

Cosmic Dawn and Epoch of Reionization Foreground Removal with the SKA

Emma Chapman^{*a}, Anna Bonaldi^b, Geraint Harker^a, Vibor Jelić^{cd}, Filipe B. Abdalla^{ae}, Gianni Bernardi^{feh}, Jérôme Bobinⁱ, Fred Dulwich^j, Benjamin Mort^j, Mario Santos^{fkl}, and Jean-Luc Starckⁱ

^aDepartment of Physics & Astronomy, University College London; ^bJodrell Bank Centre for Astrophysics, University of Manchester; ^cKapteyn Astronomical Institute, University of Groningen; ^dASTRON - the Netherlands Institute for Radio Astronomy; ^eRhodes University, South Africa; ^fSKA South Africa; ^hHarvard-Smithsonian Center for Astrophysics; ⁱCEA, IRFU, Service d'Astrophysique, France; ^jOxford e-Research Centre, University of Oxford ^kPhysics Department, University of the Western Cape, South Africa; ^lCENTRA, Instituto Superior Tecnico, Universidade de Lisboa, Portugal

E-mail: eow@star.ucl.ac.uk

The exceptional sensitivity of the SKA will allow observations of the Cosmic Dawn and Epoch of Reionization (CD/EoR) in unprecedented detail, both spectrally and spatially. This wealth of information is buried under Galactic and extragalactic foregrounds, which must be removed accurately and precisely in order to reveal the cosmological signal. This problem has been addressed already for the previous generation of radio telescopes, but the application to SKA is different in many aspects. We summarise the contributions to the field of foreground removal in the context of high redshift and high sensitivity 21-cm measurements. We use a state-of-the-art simulation of the SKA Phase 1 observations complete with cosmological signal, foregrounds and frequency-dependent instrumental effects to test both parametric and non-parametric foreground removal methods. We compare the recovered cosmological signal using several different statistics and explore one of the most exciting possibilities with the SKA — imaging of the ionized bubbles. We find that with current methods it is possible to remove the foregrounds with great accuracy and to get impressive power spectra and images of the cosmological signal. The frequency-dependent PSF of the instrument complicates this recovery, so we resort to splitting the observation bandwidth into smaller segments, each of a common resolution. If the foregrounds are allowed a random variation from the smooth power law along the line of sight, methods exploiting the smoothness of foregrounds or a parametrization of their behaviour are challenged much more than non-parametric ones. However, we show that correction techniques can be implemented to restore the performances of parametric approaches, as long as the first-order approximation of a power law stands.

Advancing Astrophysics with the Square Kilometre Array
June 8-13, 2014
Giardini Naxos, Italy

*Speaker.

1. Introduction

The statistical detection of the 21-cm reionization signal depends on an accurate and robust method for removing the foregrounds from the total signal. The first attempts to address this problem focused on exploiting the angular fluctuations of the 21-cm signal, e.g. Di Matteo et al. (2002); Oh & Mack (2003); Di Matteo, T. and Ciardi, B. and Miniati, F. (2004), but the 21-cm signal was found to be swamped by various foregrounds. The focus then moved on to the frequency correlation of the foregrounds, with the cross-correlation of pairs of maps used as a cleaning step (Zaldarriaga et al. 2004; Santos et al. 2005). This naturally evolved into methods which exploited the correlation of the foreground across whole segments of, or the entire bandwidth of, the observation: line of sight (LOS) fitting. This LOS fitting (to be discussed in much greater detail in Sec. 2) has been numerically shown to be the optimal method for power spectrum recovery (Liu & Tegmark 2011).

For this Chapter, we choose to concentrate on comparing the major LOS methods in the field, alongside the much more recent idea of avoiding the foregrounds altogether (e.g., Dillon et al. 2014) by focusing analysis on the area of Fourier space where the foregrounds are sub-dominant to the cosmological signal (see Sec. 2.3).

One of the ‘saving graces’ of foreground contamination is its smoothness over frequency space. While the foregrounds are expected to be highly correlated on the order of MHz, the cosmological signal in comparison is expected to be highly uncorrelated, e.g. Di Matteo et al. (2002); Gnedin & Shaver (2004). LOS methods can be divided into subcategories of parametric and non-parametric methods. Both aim to find the form of the smooth foreground function along frequency for each line of sight and subtract this from the total signal leaving residuals of noise, fitting errors and the cosmological signal.

