Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **452,** 1587–1600 (2015) doi:10.1093/mnras/stv1355

A Bayesian analysis of redshifted 21-cm H I signal and foregrounds: simulations for LOFAR

Abhik Ghosh,[1]★ Léon V. E. Koopmans,[1] E. Chapman[2] and V. Jelić[1,3,4]

[1]*Kapteyn Astronomical Institute, University of Groningen, PO Box 800, NL-9700 AV Groningen, the Netherlands*
[2]*Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT*
[3]*ASTRON, PO Box 2, NL-7990 AA Dwingeloo, the Netherlands*
[4]*Ruđer Bošković Institute, Bijenička cesta 54, 10000 Zagreb, Croatia*

Accepted 2015 June 16. Received 2015 June 11; in original form 2015 April 15

ABSTRACT

Observations of the epoch of reionization (EoR) using the 21-cm hyperfine emission of neutral hydrogen (H I) promise to open an entirely new window on the formation of the first stars, galaxies and accreting black holes. In order to characterize the weak 21-cm signal, we need to develop imaging techniques that can reconstruct the extended emission very precisely. Here, we present an inversion technique for LOw Frequency ARray (LOFAR) baselines at the North Celestial Pole (NCP), based on a Bayesian formalism with optimal spatial regularization, which is used to reconstruct the diffuse foreground map directly from the simulated visibility data. We notice that the spatial regularization de-noises the images to a large extent, allowing one to recover the 21-cm power spectrum over a considerable $k_\perp$–$k_\parallel$ space in the range $0.03\,\mathrm{Mpc}^{-1} < k_\perp < 0.19\,\mathrm{Mpc}^{-1}$ and $0.14\,\mathrm{Mpc}^{-1} < k_\parallel < 0.35\,\mathrm{Mpc}^{-1}$ without subtracting the noise power spectrum. We find that, in combination with using generalized morphological component analysis (GMCA), a non-parametric foreground removal technique, we can mostly recover the spherical average power spectrum within 2σ statistical fluctuations for an input Gaussian random root-mean-square noise level of 60 mK in the maps after 600 h of integration over a 10-MHz bandwidth.

Key words: methods: data analysis – techniques: interferometric – cosmology: general – diffuse radiation – radio continuum: general.

1 INTRODUCTION

Observations of redshifted 21-cm radiation from the large-scale distribution of neutral hydrogen (H I) provide one of the most promising probes with which to study the high-redshift Universe. The epoch of reionization (EoR) marks an important milestone in the large-scale history of the Universe. The timing, duration and character of subsequent events that led to reionization of the Universe contain an enormous amount of information about the first galaxies and stars (Furlanetto, Oh & Briggs 2006; Morales & Wyithe 2010; Mellema et al. 2013). Evidence from quasar absorption spectra (Becker et al. 2001; Fan et al. 2002) and cosmic microwave background radiation (CMBR: Page et al. 2007; Spergel et al. 2007; Larson et al. 2011) implies that neutral hydrogen was reionized over a redshift range $6 \leq z \leq 15$, but the exact details of how the Universe was slowly reionized, the first sources of radiation in the Universe and the way in which the EoR influenced structure formation (Barkana & Loeb 2001; Loeb & Barkana 2001; Pritchard & Loeb 2012; Robertson et al. 2013) are presently unknown. Through observations of the redshifted 21-cm line as a function of redshift and angular position, we aim to measure the H I distribution in the Universe directly, which will eventually shed light on large-scale structure formation during this epoch. Currently, several low-frequency instruments such as the LOw Frequency ARray (LOFAR),[1] Murchison Widefield Array (MWA),[2] Precision Array for Probing the Epoch of Reionization (PAPER)[3] and Giant Metrewave Radio Telescope (GMRT)[4] are carrying out observations specifically to detect the redshifted 21-cm H I signal. There are also ongoing efforts to build up interferometers such as MIT Epoch of Reionization (MITEoR; Zheng et al. 2014) and the Hydrogen Epoch of Reionization Array (HERA)[5], which use massive baseline redundancy to enable automated precision calibration,

★ E-mail: ghosh@astro.rug.nl

[1] http://www.lofar.org/
[2] http://www.mwatelescope.org/
[3] http://astro.berkeley.edu/∼dbacker/eor/
[4] http://www.gmrt.ncra.tifr.res.in
[5] http://reionization.org/

reduce system noise and cut down on the computational cost of the correlator.

Unfortunately, the cosmological 21-cm signal is very faint, so that no current instrument can detect it directly. It is perceived that it is possible to observe the high-redshift 21-cm signal through a statistical analysis of the fluctuations in 21-cm radiation, though currently only upper limits have been placed (Paciga et al. 2013; Dillon et al. 2014a; Ali et al. 2015). Regarding the foregrounds, which are at least four orders of magnitude larger than the 21-cm signal, it is expected that individual point sources can be identified in the images and can subsequently be removed to a flux level depending on the sensitivity of the instrument. The contribution from synchrotron emission (Shaver et al. 1999) from our Galaxy is another major diffuse foreground component for detecting the redshifted 21-cm H I signal. Other foreground sources include free–free emission from ionizing haloes (Oh & Mack 2003) and unresolved extragalactic sources, which could also be large enough compared with the H I signal. Confusion noise generated from the sea of point sources below the detection limit can also be a major obstacle in detecting the H I signal (Di Matteo et al. 2002). For a dipole array such as PAPER, which has a wide field of view, Parsons & Backer (2009) have developed a delay rate filtering technique that is used to remove smooth foregrounds and a beam-sculpting method depending on the fringe rate of the sources (Parsons et al. 2015) that can be used to select a restricted area near the phase centre to minimize the side-lobe contribution coming from sea of point sources away from the field centre.

Radio surveys at 408 MHz (Haslam et al. 1982), 1.42 GHz (Reich 1982; Reich & Reich 1988) and 2.326 GHz (Jonas, Baart & Nicolson 1998) have measured the diffuse Galactic synchrotron radiation at angular scales larger than $\sim 1^{\circ}$. At relatively high frequencies, Giardino et al. (2001) and Giardino et al. (2002) have analysed the fluctuations in Galactic synchrotron radiation using the 2.3-GHz Rhodes Survey and the 2.4-GHz Parkes radio continuum and polarization survey. The structure of Galactic synchrotron emission is not well quantified at the frequencies and angular scales relevant for detecting the cosmological 21-cm signal. However, the Global Sky Model (GSM) of de Oliveira-Costa et al. (2008), which is compiled from publicly available total-power large-area radio surveys at frequencies 10, 22, 45 and 408 MHz and 1.42, 2.326, 23, 33, 41, 61 and 94 GHz, gives us a representative idea of diffuse Galactic radio emission at a common angular resolution of 5° and with an accuracy of 1–10 per cent. Recently, Bernardi et al. (2009) and Ghosh et al. (2012) have analysed 150-MHz Westerbork Synthesis Radio Telescope (WSRT) and GMRT observations, respectively, where they find that Galactic synchrotron emission is the most dominant foreground at angular scales >10 arcmin after point source subtraction at the 10–20 mJy level. Precise characterization and detailed understanding of the Galactic synchrotron emission are needed to remove foregrounds in 21-cm experiments faithfully. The study of Galactic synchrotron emission is itself interesting. It will eventually shed light on the cosmic-ray electron distribution, the strength and structure of the Galactic magnetic field and magnetic turbulence (Waelkens, Schekochihin & Enßlin 2009; Lazarian & Pogosyan 2012; Iacobelli et al. 2013). It is also interesting to point out that Galactic polarized emission can mimic the EoR signal in the frequency direction and a proper calibration of the instrument is essential so that leakages of polarized to total signal are minimal (Jelić et al. 2010; Geil, Gaensler & Wyithe 2011; Asad et al. 2015).

However, interferometers inherently give us a dirty map, which is the convolution of the true map and direction-dependent point-spread functions (PSFs) or synthesized beams corresponding to the sampling function of the telescope. In general, PSFs are direction-dependent and typically not invertible (Dillon et al. 2014b). Therefore, we need to develop techniques to characterize (and subsequently remove) the diffuse foregrounds accurately for every image pixel in the map. The ultimate aim is to develop calibration and imaging software that will take into account all direction-dependent effects such as those due to the ionosphere (Vedantham & Koopmans 2014) or the primary beam and direction-independent (such as instrument gain) calibration errors (Yatawatta et al. 2008). We note that, in the current generation of imaging software for LOFAR, the calibration solution in the radio interferometric measurement equation that describes the instrument and the ionosphere is only applied to the model components (mainly compact sources) in specific directions. In general, only a single directionally independent station-based solution is applied to the residual image. Hence, all images produced by the available imaging software will include some residual calibration errors, which vary from pixel to pixel and can pose a potential obstacle for EoR detection and its characterization. Also, the polarized foregrounds leaked from Stokes Q, U and I maps can mimic the EoR signal if not properly calibrated. However, recent results from Asad et al. (2015) show that in the case of LOFAR a modest polarimetric calibration is sufficient to ensure that the polarization leakage remains below the EoR signal inside the full width at half-maximum (FWHM) of the primary beam (PB) of LOFAR. The polarization leakage will also influence the diffuse foreground characterization and modelling for Stokes I images. In this article, we examine a maximum-likelihood (ML) inversion methodology based on a Bayesian formalism to infer the deconvolved map directly from the simulated visibilities for LOFAR. We note that the technique is quite general and can easily be exported for other interferometers to deconvolve the effects of the instrumental point-spread function effectively.

An important point to note here is that the fundamental concept of the proposed method is based on a Bayesian inference network, where we can incorporate two levels of inference regarding the estimation of the foreground model. Initially, we fit a foreground model to the data to determine the free model parameters and in the second level of inference we can also rank these models. This analysis technique automatically incorporates 'Occam's razor', which ensures that overly complex foreground models will not be preferred over simpler models unless the data support them (MacKay 1992). Regarding the choice of prior in the Bayesian framework, the proposed technique does not depend strongly on the inclusion of the right priors that will influence the final outcome; rather, many different priors can be tried (such as identity, gradient or curvature regularization functions) and finally the data will inform us which is the most appropriate prior based on the calculated evidence values. We can also maximize the evidence in light of the data for finding the regularization constant, which will set the strength of the prior in the penalty function.

Based on our simulated ML solutions, we also study how good we are at recovering the input EoR power spectrum if we only have a diffuse foreground model with unresolved extragalactic sources, EoR signal and Gaussian random noise. We use a non-parametric foreground removal technique (generalized morphological component analysis, hereafter GMCA) to remove the foregrounds. GMCA uses a sparsity-based blind source separation (BSS) technique (Bobin et al. 2008) to describe the different foreground components, which are finally combined in different ratios according to the frequency of the maps to model the foregrounds (Chapman et al. 2013).

The article is organized as follows. In Section 2, we describe our approach in mathematical formalism. Section 3 discusses the

templates used for generating the visibility data corresponding to the EoR signal and diffuse foreground emission. In Section 4 we present our inversion results, while in Sections 5 and 6 we describe our foreground removal results and power-spectrum comparison. Finally, in Section 7 we present a summary and possible future extension of the current work.

