferometric images using the Bayesian method of Gibbs sampling.

Downloaded from https://academic.oup.com/mnras/article/438/1/768/1041817 by guest on 22 April 2022

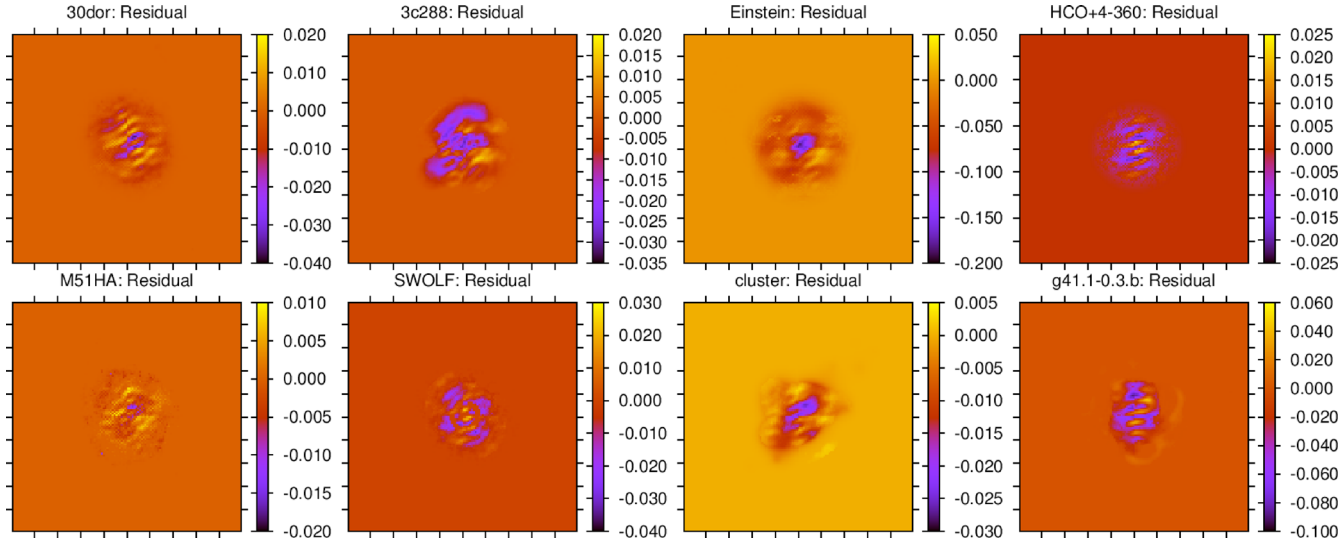


Figure 8. Residuals (differences between reconstructed signal and beamed input signal *As*) for the Gibbs sampling algorithm for all test images. All images are 20° across. The colour scale ranges from 0.0 (black) to 1.0 (white).