The majority of early line of sight methods in the literature can be termed parametric as at some point they assume a specific form for the foregrounds, for example a polynomial (see Sec. 2.1). Despite the successes of the parametric methods, the fact remains that the form of the foregrounds is not definitively known across the frequency range and resolution of interest. Too great an assumption of their spectral form risks introducing a large element of uncertainty into the cosmological signal detection. It is with this argument that, more recently, ‘blind’ methods have been investigated. These allow the data to determine the form of the foregrounds, without assuming any particular shape beforehand (see Sec. 2.2). This has obvious advantages for a cosmological era so far not directly observed, but results are often not as promising as parametric results when applied to simulations. Arguably this is common sense, since in parametric methods one has modelled the foregrounds based on the simulation knowledge. If these methods were applied to foregrounds of different shape to the accepted form, they would risk suffering a large drop in accuracy.

Though there are now multiple proof-of-principle papers relating to LOS fitting for the EoR signal recovery, there has been little consideration given to which method aids the recovery of CD/EoR information most efficiently, accurately or precisely. Though the comparison of polynomial-fitting methods to any new method introduced is fairly common, the comparison between more complicated parametric, non-parametric methods and indeed foreground avoidance, is rare, with the exception of brief comparisons in Gu et al. (2013); Patil et al. (2014); Chapman et al. (2014). In this Chapter we aim to compare non-parametric and parametric methods on a SKA Phase 1 CD/EoR simulation, comparing the recovered signals both in terms of statistics and imaging.

In Sec. 2 we introduce the five methods used in this chapter to mitigate the simulated foreground contamination. In Sec. 3 we describe the state-of-the-art SKA Phase 1 simulation, consisting of cosmological signal, foregrounds and instrumental noise, used in this chapter. We show the results of applying the methods to these simulations in Sec. 4 before making our conclusions in Sec. 5.

2. Comparing Foreground Removal Methods

In this section we describe the methods of LOS foreground removal applied in this chapter. We divide the methods into those which assume a functional form for the foreground signal (parametric) and those which loosen the constraints on this form somewhat (non-parametric).

2.1 Parametric Methods

2.1.1 Polynomial fitting

The simplest method for foreground removal in total intensity is polynomial fitting in frequency or log frequency, e.g. McQuinn et al. (2006); Morales et al. (2006); Gleser et al. (2008); Jelić et al. (2008); Bowman et al. (2006); Liu et al. (2009); Petrovic & Oh (2011).

The usual method of polynomial fitting is to fit the total observed spectrum along the line of sight with a smooth function such as a n -th order polynomial: $\log T_{b,fg}(v) = a_0 + \sum_{i=1}^n a^i \log v^i$. The order of polynomial varies slightly between different papers, for example Wang et al. (2006) set $n = 2$ while Jelić et al. (2008) sets $n = 3$.

One should be careful in choosing the order of the polynomial to perform the fitting. If the order of the polynomial is too small, the foregrounds will be under-fitted and the EoR signal could be dominated and corrupted by the fitting residuals. If the order of the polynomial is too big, the EoR signal could be fitted out. For this work we will follow Jelić et al. (2008) and perform the fitting in log space using a 3rd order polynomial.

2.1.2 CCA (Correlated Component Analysis)

In this section we describe the main principles of operation of the Fourier-domain Correlated Component Analysis (CCA) method. More details on the method and on its application to the HI signal can be found in Ricciardi et al. (2010) and Bonaldi & Brown (2015), respectively. The CCA is a “model learning” algorithm, which estimates the frequency spectrum of the foreground components from the data exploiting second-order statistics. This method was developed for the Cosmic Microwave Background (CMB); its ability to improve the modelling of the poorly known anomalous microwave emission has been demonstrated in Bonaldi et al. (2007), Bonaldi & Ricciardi (2012) and Planck Collaboration (2013).

We start by modelling the data in the uv plane as a linear mixture of the foreground components. For each point in the uv plane we write

$$\vec{x} = \mathbf{B}\vec{s} + \vec{n}. \quad (2.1)$$

The vectors \vec{x} and \vec{n} contain the data and the noise in Fourier space, respectively; the vector \vec{s} contains the astrophysical foregrounds; the diagonal matrix \mathbf{B} contains the instrumental dirty beams

in Fourier space and the matrix \mathbf{A} , called the mixing matrix, contains the intensity of the foreground components at all frequencies. The 21-cm signal is modelled as a noise term, contributing to \vec{n} together with the instrumental noise.

The additional assumptions made by the CCA are that the mixing matrix is constant within the considered area of the sky, and that its unknown elements can be reduced by adopting a suitable parametrization $\mathbf{A} = \mathbf{A}(\vec{p})$. For example, in the following, a power law is assumed for the synchrotron component with unknown, spatially constant, spectral index. For the free-free, we adopt a power-law behaviour with fixed spectral index of -2.08. When necessary, we can adopt other parametric models having more degrees of freedom. Though CCA allows the data to estimate the parameters of the model, it does exploit a parametrization, and therefore it is classified as a parametric method.