2 BAYESIAN INFERENCE: FORMALISM

The fundamental concept of our proposed method is based on a Bayesian framework (MacKay 1992), which automatically ensures that overly complex models will not be preferred over simpler models unless the data support them. The Bayesian framework adopted in this work automatically incorporates 'Occam's razor'. In this section, we give a general overview of the mathematical formalism used to develop the ML technique. Let us consider a linear system,

$$\boldsymbol{V} = \mathbf{F}\boldsymbol{s} + \boldsymbol{n}, \tag{1}$$

where $\boldsymbol{V}$ is a vector of data points, $\mathbf{F}$ represents the response function, $\boldsymbol{s}$ are the source flux parameters that we want to infer given the data and $\boldsymbol{n}$ are the noise in the data characterized by the covariance matrix $\mathbf{C}_\mathrm{D}$ (see e.g. Suyu et al. 2006). In our case, $\boldsymbol{V}$ are the measured visibilities $\mathcal{V}_\nu(u, v, w)$, $\mathbf{F}$ is the Fourier transform kernel $\mathrm{e}^{-2\pi\mathrm{i}[ul+vm+w(\sqrt{1-l^2-m^2}-1)]}$ and $\boldsymbol{s}$ are the intensities $I_\nu(l, m)/[\sqrt{1-l^2-m^2}] \times \mathrm{d}\Omega_\mathrm{p}$ in the sky that we want to infer ($\mathrm{d}\Omega_\mathrm{p}$ is the pixel area in solid angle; for details we refer the reader to Section 3.4). Assuming the noise to be Gaussian, the probability of the data given the model parameters $\boldsymbol{s}$ is

$$P(\boldsymbol{V}|\boldsymbol{s}) = \frac{\exp(-E_\mathrm{D}(\boldsymbol{V}|\boldsymbol{s}))}{Z_\mathrm{D}}, \tag{2}$$

where

$$\begin{aligned} E_\mathrm{D}(\boldsymbol{V}|\boldsymbol{s}) &= \frac{1}{2}\left(\mathbf{F}\boldsymbol{s} - \boldsymbol{V}\right)^\mathrm{H} \mathbf{C}_\mathrm{D}^{-1}\left(\mathbf{F}\boldsymbol{s} - \boldsymbol{V}\right) \\ &\equiv \frac{1}{2}\chi^2 \end{aligned} \tag{3}$$

and $Z_\mathrm{D} = (2\pi)^{N_\mathrm{d}/2}(\det \mathbf{C}_\mathrm{D})^{1/2}$ is the normalization factor.

By definition, the most likely solution ($\boldsymbol{s}_\mathrm{ML}$) maximizes the likelihood or minimizes E_D. By setting $\nabla E_\mathrm{D}(\boldsymbol{s}_\mathrm{ML}) = \mathbf{0}$, where $\nabla \equiv \partial/\partial\boldsymbol{s}$, we find the ML solution to be

$$\boldsymbol{s}_\mathrm{ML} = \left(\mathbf{F}^\mathrm{H}\mathbf{C}_\mathrm{D}^{-1}\mathbf{F}\right)^{-1}\mathbf{F}^\mathrm{H}\mathbf{C}_\mathrm{D}^{-1}\boldsymbol{V}. \tag{4}$$

In many cases, the problem of finding the most likely solution that minimizes half of χ^2 (here, E_D) is ill-posed and we need to introduce priors to regularize the solution of $\boldsymbol{s}$.

The posterior probability of the source flux pixels $\boldsymbol{s}$, given the visibility data $\boldsymbol{V}$, a fixed form of regularization function $\mathbf{R}$ and a level λ of regularization follows from Bayes' theorem:

$$P(\boldsymbol{s}|\boldsymbol{V}, \lambda, \mathbf{R}) = \frac{P(\boldsymbol{V}|\boldsymbol{s})P(\boldsymbol{s}|\mathbf{R}, \lambda)}{P(\boldsymbol{V}|\lambda, \mathbf{R})}, \tag{5}$$

where $P(\boldsymbol{V}|\lambda, \mathbf{R})$ is called the *evidence*, which depends on the model parameters $\{\lambda, \mathbf{R}\}$. We choose the model that maximizes the evidence. In the first level of inference, the most probable (MP) solution maximizes the posterior, which can be written as

$$P(\boldsymbol{s}|\boldsymbol{V}, \lambda, \mathbf{R}) = \frac{\exp(-M(\boldsymbol{s}))}{Z_\mathrm{M}(\lambda)}, \tag{6}$$

where $M(\boldsymbol{s}) = E_\mathrm{D}(\boldsymbol{s}) + \lambda E_\mathrm{S}(\boldsymbol{s})$, λ is the regularization constant, E_S is called the regularizing function and $Z_\mathrm{M}(\lambda) = \int \mathrm{d}^{N_\mathrm{s}}\boldsymbol{s}\exp(-M(\boldsymbol{s}))$ is the normalizing evidence function.

In our analysis we choose the three most common quadratic forms of regularization functions (e.g. zero-order, gradient and curvature). We subsequently maximize the evidence to fix the optimal regularization function and regularization parameter. This allows for a fair comparison between any form of regularization. Although we do not claim to have found the 'best' form, we note that, for reconstructing smooth diffuse foregrounds or the 21-cm signal, higher order regularization functions (i.e. gradient or curvature) in general work better compared with the zeroth-order regularization (Suyu et al. 2006; see also Koopmans 2005, where these function were also used for lensing purposes). Any other form that can be conceived can in that case be compared objectively with the ones we chose to test, which are the simplest forms possible. In index and summation notation, the zero-order regularization can be expressed as

$$E_\mathrm{S}(\boldsymbol{s}) = \frac{1}{2}\sum_{i=1}^{N_\mathrm{s}} s_i^2, \tag{7}$$

which tries to minimize the flux at every source pixel.

For gradient and curvature forms of regularization, let s_{i_1,i_2} be the source flux at pixel (i_1, i_2), where i_1 and i_2 lie in range $i_1 = 1, \ldots, N_{1\mathrm{s}}$ and $i_2 = 1, \ldots, N_{2\mathrm{s}}$ in the two dimensions, respectively. $N_\mathrm{s} = N_{1\mathrm{s}}N_{2\mathrm{s}}$ is the total number of flux pixels. We use the following form of gradient regularization:

$$\begin{aligned} E_\mathrm{S}(\boldsymbol{s}) = &\frac{1}{2}\sum_{i_1=1}^{N_{1\mathrm{s}}-1}\sum_{i_2=1}^{N_{2\mathrm{s}}}\left[s_{i_1,i_2} - s_{i_1+1,i_2}\right]^2 \\ &+ \frac{1}{2}\sum_{i_1=1}^{N_{1\mathrm{s}}}\sum_{i_2=1}^{N_{2\mathrm{s}}-1}\left[s_{i_1,i_2} - s_{i_1,i_2+1}\right]^2, \end{aligned} \tag{8}$$

which tries to minimize the difference in the flux values between adjacent pixels in both dimensions.

Similarly, the curvature regularization function that we use is

$$\begin{aligned} E_\mathrm{S}(\boldsymbol{s}) = &\frac{1}{2}\sum_{i_1=1}^{N_{1\mathrm{s}}-2}\sum_{i_2=1}^{N_{2\mathrm{s}}}\left[s_{i_1,i_2} - 2s_{i_1+1,i_2} + s_{i_1+2,i_2}\right]^2 \\ &+ \frac{1}{2}\sum_{i_1=1}^{N_{1\mathrm{s}}}\sum_{i_2=1}^{N_{2\mathrm{s}}-2}\left[s_{i_1,i_2} - 2s_{i_1,i_2+1} + s_{i_1,i_2+2}\right]^2, \end{aligned} \tag{9}$$

which minimizes the sum of the curvature in both dimensions.

The most probable solution can be derived from the most likely solution. For mathematical convenience, we introduce some parameters $\mathbf{B}$ and $\mathbf{C}$, which are the Hessian of E_D and E_S respectively ($\mathbf{B} = \nabla\nabla E_\mathrm{D}(\boldsymbol{s})$ and $\mathbf{C} = \nabla\nabla E_\mathrm{S}(\boldsymbol{s})$). We define the Hessian of M as $\mathbf{A} = \nabla\nabla M(\boldsymbol{s})$, which is equivalent to $\mathbf{A} = \mathbf{B} + \lambda\mathbf{C}$. Following Suyu et al. (2006), $\nabla E_\mathrm{D}(\boldsymbol{s}_\mathrm{ML}) = \mathbf{0}$ gives $\boldsymbol{s}_\mathrm{ML} = \mathbf{B}^{-1}\mathbf{F}^\mathrm{H}\mathbf{C}_\mathrm{D}^{-1}\boldsymbol{V}$ and the most probable (MP) solution can be estimated by $\nabla M(\boldsymbol{s}_\mathrm{MP}) = \mathbf{0}$, which can be written as $\boldsymbol{s}_\mathrm{MP} = \mathbf{A}^{-1}\mathbf{B}\boldsymbol{s}_\mathrm{ML}$. We note that the MP solution depends on the regularization constant λ, since the Hessian $\mathbf{A}$ depends on λ.

To find the value of the optimal regularization parameter, we maximize

$$P(\lambda|\boldsymbol{V}, \mathbf{R}) = \frac{P(\boldsymbol{V}|\lambda, \mathbf{R})P(\lambda)}{P(\boldsymbol{V}|\mathbf{R})}. \tag{10}$$

Assuming a flat prior in $\log\lambda$, we maximize the evidence in order to optimize λ.

For any quadratic form of the regularization function $E_\mathrm{S}(\boldsymbol{s})$, we can write the evidence as

$$P(\boldsymbol{V}|\lambda, \mathbf{R}) = \frac{Z_\mathrm{M}(\lambda)}{Z_\mathrm{D}Z_\mathrm{S}(\lambda)}, \tag{11}$$

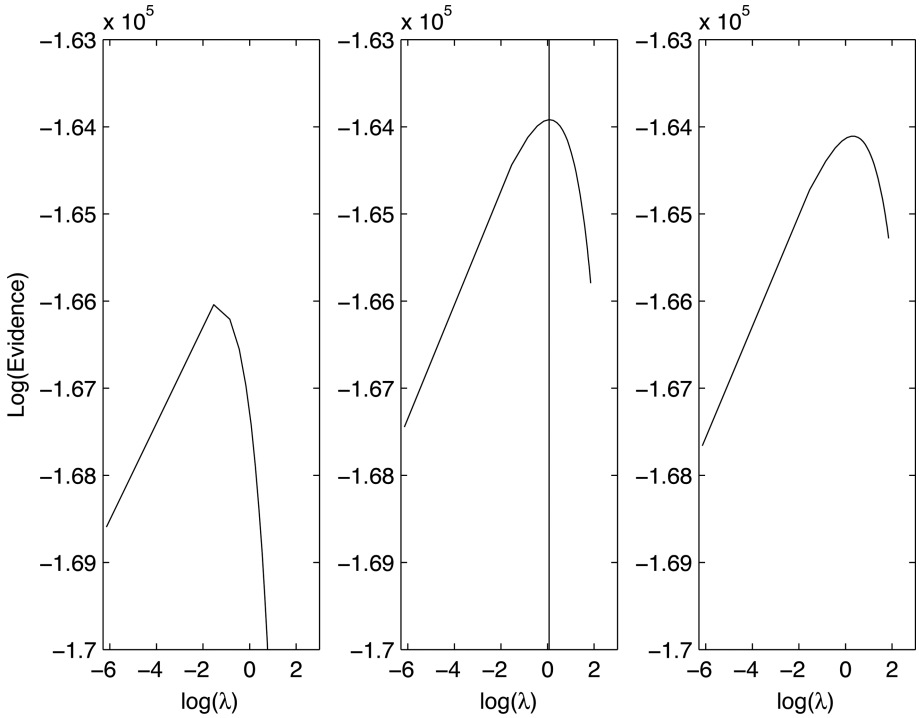

Figure 1. From left to right, this figure shows the log evidence as a function of the log of the regularization parameter λ for identity, gradient and curvature regularization function, respectively. The log evidence is maximum for gradient regularization and the vertical line in the second panel displays the peak of log evidence, which is around $\log(\lambda) = 0.0582$ or equivalently $\lambda = 1.06$.