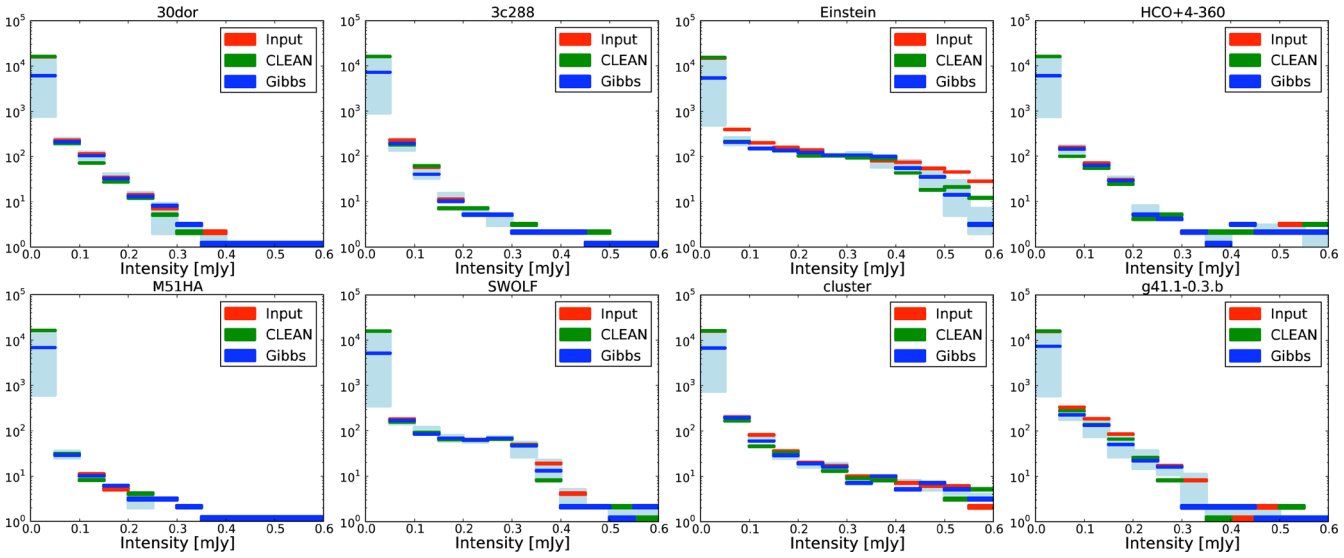


Figure 9. Histograms of sky intensities in the beamed input signal *As* (red), CLEAN reconstructed sky (green) and Gibbs sampling reconstructed sky (blue). Error bars (light blue) on the Gibbs sampling intensities are 2σ uncertainties calculated from the posterior signal samples.

Our method naturally offers several advantages over traditional deconvolution approaches. It fully accounts for signal and noise mode coupling and incomplete *uv*-plane coverage in an automatic and well-motivated fashion, requires no fine-tuning or supervision as the method progresses and provides an informative description of the uncertainty information in the signal reconstruction. We choose an isotropic Gaussian process image prior, though we do not specify the signal covariance in advance.

We have tested our method with a realistic interferometric observing scenario of a wide variety of source images. These images represent typical targets, including protoplanetary discs and AGN jet and lobe systems. Note that our choice of prior is wrong for *all test cases*, but the iterative application of the Wiener filter allows us to discover the source images within the noise and incomplete *uv*-plane coverage. We find that our method is quite robust: regardless of the structure of the source image we are able to recover the intensity distribution to a very high fidelity. Our method outperforms traditional point source-based CLEAN in terms of intensity distributions, rms error and reconstruction SNR. As expected given our isotropic Gaussian process prior, we perform best on images with large amount of symmetry, though the Wiener filter provides a route for reliable reconstructions of asymmetric images in regions of strong data support regardless of source structure.

Using our method we can also easily incorporate uncertainty information in the reconstructed image. As discussed in earlier works (Wandelt et al. 2004), the error model generated from the isotropic Gaussian process prior tends to underestimate the true errors, but this situation is far preferable to no error model at all. With our error model we can construct acceptance masks based on the local (i.e. pixel-by-pixel) signal-to-variance ratios. The desired threshold can be adjusted based on the desired level of confidence in the deconvolution. Since the sample variance is directly tied to the relative level of noise by way of the Wiener filter, our method naturally and self-consistently determines this mask.

Downloaded from https://academic.oup.com/mnras/article/438/1/768/1041817 by guest on 22 April 2022

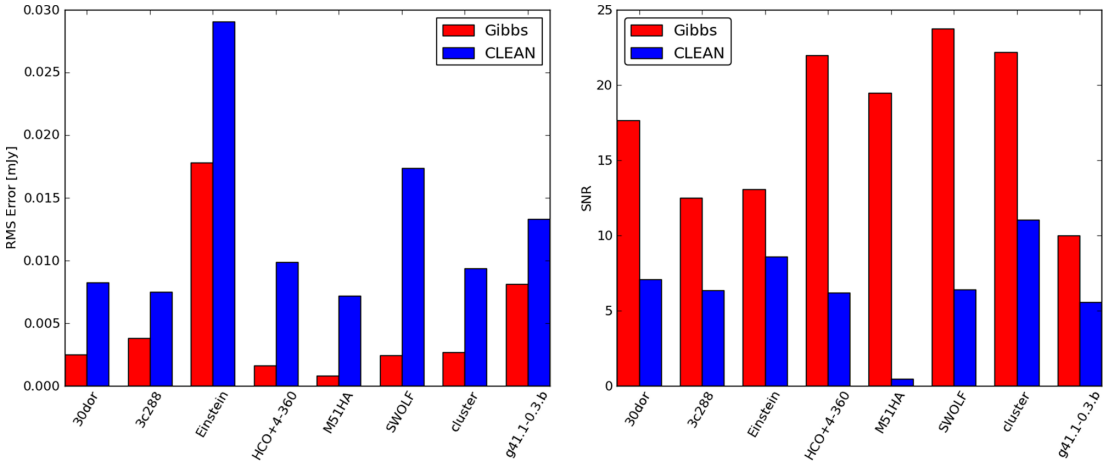


Figure 10. Quantitative assessment of the errors. We show rms error between the beamed input signal As and the reconstructed images in the left-hand plot. In the right-hand plot we show the SNR as defined by equation (10). For each image we compare the Gibbs sampling reconstructed sky (red) to the CLEAN reconstruction (blue). For the rms (SNR) error measurement, smaller (larger) values indicate better performance.