where

$$Z_{\rm S}(\lambda) = \left(\frac{2\pi}{\lambda}\right)^{N_{\rm s}/2} (\det \mathbf{C})^{-1/2} {\rm e}^{-\lambda E_{\rm S}(\mathbf{0})}, \tag{12}$$

$$Z_{\rm M}(\lambda) = (2\pi)^{N_{\rm s}/2} (\det \mathbf{A})^{-1/2} {\rm e}^{-M(\boldsymbol{s}_{\rm MP})}. \tag{13}$$

We optimize the log of evidence to find the optimal value of regularization constant λ. Using equations (12) and (13), the log of evidence can be expressed as

$$\begin{aligned} \log P(\boldsymbol{V}|\lambda, \mathbf{R}) = & -\lambda E_{\rm S}(\boldsymbol{s}_{\rm MP}) - E_{\rm D}(\boldsymbol{s}_{\rm MP}) \\ & - \frac{1}{2}\log(\det \mathbf{A}) + \frac{N_{\rm s}}{2}\log\lambda + \lambda E_{\rm S}(\mathbf{0}) \\ & + \frac{1}{2}\log(\det \mathbf{C}) - \frac{N_{\rm d}}{2}\log(2\pi) \\ & + \frac{1}{2}\log(\det \mathbf{C}_{\rm D}^{-1}). \end{aligned} \tag{14}$$

Here, $N_{\rm d}$ is the number of visibility samples.

Assuming a flat prior in $\log\lambda$, we maximize the evidence $P(\boldsymbol{V}|\lambda, \mathbf{R})$ to find the optimal regularization constant $\hat{\lambda}$. Solving

$$\frac{\rm d}{{\rm d}\log\lambda} \log P(\boldsymbol{V}|\lambda, \mathbf{R}) = 0$$

(Suyu et al. 2006), we obtain the following non-linear equation for $\hat{\lambda}$:

$$2\hat{\lambda} E_{\rm S}(\boldsymbol{s}_{\rm MP}) = N_{\rm s} - \hat{\lambda}{\rm Tr}(\mathbf{A}^{-1}\mathbf{C}), \tag{15}$$

where, $N_{\rm s}$ are the number of sky flux pixels. We note that solving this non-linear equation is computationally intensive for inversions on a large grid and we find that $\hat{\lambda} = 1/\sigma^2$, where σ^2 is the noise variance in the data, is a good choice for setting the value of the regularization constant. Following Suyu et al. (2006), we expect M evaluated at the optimal λ value to be equal to half the number of data points ($M(\boldsymbol{s}_{\rm MP}) \sim N_{\rm d}/2$). We find that $\hat{\lambda} = 1/\sigma^2$ follows this quite well and the difference is close to the 1 per cent level. We emphasize here that in our analysis we solve equation (15) explicitly to find the optimal λ value with different regularization functions, namely identity, gradient or curvature regularization. We also note that, regarding the choice of the functional form of prior in this framework, our technique does not depend heavily on the inclusion of very precise priors (indeed a genuine worry in many Bayesian methods, but one that is mitigated in the framework laid out by MacKay 1992, which we use here), which will influence the final outcome. Rather, different prior families can be compared quantitatively (e.g. identity, gradient or curvature regularization functions) and the data inform us which is the most appropriate prior (both shape and strength) based on the calculated evidence (i.e. the marginalized posterior). For ranking different models, the regularization factor λ is a nuisance parameter, which ends up being marginalized. In general, the distribution of λ is sharply peaked and we can approximate $P(\lambda|\boldsymbol{V}, \mathbf{R})$ by a delta function peaked at its most probable value $\hat{\lambda}$. Hence, the evidence can be approximated by $P(\boldsymbol{V}|\hat{\lambda}, \mathbf{R})$. In our case, using the non-gridded visibilities, we find that the gradient regularization works best with peak values of log evidence at the optimal λ value of −166043, −163919, −164109, respectively, for identity, gradient and curvature regularization functions (Fig. 1). Based on the Bayes factors, there is very strong evidence towards choosing the gradient regularization over the identity or curvature forms of regularization (Jeffreys 1961; Kass & Raftery 2011). For the rest of our analysis, we therefore choose the gradient regularization function.

In Fig. 2, we show the ML maps corresponding to different regularization functions. Even though the ML maps look similar, based on the evidence values, we decided to use curvature regularization for the rest of our analysis. We note that Occam's razor is implicit in our evidence optimization. If the model parameter space is overly

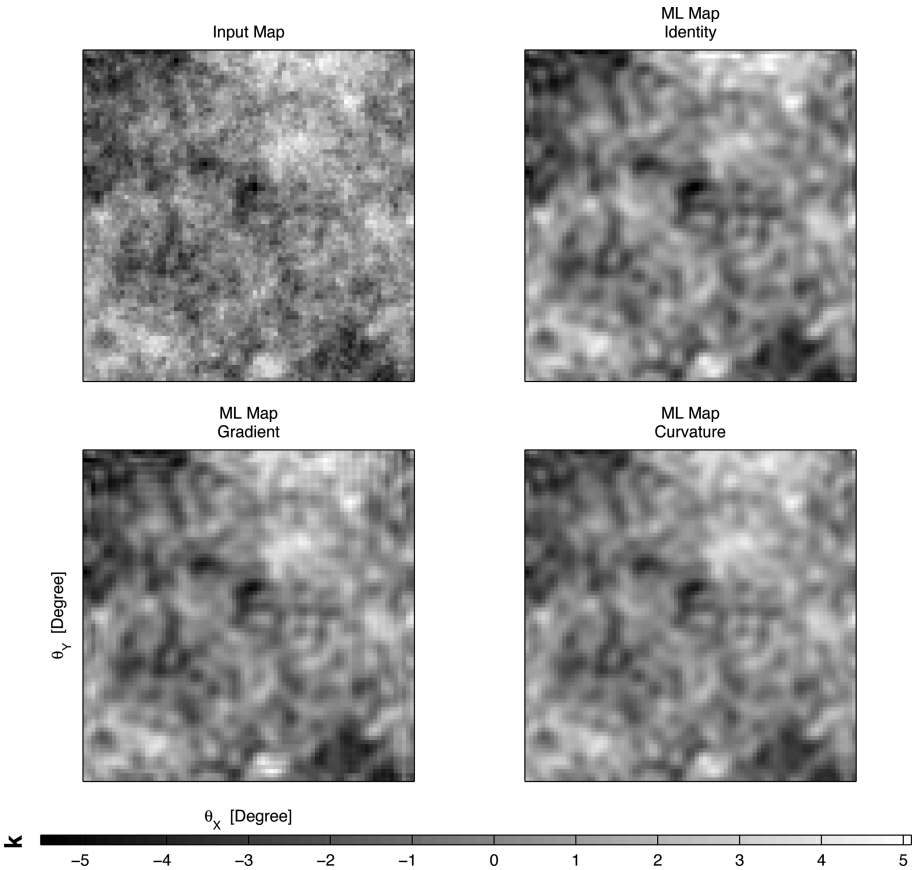

Figure 2. The top left panel shows the input map without any noise. The remaining panels show the ML inverted maps corresponding to identity, gradient and curvature regularization functions, respectively. Each panel covers a region of $2^\circ \times 2^\circ$.

large (for small values of λ), Occam's razor penalizes such an overly flexible model. At the other extreme, for overly large values of λ the model can no longer fit the data, as the model parameter space is restricted to a limited region. The optimal λ value lies somewhere in between these two extremes, which ensures that overly complex models will not be preferred over simpler models unless the data support them. Although it is in principle possible to infer σ as well, in this article we assume that we can infer it independently.

3 SIMULATED DATA TEMPLATES

In this section we describe briefly how the data are generated for the ML analysis. We consider templates for the EoR signal and diffuse foreground emission from which the visibility samples were generated and then we add Gaussian random noise to the real and imaginary part of these visibilities.

3.1 EoR signal

We have used the semi-analytic code 21CMFAST (Mesinger & Furlanetto 2007; Mesinger, Furlanetto & Cen 2011) to simulate the EoR signal. Being a semi-analytical code, 21CMFAST treats physical processes with approximate methods but, apart from smaller scales (<1Mpc), the results tend to agree well with the hydrodynamical simulations of Mesinger et al. (2011). We note that the EoR signal simulation used here is the same as in Chapman et al. (2012). We run the code using standard parameters from the seven-year Wilkinson Microwave Anisotropy Probe (WMAP) observations, (Ω_Λ, Ω_m, Ω_b, n, σ_8, h) = (0.72,0.28,0.046,0.96,0.82,73) (Komatsu et al. 2011). We initialized the simulation with 1800^3 dark matter particles at $z = 300$. The code uses linear perturbation theory to evolve the initial density and velocity fields to the redshifts of the EoR. The velocity fields used to perturb the initial conditions were formed on a grid of 450^3 and then interpolated to a finer grid of 512^3. 21CMFAST utilizes an excursion set formalism to find dark matter haloes. For our simulation, we set $10^9\,\mathrm{M}_\odot$ to be the threshold for haloes contributing ionizing photons. Once the evolved density, velocity and ionization fields are obtained, the code computes the brightness temperature δT_b box at each redshift based on the following equation:

$$\delta T_\mathrm{b} = 28\,\mathrm{mK} \times (1+\delta) x_{\mathrm{H\,I}} \left(1 - \frac{T_\mathrm{CMB}}{T_\mathrm{spin}}\right) \left(\frac{\Omega_\mathrm{b} h^2}{0.0223}\right) \times \sqrt{\left(\frac{1+z}{10}\right)\left(\frac{0.24}{\Omega_\mathrm{m}}\right)}, \quad (16)$$

where the brightness temperature fluctuation δT_b is detected as a difference from the background cosmic microwave background (CMB) temperature T_CMB (Field 1958, 1959; Ciardi & Madau 2003), h is the Hubble constant in units of $100\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, $x_{\mathrm{H\,I}}$ is the neutral hydrogen fraction and Ω_b and Ω_m are the baryon and matter densities in critical density units. For simplicity, we ignore the gradient of the peculiar velocity, the contribution of which to the brightness temperature is relatively small (Ghara, Choudhury & Datta 2012; Shimabukuro et al. 2015). We also neglected spin temperature fluctuations by assuming $T_\mathrm{spin} \gg T_\mathrm{CMB}$, i.e. the neutral

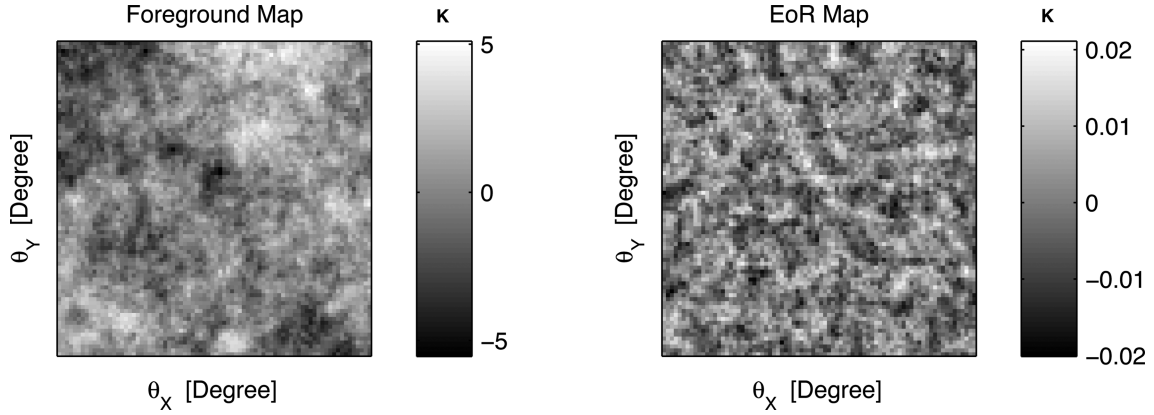

Figure 3. This figure shows a slice at 165 MHz for the diffuse foreground emission and the EoR signal at a pixel resolution of 3 arcmin. Note that the mean is subtracted from the maps. Each panel covers a region of $2^\circ \times 2^\circ$.

gas has been heated well above the CMB temperature during the EoR epoch (Pritchard & Loeb 2008).

3.2 Diffuse foregrounds

The diffuse foreground model used in this article is described in great detail in Jelić et al. (2008, 2010). These simulations include Galactic diffuse synchrotron emission (GDSE), Galactic localized synchrotron emission, Galactic diffuse free–free emission and unresolved extragalactic foregrounds. The GDSE is produced by cosmic-ray electrons in the magnetic field of the Galaxy (Ginzburg & Syrovatskii 1969). The intensity and the spectral index of the GDSE are modelled as Gaussian random fields. The Gaussian fields are generated assuming a power-law spatial power spectrum with 2D index of −2.7. The GDSE is modelled as a power law as a function of frequency with a spectral index of -2.55 ± 0.1 (Shaver et al. 1999). The amplitude of the brightness temperature is set around to 253 K at 120 MHz.

Another part of the synchrotron emission comes from Galactic localized supernova remnants (SNRs). Along with the GDSE, SNRs make up 70 per cent of the total foreground emission after point sources are removed. Inside the observational window of $10^\circ \times 10^\circ$, eight SNRs were placed randomly. We note that, due to the large computational requirement for our current ML inversion, we have restricted our field of view (FoV) to $4^\circ \times 4^\circ$, which corresponds to the FWHM of the LOFAR primary beam. The SNRs are modelled as extended discs and their angular size, flux density and spectral index are chosen randomly from the observed SNR catalogue of D.A. Green (A Catalogue of Galactic Supernova Remnants, 2006 April version[6]).

The free–free emission is generated through thermal bremsstrahlung radiation from our Galaxy and external galaxies from the diffused ionized gas. It is also modelled as a Gaussian random field similar to GDSE, but with a different frequency spectral index fixed to −2.15 (Tegmark et al. 2000; Santos, Cooray & Knox 2005). Compared with other foreground components, the free–free emission only contributes 1 per cent to the total foreground emission (Shaver et al. 1999), but is orders of magnitude higher than the 21-cm signal.

Unresolved extragalactic sources, mainly radio galaxies and clusters, contribute almost ∼27 per cent of the total foreground emission after discrete point sources are removed from the map (Jelić et al. 2008). The simulated radio galaxies have a power-law spectrum and are clustered based on a random-walk algorithm. The radio clusters are chosen from the cluster catalogue of the Virgo Consortium[7] and using the corresponding relation between the observed mass luminosity and X-ray radio luminosity functions.

Fig. 3 shows one of the slices at 165 MHz for the diffuse foreground and the EoR signal that we use in this analysis.

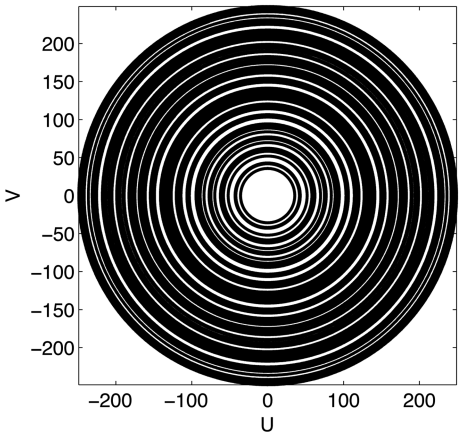

Figure 4. The figure shows the uv coverage of our LOFAR simulation. We do not show the w component here, but in our analysis we take the w component into account. The uv points are generated within a 30–250 uv cut. The uv points and the Hermitian uv values are shown in this figure. Here (u, v) are the antenna separations in wavelength units at a frequency of 165 MHz.

3.3 Visibility data

We use the LOFAR–High Band Antenna (HBA) positions (van Haarlem et al. 2013) to generate the baseline components (u, v, w) towards the North Celestial Pole (NCP), which were used to generate the visibility data for our simulation. Fig. 4 shows the uvw coverage that we have used for our simulation. To ensure a manageable data set, we have sampled visibilities for each 100-s interval for a uv track covering the full ±6 h angle. The components u and v specify the baseline coordinates projected on the plane

[6] https://www.mrao.cam.ac.uk/surveys/snrs/

[7] http://www.mpa-garching.mpg.de/galform/virgo/hubble/

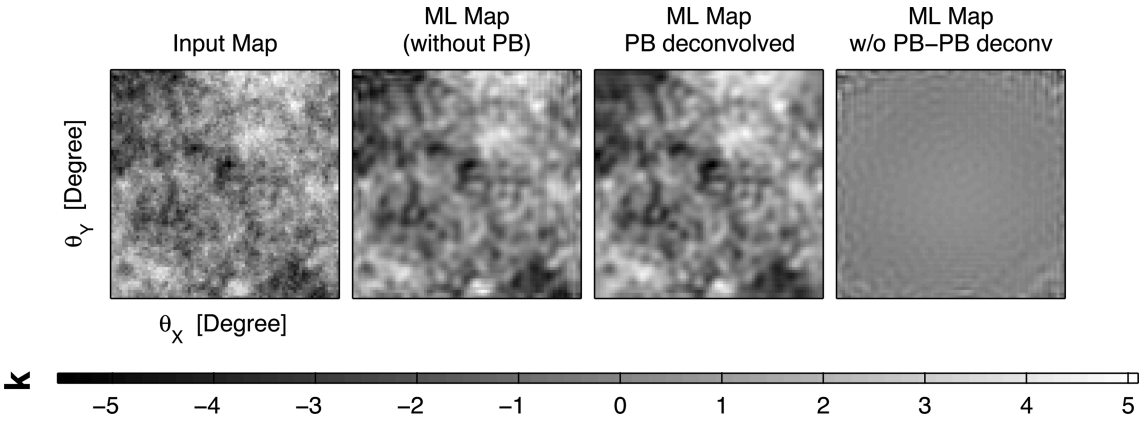

Figure 5. From left to right, the $2^\circ \times 2^\circ$ panels respectively show the true input map without any noise, the inverted map without the any effect of PB, the PB-deconvolved map (where the PB was applied during creation of visibility data and the beam was included in the Fourier transform kernel to find the corresponding beam-deconvolved ML solutions) and the difference between the second and third panels. We see that the difference map is very close to zero, apart from the corner of the field where the PB drops significantly from the centre of the field and the noise increases considerably, which may have some effect on the inversion.

tangent to the NCP, while the third component w of the coordinate frame points towards the direction of the NCP. We note that the actual sampling of uvw points in the LOFAR observation will be higher, but here we restrict our number of visibility samples, as we are limited by the present computational resources. Note that the number of visibilities is chosen to be about equal to the number of independent uvw cells, hence after gridding the computational effort remains similar even if the number of visibilities goes up.

The maximum baseline length is restricted to 250λ, which corresponds to a maximum wavelength on the sky of $\sim$13 arcmin. In our inversion, we choose a pixel scale of $\sim$3 arcmin, which ensures that we sample the PSF reasonably well.

3.4 Fourier kernel matrix

A typical radio interferometric array measures visibilities at a large number of baselines and frequency channels simultaneously. Ideally, each visibility records a single mode of the Fourier transform of the specific intensity distribution $I_\nu(\ell, m)$ on the sky. Representing the celestial sphere by a unit sphere, the component n can be expressed in terms of (l, m) by $n(l, m) = \sqrt{1 - l^2 - m^2}$. Then the measured visibility for a monochromatic, unpolarized signal is related to the specific intensity distribution through the following relation (Perley, Schwab & Bridle 1989):

$$\mathcal{V}_\nu(u, v, w) = \int \frac{I_\nu(l, m)}{\sqrt{1 - l^2 - m^2}} \times e^{-2\pi i(ul+vm+w(\sqrt{1-l^2-m^2}-1))}\, dl\, dm. \quad (17)$$

We note that, without loss of generality (assuming that all stations have identical beams), we have not incorporated any primary beam pattern in the visibility definition, which in general can be absorbed into the specific intensity distribution. To test this, we have introduced a Gaussian primary beam (PB) in the sky with a FWHM of 2° within a field of view of 4° by 4° and subsequently created the input visibility data. During the inversion, we absorb the PB in the response matrix (as a directionally dependent amplitude to the Fourier kernel) to obtain the corresponding 'PB-deconvolved' true sky. We compare these solutions with the solutions presented in this article where we have not introduced any PB in our analysis and find that these two sets of solutions are very similar (Fig. 5). We note that here we assume our inversion grid is in l, m coordinates and the sky phase centre corresponds to the centre of the grid. Each visibility corresponding to a uvw value has a contribution from all pixels in the field of view of 4° with a pixel resolution of 3 arcmin. To relate it to our matrix notation (in Section 2), we recall that here $\mathcal{V}$ is our data matrix $\boldsymbol{V}$, the Fourier kernel $e^{-2\pi i[ul+vm+w(\sqrt{1-l^2-m^2}-1)]}$ is the response matrix $\mathbf{F}$ and $\frac{I_\nu(l,m)}{\sqrt{1-l^2-m^2}} \times d\Omega_p$ is the model sky parameter that we want to infer from the visibility data.