While the test cases we have presented in this work use an interferometric set-up with realistic noise levels, primary beam shape and antenna array, they do represent a relatively simplistic observing scenario. A fully implemented method would include simultaneous solutions of multiple frequencies, mosaicked images and wide bandwidth observations. Also, we have specified an overall SNR of 10 for our fiducial observations, and we must examine the performance of this method in different regimes.

Additionally, future large-scale interferometers will deliver incredibly high volumes of data. Image deconvolution from even a single pointing at a single frequency will tax most computing systems. To accommodate future data sets we have implemented our algorithm using the MPI-parallelized PETSc (Balay et al. 2011) toolkit, so our approach automatically grows with the size of the data without loss of scalability. Algorithmically, the most expensive portion of our approach is the solution to equation (5), which scales as $\mathcal{O}(n_{\rm p} \log n_{\rm p})$ in the ideal pre-gridded flat-sky approximation we have presented here.

While we have not discussed foreground removal in this work, this method easily accommodates modelling in two fashions. First, partial signal or foreground information, if known in advance, can enter as an additional prior. Alternatively, foreground models can be added as an additional sampling step within the algorithm (Wandelt et al. 2004). The resulting posterior mean signal will thus *automatically* include marginalizations over the unconstrained parameters of the model. The same approach can be taken with the noise: if, for example, the noise spectrum is known but the absolute amplitude is not, we can sample over that amplitude. This flexibility, in addition to the other strengths discussed above, suggests that Gibbs sampling is a promising response to the challenges of contemporary and future radio interferometric observations.

ACKNOWLEDGEMENTS

PMS and BDW acknowledge support from NSF Grant AST-0908902. BDW also acknowledges funding from an ANR Chaire d'Excellence, the UPMC Chaire Internationale in Theoretical Cosmology and NSF AST-0708849. JDM is supported in part by a Newton International Fellowship from the Royal Society and the British Academy. LZ and PT acknowledge support from NSF Grant AST-0908900. EFB acknowledges support from NSF Grant AST-0908900. GST and AtK acknowledge support from NSF Grant AST-0908844.

REFERENCES

Ables J. G., 1974, A&AS, 15, 383
Ayasso H., Rodet T., Abergel A., 2012, Inverse Problems, 28, 125005
Balay S., Gropp W. D., McInnes L. C., Smith B. F., 1997, in Arge E., Bruaset A. M., Langtangen H. P., eds, Modern Software Tools in Scientific Computing Efficient Management of Parallelism in Object Oriented Numerical Software Libraries. Birkhäuser Press, Boston, MA, p. 163
Balay S. et al., 2010, PETSc Users Manual, Tech. Rep. ANL-95/11, revision 3.1. Argonne National Laboratory
Balay S. et al., 2011, PETSc, available at: http://www.mcs.anl.gov/petsc
Baron F., Kloppenborg B., Monnier J., 2012, Proc. SPIE, 8445, 84451D
Bhatnagar S., 2009, in Saikia D. J., Green D. A., Gupta Y., Venturi T., eds, ASP Conf. Ser. Vol. 407, The Low-Frequency Radio Universe. Astron. Soc. Pac., San Francisco, p. 375
Bhatnagar S., Cornwell T. J., 2004, A&A, 426, 747
Bhatnagar S., Rau U., Green D. A., Rupen M. P., 2011, ApJ, 739, L20
Carrillo R. E., McEwen J. D., Wiaux Y., 2012, MNRAS, 426, 1223
Carrillo R. E., McEwen J. D., Ville D. V. D., Thiran J.-P., Wiaux Y., 2013a, IEEE Signal Process. Lett., 20, 591
Carrillo R. E., McEwen J. D., Wiaux Y., 2013b, MNRAS, preprint (arXiv:1307.4370)
Cassano R., Brunetti G., Röttgering H. J. A., Brüggen M., 2010, A&A, 509, A68
Combettes P., Pesquet J.-C., 2007, IEEE J. Selected Topics Signal Processing, 1, 564
Cornwell T. J., 2008, IEEE J. Selected Topics Signal Processing, 2, 793
Cornwell T. J., Evans K. F., 1985, A&A, 143, 77
Frigo M., Johnson S., 2005, Proc. IEEE, 93, 216
Gelman A., Rubin D., 1992, Stat. Sci., 7, 457
Gelman A., Carlin J., Stern H., Rubin D., 2004, Bayesian Data Analysis. Chapman & Hall/CRC, London
Giovannelli J.-F., Coulais A., 2005, A&A, 439, 401
Gratier P. et al., 2010, A&A, 522, A3
Gull S. F., Daniell G. J., 1978, Nature, 272, 686
Heesen V., Rau U., Rupen M. P., Brinks E., Hunter D. A., 2011, ApJ, 739, L23
Högbom J. A., 1974, A&AS, 15, 417
Jaeger S., 2008, in Argyle R. W., Bunclark P. S., Lewis J. R., eds, ASP Conf. Ser. Vol. 394, Astronomical Data Analysis Software and Systems XVII. Astron. Soc. Pac., San Francisco, p. 623