In our simulation, we generate visibilities at exactly the same (u, v, w) coordinates for all frequency channels. This ensures that, for a limited sampling of uvw points, the PSF will not vary with frequency, which in principle can be a major obstacle for foreground removal (Bowman, Morales & Hewitt 2009). We note that the case of real LOFAR observations, where the uvw sampling is quite dense, ensures that the PSF variation with frequency can be made very small with the proper weighting and gridding. LOFAR observations of the NCP show that this can be done to a level of better than 30 dB (i.e. $<10^{-3}$). On scales of a few arcmin, this leads to <1 mK chromatic effects from any remaining compact source of a few mJy. For the extended emission we are simulating, this will be even less. Therefore, in this simulation we choose not to scale the uvw values with frequency, so as not to introduce any additional effects due to our (currently) limited uvw samples. We also performed a test where we generated another set of uvw samples by scaling the uvw values (U_0) with frequency as $U_n = U_0(1 + n\Delta\nu/\nu_0)$. (The number of frequency channels $n = 20$, channel width $\Delta\nu = 0.5$ MHz, middle of the frequency band $\nu_0 = 170$ MHz.) Then we use only those visibilities within a maximum uv cut of 250 and redo the ML analysis. We find that the resultant ML solutions are quite similar to uvw values (those with U_0) with no frequency scaling (Fig. 6). Moreover, proper filtering of side lobes via delay-rate filtering (Vedantham, Udaya Shankar & Subrahmanyan 2012) can mitigate the remaining side-lobe leakage. We therefore believe our current approximation is justified.

We simulated visibilities over 20 frequency channels of 0.5-MHz resolution covering a frequency range of 165–175 MHz, where the EoR signal peaks in the simulation (Mesinger & Furlanetto 2007; Mesinger et al. 2011). The restricted bandwidth ensures that the evolution of the 21-cm signal along the line-of-sight direction due to the light-cone effect (Datta et al. 2012) will not be a major concern.

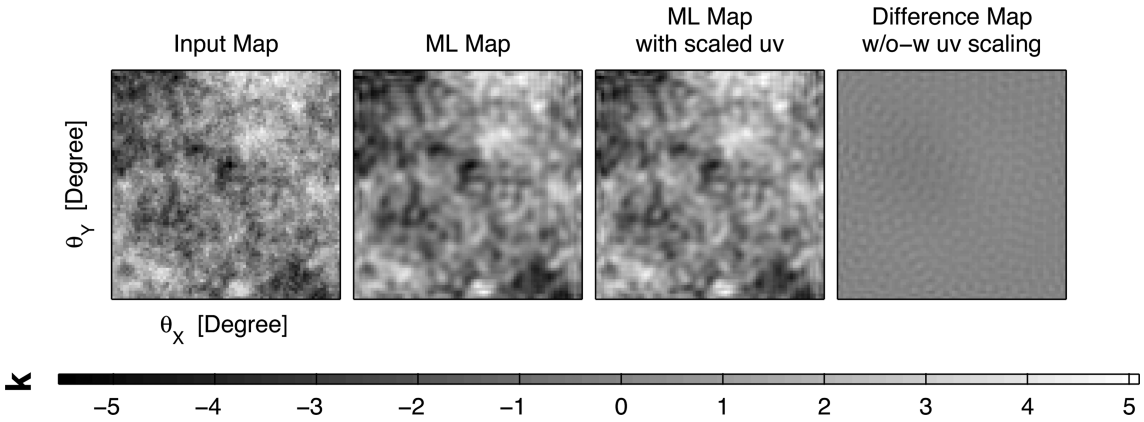

Figure 6. The top left panel shows the input map without any noise. The remaining panels are as follows: the ML inverted map (w/o PB) where the uv values (U_0) correspond to the start of the frequency band; the ML solutions where the uv values (U_n) are scaled according to $U_n = U_0(1 + n\Delta\nu/\nu_0)$ and we have selected a subsection of visibilities from these scaled uv values that lie below $U_{\rm max} = 250$ to reconstruct the ML solutions; the last panel show the difference between the second and third panels, which is mostly close to 0. Each panel covers a region of $2^\circ \times 2^\circ$.

In addition to the signal component, each visibility also has a random noise contribution. We added Gaussian noise of root-mean-square error (rms) 60 mK in the real and imaginary parts of the sampled visibilities to mimic actual LOFAR observations. This is close to the expected rms level of noise after 600 h of LOFAR observations (Chapman, Zaroubi & Abdalla 2014). We note that this noise value is indicative only and may change somewhat depending upon the observations.

4 ML INVERSION

In this section, we present our ML solutions based on the visibility data that we generated (Section 3.3). The computational effort of this linear inversion problem is in general of order $\sim N_{\rm s}^3$. For $N_{\rm s} = 80 \times 80$ pixels, this leads to 10^{11-12} floating-point operations per frequency channel. To overcome this, we run this inversion on a parallel 2-TB memory machine. Also creating the matrices from the data itself takes a similar effort.

However, if we have a good estimate of an initial solution (which we can in principle obtain via FFTs from the dirty images for different channels within the frequency bands), then the inversion is reduced to $\sim N_{\rm s}^2$, where $N_{\rm s}$ is the number of sky pixels, where the '$\sim$' stands for 4–5 iterations in a conjugate gradient (or quasi-Newtonian) optimization scheme. Here, because each frequency slice looks similar to the adjacent one, we use the ML solution of the previous slice as a starting point for the subsequent slice. In this article, we have performed our analysis with both non-gridded and gridded data to assess whether the solutions differ substantially when creating the visibility data matrix, the response matrix (the Fourier kernel that is a filled matrix) and the noise covariance matrix. For a large number of visibility samples, this is quite time-consuming and requires a lot of computer memory. Therefore, we restrict our visibility data to a size that we expect in real gridded data (about 128^2 gridded visibilities) as well choosing to restrict our field of view to 4° by 4° with a pixel resolution of 3 arcmin (3 arcmin and above, due to the increase of noise below this scale), which will mostly be sensitive enough to detect the 21-cm signal in LOFAR. Overall, we think these choices are fairly reasonable and comparable to grid sampling and fields of view that are used in real observations.

4.1 Resulting inversions

Fig. 7 shows the ML solutions based on the visibility data sets that we use. The topmost left panel shows the input map, where the mean is subtracted from the input template maps, as interferometers such as LOFAR are only sensitive to fluctuations in the sky. Subsequent maps are the results of the ML inversion with non-gridded and gridded visibility data with grid pixel sizes of 1, 2, 4 and 8λ respectively. For each of these, we re-optimize the regularization parameter to maximize the Bayesian evidence. We note that finer binning is always useful, since it provides a better approximation of the visibility positions in the uvw cube. We averaged the visibilities according to their w components, where the number of w slices is set by $N_w = \lambda B/D^2$, where λ is the observing wavelength, B is the maximum baseline length and D is the station diameter (Perley, Schwab & Bridle 1989). The pixel resolution of the ML simulation is set to a 3-arcmin scale, sufficient to recover the EoR signal (Mesinger & Furlanetto 2007). This also ensures that we subsample the PSF reasonably well (more than four times). We note that the ML inversion maps look smoother relative to the input map, because we do not sample all the small angular scales with our limited uvw. Hence, a direct comparison in the image domain remains difficult because of this 'filtering'.

4.2 Data and model coherency and foreground power spectrum

Next, we computed the visibility coherency matrix between the input data and reconstructed ML solutions. We define the coherency matrix as

$$V_{\rm coh} = \frac{|\langle V_{\rm data} V_{\rm ML}^* \rangle|^2}{\langle |V_{\rm data}|^2 \rangle \langle |V_{\rm ML}|^2 \rangle} \quad (18)$$

and evaluate this in bins of baselines and across frequency channels. For a perfect reconstruction, we expect $V_{\rm coh}$ to be 1, although the effect of noise will lower this value a little (although the noise is small compared with the foreground flux). The results are shown in Fig. 8. We notice that, apart from $\Delta u = 8$, the coherency matrix is mostly within 0.99–1.0 and this suggests that the inversion works well when the data are not averaged heavily.

We also computed the power spectrum (PS) of the input map and the ML inverted maps corresponding to the gridded and non-gridded viability data. We interpolated the PS to common baseline

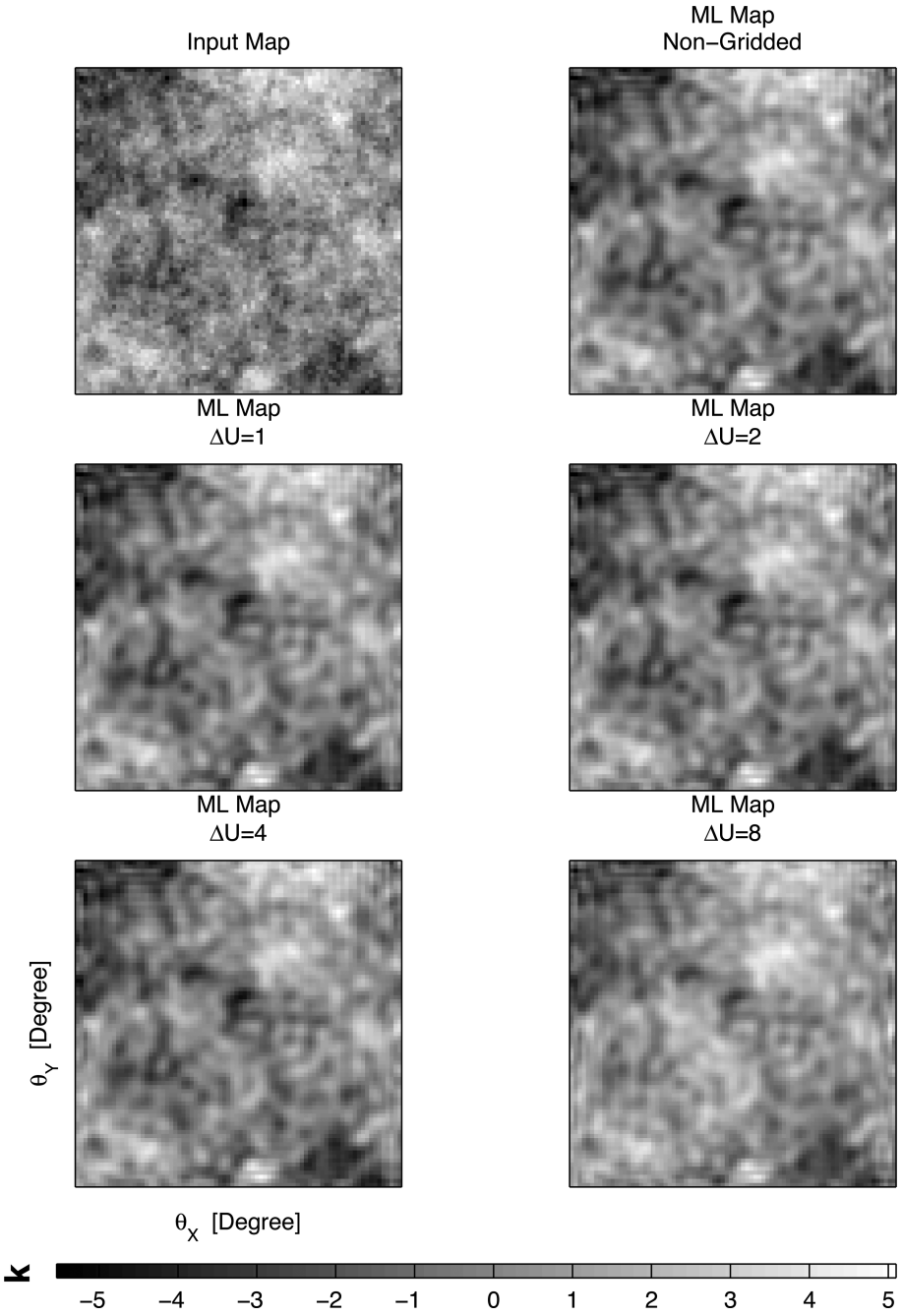

Figure 7. The top left panel shows the input map without any noise. The remaining panels show the ML inverted maps for non-gridded and gridded visibility data with grid sizes of $\Delta u = 1$, 2, 4 and 8λ, respectively. For the ML inverted maps, we have added 60 mK of noise to the real and imaginary part of the visibility data, as expected for LOFAR after 600 h of integration. Each panel covers a region of $2^\circ \times 2^\circ$.

values so that we can compare different maps. The resulting PS are shown in Fig. 9. We find that, due to the averaging of visibilities with different grid sizes, there is an overall drop in the power spectrum on all scales, though the effect is more pronounced on smaller scales.