Downloaded from https://academic.oup.com/mnras/article/438/1/768/1041817 by guest on 22 April 2022

Jarvis M. J., 2007, in Afonso J., Ferguson H. C., Mobasher B., Norris R., eds, ASP Conf. Ser. Vol. 380, At the Edge of the Universe: Latest Results from the Deepest Astronomical Surveys. Astron. Soc. Pac., San Francisco, p. 251
Jewell J., Levin S., Anderson C. H., 2004, ApJ, 609, 1
Johnston S. et al., 2008, Exp. Astron., 22, 151
Li F., Cornwell T. J., de Hoog F., 2011, A&A, 528, A31
Lonsdale C. J. et al., 2009, Proc. IEEE, 97, 1497
McClure-Griffiths N. M., Dickey J. M., Gaensler B. M., Green A. J., Green J. A., Haverkorn M., 2012, ApJS, 199, 12
McEwen J. D., Wiaux Y., 2011, MNRAS, 413, 1318
Mackay D. J. C., 2003, Information Theory, Inference and Learning Algorithms. Cambridge Univ. Press, Cambridge
Marsh K. A., Richardson J. M., 1987, A&A, 182, 174
Myers S. T. et al., 2003, ApJ, 591, 575
Norris R. P. et al., 2013, Publ. Astron. Soc. Aust., 30, 20
Nyman L.-Å., Andreani P., Hibbard J., Okumura S. K., 2010, Proc. SPIE, 7737, 77370G
Parsons A. R. et al., 2010, AJ, 139, 1468
Pearson T. J. et al., 2005, in Lasenby A., Wilkinson A., eds, Proc. IAU Symp. 201, New Cosmological Data and the Values of the Fundamental Parameters. Astron. Soc. Pac., San Francisco, p. 23
Perley R. A., Chandler C. J., Butler B. J., Wrobel J. M., 2011, ApJ, 739, L1
Press W. H., Flannery B. P., Teukolsky S. A., 1986, Numerical Recipes. The Art of Scientific Computing. Cambridge Univ. Press, Cambridge
Puetter R. C., Gosnell T. R., Yahil A., 2005, ARA&A, 43, 139
Rau U., 2012, Proc. SPIE, 8500, 85000N
Rau U., Cornwell T. J., 2011, A&A, 532, A71
Starck J.-L., Murtagh F., Querre P., Bonnarel F., 2001, A&A, 368, 730
Stewart I. M., Fenech D. M., Muxlow T. W. B., 2011, A&A, 535, A81
Suksmono A., 2009, ICEEI'09 Vol. 1, Deconvolution of VLBI images based on compressive sensing. Electr. Eng. Inform., Bandung, Indonesia, p. 110
Sutter P. M., Wandelt B. D., Malu S. S., 2012, ApJS, 202, 9
Sutton E. C., Wandelt B. D., 2006, ApJS, 162, 401
Thompson A. R., Moran J. M., Swenson G. W., Jr, 2001, Interferometry and Synthesis in Radio Astronomy. Wiley, New York
Wandelt B., Larson D., Lakshminarayanan A., 2004, Phys. Rev. D, 70, 12
Wenger S., Magnor M., Pihlström Y., Bhatnagar S., Rau U., 2010, PASP, 122, 1367
Wiaux Y., Jacques L., Puy G., Scaife A. M. M., Vandergheynst P., 2009, MNRAS, 395, 1733
Wiaux Y., Puy G., Vandergheynst P., 2010, MNRAS, 402, 2626
Wolz L., McEwen J. D., Abdallah F. B., Carrillo R. E., Wiaux Y., 2013, MNRAS, 436, 1993
Zhang H.-X., Hunter D. A., Elmegreen B. G., 2012, ApJ, 754, 29

This paper has been typeset from a TEX/LATEX file prepared by the author.

Downloaded from https://academic.oup.com/mnras/article/438/1/768/1041817 by guest on 22 April 2022