In Table 1, we tabulated the reduced χ^2 values for the non-gridded and gridded visibility data sets. We note that here we do not compute the evidence values, as the visibility data change each time for different grid sizes (Δu). We define the reduced χ^2 using the number of degrees of freedom as $N_\mathrm{b} - \gamma$, where N_b is the number of visibility data (non-gridded or gridded) that is used in the inversion and γ is the right-hand side of equation (15), which determines the number of 'good' parameters determined by the visibility data. The reduced χ^2 commonly determines the goodness of the fit and its value is close to 1. We note that the reduced χ^2 is not strictly 1.0 because Bayesian analysis ensures that evidence is maximized instead of trying to reach a reduced χ^2 of 1.0. We find that for the non-gridded visibility data the reduced χ^2 is very close to unity.

Up to $\Delta u = 2$, the reduced χ^2 value remains near 1, after which it increases. Based on this, we decided to carry on our subsequent analysis with the visibility data that have been gridded with $\Delta u = 2$, as a good compromise between computational speed in the ML inversion and evidence maximization and goodness of fit. We note

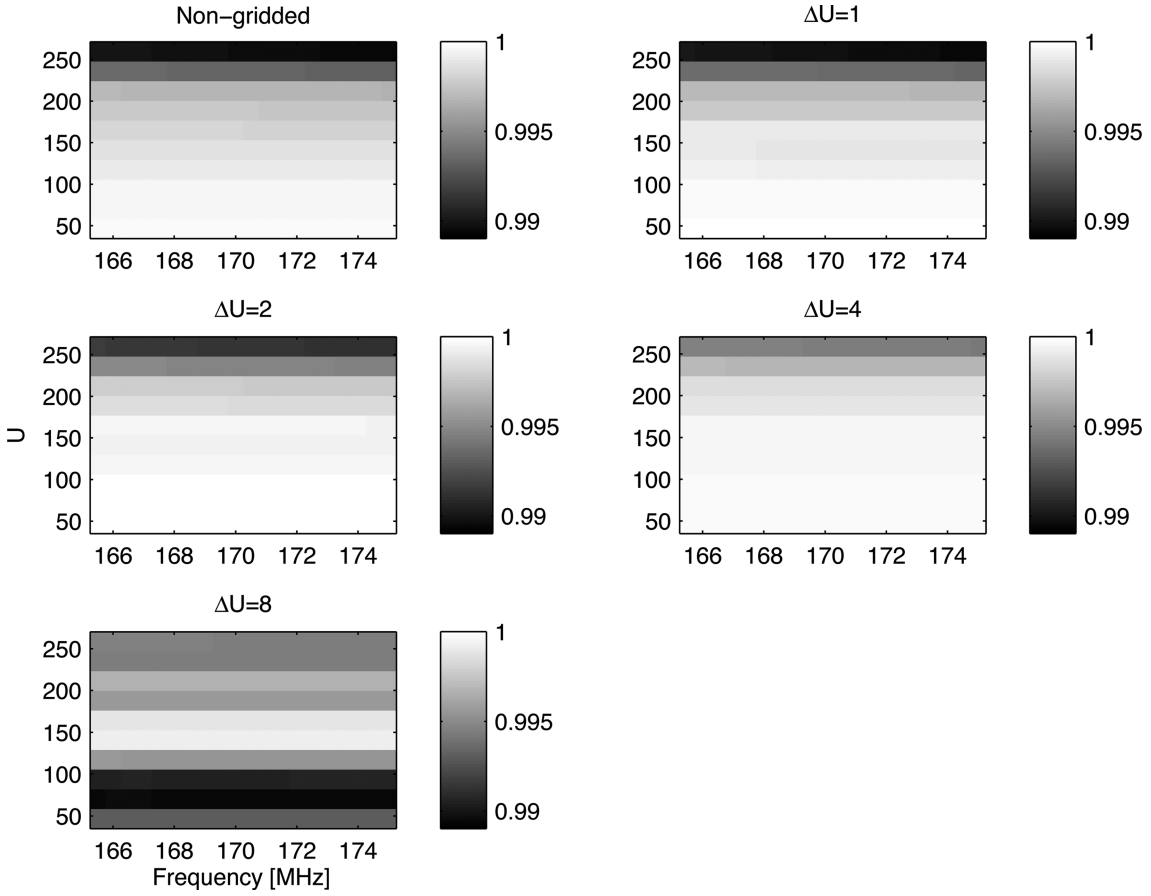

Figure 8. The visibility coherency plot between the input data and the corresponding ML non-gridded and gridded visibility data.

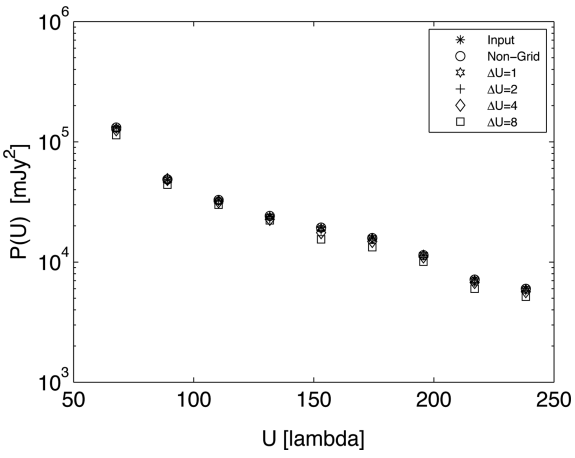

Figure 9. The spatial power spectrum corresponding to the input and ML maps as a function of baseline length. Each panel shows a region of $2^\circ \times 2^\circ$.

that ML inversion maps still contain the FG and EoR signal and that the noise in this process is largely filtered. The level to which this has succeeded will be tested in the subsequent section.

5 FOREGROUND REMOVAL

Low-frequency radio observations suggest that the foregrounds are several orders of magnitude larger than the redshifted 21-cm signal (Bernardi et al. 2009, 2010; Ghosh et al. 2011, 2012). Therefore, removing the foregrounds is possibly the biggest challenge for detecting the faint cosmological 21-cm signal. The foregrounds are expected to have smooth spectra, while the 21-cm signal is line emission, which varies rapidly with frequency (although possibly less rapidly spatially). This property of the H I signal holds the promise of allowing us to separate the signal from the foregrounds.

A possible line of approach is to represent the sky signal as an image cube and, for each angular position, use polynomial fitting to subtract out the smooth component of the sky signal that varies slowly with frequency (Santos et al. 2005; McQuinn et al. 2006;

Table 1. The reduced χ^2 values evaluated at the regularization constant ($\hat{\lambda}$) used in the simulation for non-grid and gridded visibility data for different grid sizes (Δu). The reduced χ^2 is defined as $2E_D/(N_b - \gamma)$, where $\gamma = N_x - \hat{\lambda}\mathrm{Tr}(\mathbf{A}^{-1}\mathbf{C})$ is the right-hand side of equation (15).

Gradient regularization	Non-grid	Grid ($\Delta u = 1$)	Grid ($\Delta u = 2$)	Grid ($\Delta u = 4$)	Grid ($\Delta u = 8$)
χ^2	1.046	1.07	1.14	1.83	53.21

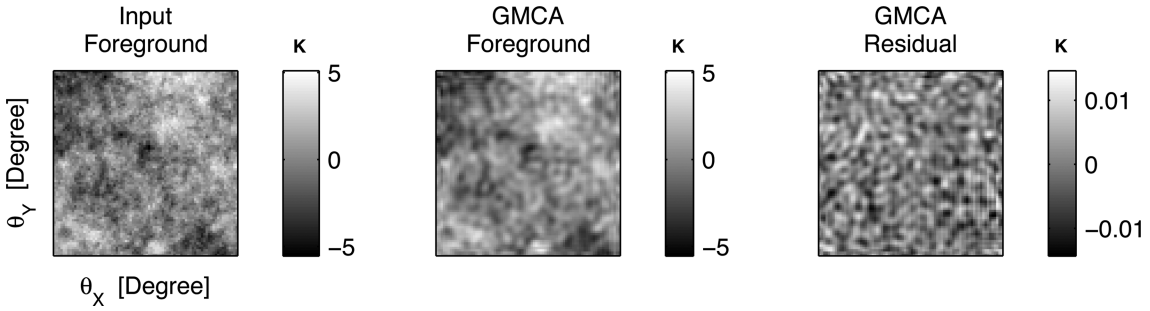

Figure 10. A slice at 165 MHz of the input foreground, GMCA reconstructed foreground model and the residual after the foregrounds are subtracted. Each panel covers a region of $2^\circ \times 2^\circ$.

Wang et al. 2006; Jelić et al. 2008; Liu et al. 2009b; Liu, Tegmark & Zaldarriaga 2009a; Petrovic & Oh 2011). The residual sky signal is expected to contain only the 21-cm signal and noise. We note that all these parametric foreground-fitting techniques depend heavily on prior knowledge of the foreground structure, which is largely unknown at the spatial resolution of the low frequencies of interest. Liu et al. (2009b) also showed that this method of foreground removal has problems, which could be particularly severe for large baselines if the uv sampling is sparse. On the other hand, non-parametric methods do not assume any specific form of the foregrounds and use the data to decide on the foreground model (Harker et al. 2009; Chapman et al. 2012, 2013). Here, we use GMCA, which is a blind source separation technique (BSS) that assumes a wavelet basis exists in which spectrally smooth foreground components are sparsely represented with minimal basis coefficients (Chapman et al. 2013) that can be separated from the 21-cm signal.

GMCA utilizes a component separation technique to define the foregrounds in order to subtract them, treating the 21-cm signal as noise. We model the foregrounds using four components, which refers to the number of foreground contributions that can be described by unique sparse descriptions. We note that, as more components are added to the model, we might expect the 21-cm signal itself to leak into the foreground model, although 4–5 components seem to work well in recovering the 21-cm signal (Chapman et al. 2013). In Fig. 10, we show one slice of the reconstructed foreground model using GMCA and the residual solution. We find an average correlation of >70 per cent between the true input foreground map and the GMCA reconstructed foreground model across the 10-MHz bandwidth.

6 DENOISING AND POWER SPECTRA

In an ideal situation, one would prefer that after ML inversion the residual visibilities are mostly dominated by noise and the ML solutions are noise-free. We find that we recover most of the input noise in the residual visibilities and the difference is around 0.1 per cent compared with the input noise for the non-gridded case. We note that this is an idealistic situation, where we assume that we know the noise level in the data and we build up the noise covariance matrix using the same noise values during the ML inversion. The GMCA residuals retain a large part of the EoR signal and we find a clear correlation (≥50 per cent) between the input EoR signal we use in our simulation and the residual GMCA solutions (Fig. 11). The mean correlation between different slices does not change significantly if we smooth the solution with Gaussian filters of different widths (Fig. 12), suggesting that the de-noising seems scale-independent. It is interesting to highlight that if we use a dirty map as an input template to GMCA then the residual map will have contributions from system noise, the 21-cm signal and possibly some foreground residuals. Moreover, the success of the signal recovery requires the noise statistics at different frequency channels to be determined at an extremely high level of precision. On the other hand, we find that the ML solutions are heavily de-noised and the residual solutions after GMCA are mostly dominated by the EoR signal. During real observations, however, EoR signal detection can be more complicated, as any unmodelled foreground power and any uncertainties in the bandpass response can prohibit such a high level of precision being achieved.

6.1 Power spectrum: recovered and input EoR

This section describes our comparison between the input EoR and the recovered EoR power spectrum after the ML maps are passed through GMCA.

6.1.1 2D power spectrum

To test the power-spectrum recovery, we compute the cylindrically binned 2D power spectra from the GMCA residual maps and compare this with the input EoR power spectrum. We find that the power spectrum is still dominated by system noise at higher $k_\perp$, where the baseline density drops. In addition, GMCA removes large-scale modes along the frequency direction and this suppresses the power spectrum severely on low $k_\parallel$ scales. However, this can be mitigated by running GMCA on wider bandwidth cubes. In our comparison, we do not include these high $k_\perp$ and low $k_\parallel$ modes. Fig. 13 shows the cylindrically binned 2D power spectra versus the input EoR signal. Note that, although we did not subtract any noise power spectrum from the recovered power spectrum, we recover the input EoR power spectrum quite well. This suggests that the ML solutions are indeed de-noised. Based on the $k_\perp$ scales where we have a reasonable match with the input EoR power spectrum and using $k_\perp = \ell/r$, where r corresponds to the comoving distance at a redshift of 7.35, we find that we are most sensitive in recovering the EoR signal at angular scales of ∼6–17 arcmin.

To inspect the PS recovery in the $k_\perp$ and $k_\parallel$ plane, we calculated the ratio of the input and recovered EoR PS. Ideally, we expect the ratio of the PS to be close to unity. In Fig. 14, we find that in most of the $k_\perp$–$k_\parallel$ plane the ratio plot is close to unity. We note that in the lowermost $k_\parallel$ bin, where LOFAR is sensitive enough to measure the PS, we find that GMCA removes most of the large-scale smooth frequency mode and the recovered EoR signal is heavily suppressed on these scales. We have therefore avoided these low $k_\parallel$ modes in Fig. 14. We find that in the range of k scales shown in Fig. 14

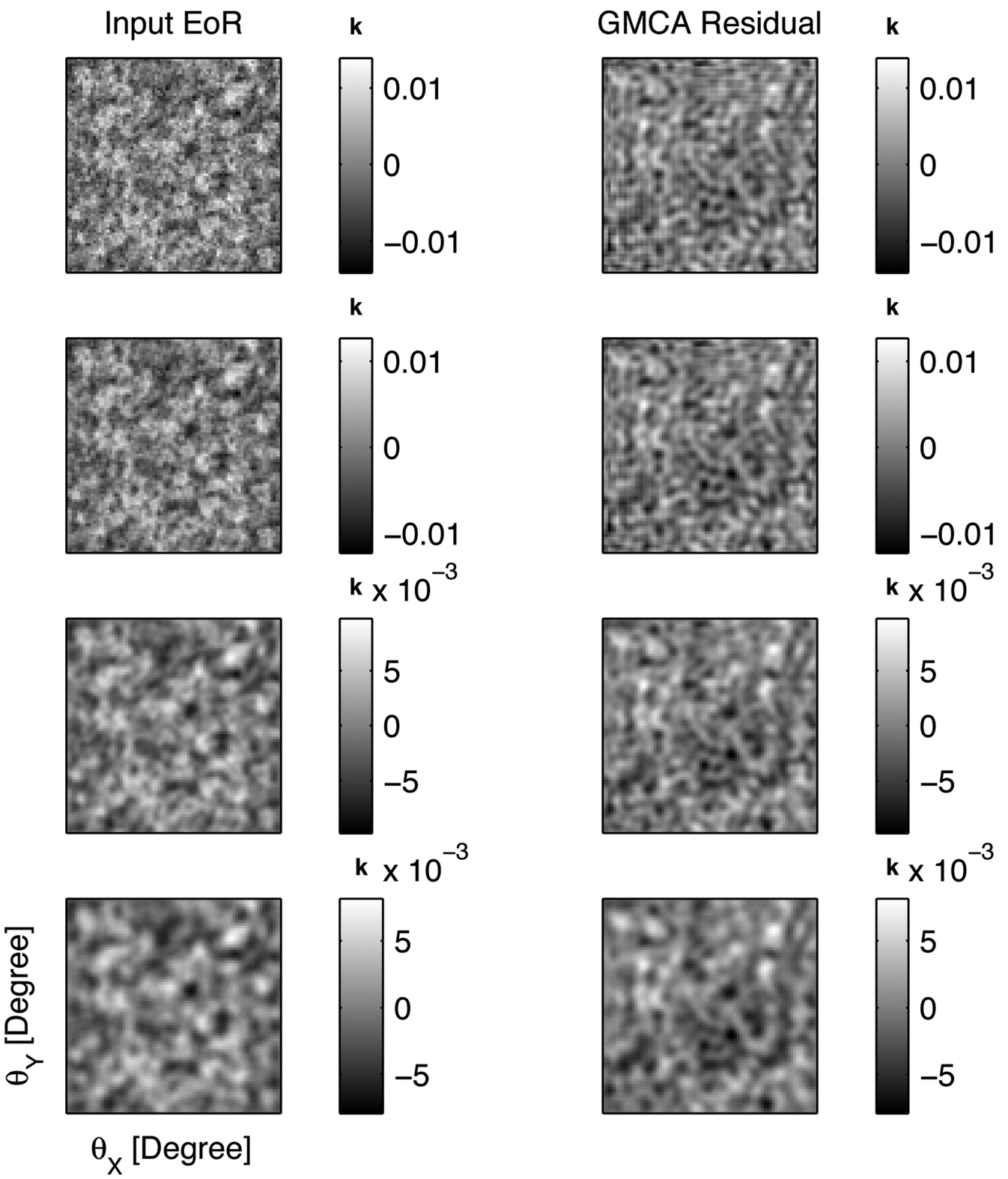

Figure 11. The left panel shows the input EoR maps and the right panel shows the GMCA residual maps. In the top row, no smoothing has been applied and in subsequent rows we applied smoothing with sigma = 0.5, 1 and 2, respectively, for a Gaussian filter with size [5 5] pixels. Each panel covers a region of $2^\circ \times 2^\circ$.

the PS ratio varies from ∼0.9–1.4, with an overall rms scatter of 0.13, and there are some k modes for which the ratio is slightly higher. It is interesting to point out here that the ML inversion filters noise spatially and GMCA removes smooth foregrounds from the frequency direction. Therefore, when comparing the recovered and input PS, we have not subtracted any noise PS from the GMCA residuals. This may lead to some extra noise bias in some of the k modes where the ratio is larger than 1. In the spherically averaged

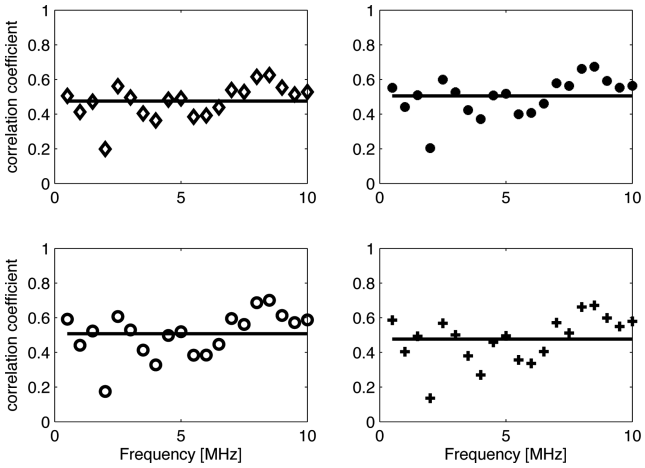

Figure 12. The correlation coefficient between the GMCA residual and input EoR as a function of frequency. The black horizontal line shows the mean of the correlation. We choose a Gaussian filter with size [5 5] pixels and different sigma values. Clockwise from the top left panel, the figures show the correlation coefficients corresponding to no smoothing and smoothing with $\sigma = 0.5$, 1 and 2, respectively.

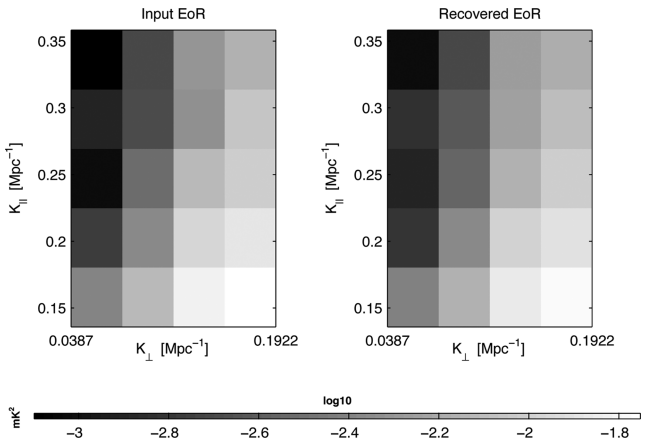

Figure 13. This figure shows the cylindrically binned 2D PS for the input and the recovered EoR.

power spectrum (Fig. 15), we find a trend where the foregrounds are over-subtracted for the small-k modes and there is a hint of some extra residual noise on the larger k modes. Despite these differences, at a level of several tens of per cent, we emphasize that the results in Fig. 14 give us a good representation of how well our Bayesian plus GMCA method recovers the EoR signal.

6.1.2 *Spherically average power-spectrum*

Next, we average the PS in spherical shells and we compute the spherically averaged dimensionless power spectrum, $\Delta^2(\mathrm{k}) = \mathrm{k}^3 P(\mathrm{k})/2\pi^2$. In Fig. 15, we show the input and recovered spherically averaged PS with 2σ noise fluctuations. The noise contribution only includes sample variance and the system noise makes a relatively smaller contribution at these low k modes, where we are most sensitive in recovering the EoR signal. We find that the recovered EoR signal is mostly within 2σ error bars in the k range shown in Fig. 15.

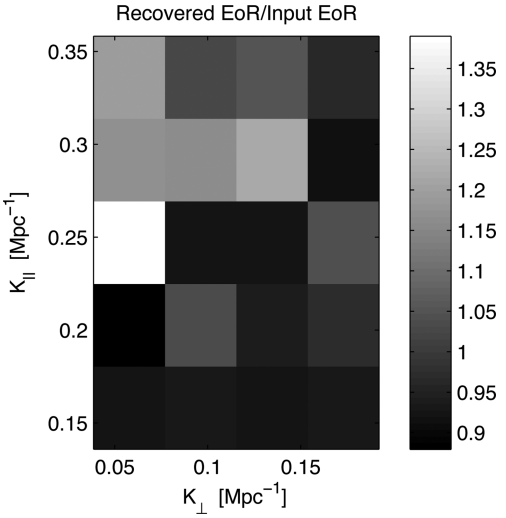

Figure 14. This figure shows the ratio of the recovered and input EoR PS.

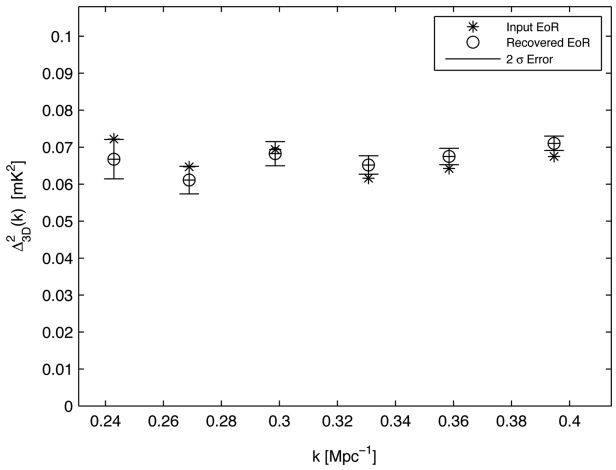

Figure 15. The spherically averaged 3D PS for the input and recovered EoR signal, along with 2σ sample variance. Note some deviations at the higher k values, probably due to left-over noise in the image residuals after ML inversion and GMCA foreground removal. Note that we have not attempted to remove the residual noise power spectrum.

7 CONCLUSIONS AND FUTURE WORK

In this article, we have introduced a Bayesian framework with spatial regularization to recover the diffuse foregrounds and the redshifted 21-cm H I signal from a set of calibrated (gridded) visibilities. We have shown that gradient regularization with an optimal regularization constant that maximizes the evidence works well in reconstructing the input signal.

Given the large size of the current data sets, it will be a large computational effort to carry out the inversion directly from the calibrated visibility data sets, but we have demonstrated that gridded visibility data can be used effectively in ML inversion, providing reduced χ^2 values close to unity. Finer binning gives better results, since it provides a better approximation to the position of the visibilities in the uv plane. However, binning on scales less than 2λ is not necessary. We note that this is comparable to the current (u, v) cell sizes in LOFAR analyses.

We used a non-parametric foreground removal technique, GMCA (Chapman et al. 2013), to remove the smooth diffuse foregrounds. In the case of a cylindrical power spectrum, we can mostly recover the

input EoR signal within a region of $0.14\,\mathrm{Mpc}^{-1} < k_{\parallel} < 0.35\,\mathrm{Mpc}^{-1}$ and $0.03\,\mathrm{Mpc}^{-1} < k_{\perp} < 0.19\,\mathrm{Mpc}^{-1}$. The high $k_{\perp}$ and low $k_{\parallel}$ modes beyond these limits that are accessible in LOFAR observations are mostly dominated by system noise and the removal of smooth large-scale modes in the frequency direction by GMCA. Scales $0.24\,\mathrm{Mpc}^{-1} < k < 0.40\,\mathrm{Mpc}^{-1}$ are largely (within 2σ) recovered in the spherically averaged 3D input EoR power spectrum. However, at higher k scales the residual noise has some effects and the recovered EoR level is slightly higher compared with the input power spectrum.

In the near future, we plan to extend our Bayesian imaging technique for a full Stokes analysis, where we will compare the ML solution with those of the more classical ones and quantify the errors and correlation between all model parameters precisely. We will also apply our technique to real LOFAR data, where most of the point sources have been subtracted and the residual visibilities are dominated by unresolved source confusion noises. Another goal is to include direction-dependent (such as ionosphere, beam etc.) and direction-independent calibration errors (such as gain variation) in predicting the corresponding sky from the measured data.

ACKNOWLEDGEMENTS

AG and LVEK acknowledge financial support from the European Research Council under ERC-Starting Grant FIRSTLIGHT – 258942 (PI: LVEK). VJ thanks the Netherlands Foundation for Scientific Research (NWO) for financial support through VENI grant 639.041.336. AG thanks Saleem Zaroubi, Ajinkya Patil and Khan M. B. Asad for many useful discussions. We thank the anonymous referee for providing us with constructive suggestions, which helped to improve the article.

REFERENCES

Ali Z. S. et al., 2015, preprint (arXiv:1502.06016)
Asad K. M. B. et al., 2015, preprint (arXiv:1503.01644)
Barkana R., Loeb A., 2001, Phys. Rep., 349, 125
Becker R. H. et al., 2001, AJ, 122, 2850
Bernardi G. et al., 2009, A&A, 500, 965
Bernardi G. et al., 2010, A&A, 522, A67
Bobin J., Starck J.-L., Moudden Y., Fadili M. J., 2008, in Hawkes P. W., ed., Advances in Imaging and Electron Physics, Vol. 152. Elsevier, Amsterdam, p. 221
Bowman J. D., Morales M. F., Hewitt J. N., 2009, ApJ, 695, 183
Chapman E. et al., 2012, MNRAS, 423, 2518
Chapman E. et al., 2013, MNRAS, 429, 165
Chapman E., Zaroubi S., Abdalla F., 2014, preprint (arXiv:1408.4695)
Ciardi B., Madau P., 2003, ApJ, 596, 1
Datta K. K., Mellema G., Mao Y., Iliev I. T., Shapiro P. R., Ahn K., 2012, MNRAS, 424, 1877
de Oliveira-Costa A., Tegmark M., Gaensler B. M., Jonas J., Landecker T. L., Reich P., 2008, MNRAS, 388, 247
Di Matteo T., Perna R., Abel T., Rees M. J., 2002, ApJ, 564, 576
Dillon J. S. et al., 2014a, Phys. Rev. D, 89, 023002
Dillon J. S. et al., 2014b, preprint (arXiv:1410.0963)
Fan X., Narayanan V. K., Strauss M. A., White R. L., Becker R. H., Pentericci L., Rix H.-W., 2002, AJ, 123, 1247
Field G. B., 1958, Proc. IRE, 46, 240
Field G. B., 1959, ApJ, 129, 536
Furlanetto S. R., Oh S. P., Briggs F. H., 2006, Phys. Rep., 433, 181
Geil P. M., Gaensler B. M., Wyithe J. S. B., 2011, MNRAS, 418, 516
Ghara R., Choudhury T. R., Datta K. K., 2012, MNRAS, 424, 1877
Ghosh A., Bharadwaj S., Ali S. S., Chengalur J. N., 2011, MNRAS, 411, 2426
Ghosh A., Prasad J., Bharadwaj S., Ali S. S., Chengalur J. N., 2012, MNRAS, 426, 3295
Giardino G., Banday A. J., Fosalba P., Górski K. M., Jonas J. L., O'Mullane W., Tauber J., 2001, A&A, 371, 708
Giardino G., Banday A. J., Górski K. M., Bennett K., Jonas J. L., Tauber J., 2002, A&A, 387, 82
Ginzburg V. L., Syrovatskii S. I., 1969, ARA&A, 7, 375
Harker G. et al., 2009, MNRAS, 397, 1138
Haslam C. G. T., Salter C. J., Stoffel H., Wilson W. E., 1982, A&AS, 47, 1
Iacobelli M. et al., 2013, A&A, 558, A72
Jeffreys H., 1961, The Theory of Probability, 3rd edn. Clarendon Press, Oxford, p. 432
Jelić V. et al., 2008, MNRAS, 389, 1319
Jelić V., Zaroubi S., Labropoulos P., Bernardi G., de Bruyn A. G., Koopmans L. V. E., 2010, MNRAS, 409, 1647
Jonas J. L., Baart E. E., Nicolson G. D., 1998, MNRAS, 297, 977
Kass R. E., Raftery A. E., 1995, J. Amer. Statist. Assoc., 90, 791
Komatsu E. et al., 2011, ApJS, 192, 18
Koopmans L. V. E., 2005, MNRAS, 363, 1136
Larson D. et al., 2011, ApJS, 192, 16
Lazarian A., Pogosyan D., 2012, ApJ, 747, 5
Liu A., Tegmark M., Zaldarriaga M., 2009a, MNRAS, 394, 1575
Liu A., Tegmark M., Bowman J., Hewitt J., Zaldarriaga M., 2009b, MNRAS, 398, 401
Loeb A., Barkana R., 2001, ARA&A, 39, 19
MacKay D. J. C., 1992, PhD thesis, California Institute of Technology, available at: http://thesis.library.caltech.edu/25/
McQuinn M., Zahn O., Zaldarriaga M., Hernquist L., Furlanetto S. R., 2006, ApJ, 653, 815
Mellema G. et al., 2013, Exp. Astron., 36, 235
Mesinger A., Furlanetto S., 2007, ApJ, 669, 663
Mesinger A., Furlanetto S., Cen R., 2011, MNRAS, 411, 955
Morales M. F., Wyithe J. S. B., 2010, ARA&A, 48, 127
Oh S. P., Mack K. J., 2003, MNRAS, 346, 871
Paciga G. et al., 2013, MNRAS, 433, 639
Page L. et al., 2007, ApJS, 170, 335
Parsons A. R., Backer D. C., 2009, AJ, 138, 219
Parsons A. R., Liu A., Ali Z. S., Cheng C., 2015, preprint (arXiv:1503.05564)
Perley R. A., Schwab F. R., Bridle A. H., eds, 1989, ASP Conf. Ser. Vol. 6, Synthesis Imaging in Radio Astronomy. Astron. Soc. Pac., San Francisco
Petrovic N., Oh S. P., 2011, MNRAS, 413, 2103
Pritchard J. R., Loeb A., 2008, Phys. Rev. D, 78, 103511
Pritchard J. R., Loeb A., 2012, Rep. Progr. Phys., 75, 086901
Reich W., 1982, A&AS, 48, 219
Reich P., Reich W., 1988, A&AS, 74, 7
Robertson B. E. et al., 2013, ApJ, 768, 71
Santos M. G., Cooray A., Knox L., 2005, ApJ, 625, 575
Shaver P. A., Windhorst R. A., Madau P., de Bruyn A. G., 1999, A&A, 345, 380
Shimabukuro H., Yoshiura S., Takahashi K., Yokoyama S., Ichiki K., 2015, MNRAS, 451, 4986
Spergel D. N. et al., 2007, ApJS, 170, 377
Suyu S. H., Marshall P. J., Hobson M. P., Blandford R. D., 2006, MNRAS, 371, 983
Tegmark M., Eisenstein D. J., Hu W., de Oliveira-Costa A., 2000, ApJ, 530, 133
van Haarlem M. P. et al., 2013, A&A, 556, A2
Vedantham H., Koopmans L. V. E., 2014, preprint (arXiv:1412.1420)
Vedantham H., Udaya Shankar N., Subrahmanyan R., 2012, ApJ, 745, 176
Waelkens A. H., Schekochihin A. A., Enßlin T. A., 2009, MNRAS, 398, 1970
Wang X., Tegmark M., Santos M. G., Knox L., 2006, ApJ, 650, 529
Yatawatta S., Zaroubi S., de Bruyn G., Koopmans L., Noordam J., 2008, preprint (arXiv:0810.5751)
Zheng H. et al., 2014, MNRAS, 445, 1084

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