Once we have an estimate of the mixing matrix, using a relation between the cross-spectra of the data, we can invert eq. (2.1) and reconstruct the foreground components as $\vec{\hat{s}} = \mathbf{W}\vec{x}$ directly in the Fourier domain.

The cleaning of the HI signal consists of subtracting the estimated foreground components $\vec{\hat{s}}$ at all frequencies. We perform the subtraction as:

$$\vec{x} - \mathbf{R}\hat{\mathbf{H}}\vec{\hat{s}} \quad (2.2)$$

where \mathbf{R} is a diagonal matrix whose elements are chosen to improve the subtraction by minimizing the power of the residuals at each frequency. This step compensates for small errors in the parametric model adopted by the CCA, which result in a slight over/underestimation of the amplitude of the foregrounds at a given frequency. The effectiveness of this approach is tested with the simulation R2 (see Sections 3.4 and 4.4). It is important to note that this minimization approach could be applied to all methods, and as such is a way of mitigating the weaknesses of the parametric method as opposed to the inherent ability for the non-parametric method to deal with foregrounds differing from our models.

2.2 Non-Parametric Methods

2.2.1 Wp smoothing

Wp smoothing is a technique, introduced to 21-cm work by Harker et al. (2009), to fit the foregrounds LOS-by-LOS. The aim is to directly exploit the physical expectation that the foregrounds are smooth, so in this sense the foreground separation is not completely blind. It does not, however, assume a specific parametric form for the foregrounds, or anything about their spatial structure. Wp penalises changes in curvature, with roughness measured ‘apart from inflection points (IPs)’, hence the name ‘Wendepunkt’ (Wp), the German word for ‘inflection point’.

In the i -th LOS, we have a set $\{(v_1^i, s_1^i), (v_2^i, s_2^i), \dots, (v_n^i, s_n^i)\}$ of observations in n frequency channels, which we wish to fit with a smooth function $f(v)$. Since Wp smoothing always applies to one LOS, from now on we will drop the superscript for clarity. Wp smoothing takes as its measure of roughness the integrated change of curvature. If κ is the radius of curvature, the standardized change of curvature is $\kappa'/\kappa \approx f'''/f''$, where the approximation, which we adopt here, holds exactly at local extrema ($f' = 0$) and becomes singular at IPs ($f'' = 0$).

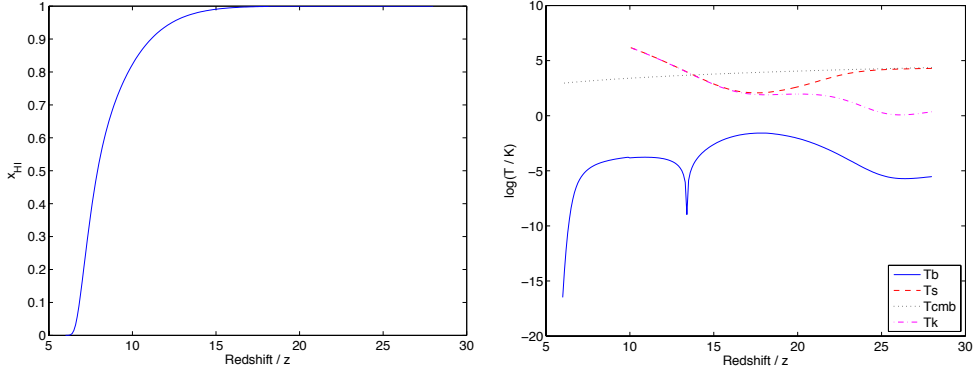


Figure 1: Left: The evolution of neutral hydrogen fraction with redshift. Right: The evolution of T_s , T_{cmb} , T_K and T_b with redshift. The absolute log is taken of T_b .

We therefore separate out the IPs, writing f'' as a smooth function $g(v)$ multiplied by a polynomial with the IPs as its roots. With the IPs specified, we find f by performing a penalised fit to the data, where the penalty term is given by a measure of the integrated change in curvature of g multiplied by a smoothing parameter, λ .

This formulation of Wp smoothing is given by Mächler (1993, 1995), who derived a boundary value problem, the solution of which is the function f we seek. Different algorithms have been proposed to solve this system (Mächler 1989; Gu et al. 2013), but we use that outlined by Harker et al. (2009). Unfortunately, these methods currently take ~ 1 s to solve for a single sightline, depending on the value of λ , making Wp smoothing relatively slow for large data cubes.

In principle, the choice of a value for λ should be determined by the level of smoothness we expect in our foregrounds. In the limit of $\lambda \rightarrow 0$, f becomes the best-fitting function with the given inflection points, while for $\lambda \rightarrow \infty$ it becomes the best-fitting polynomial of degree $n_w + 2$. Here, we fix $n_w = 0$ and choose λ based on the performance of the method in simulations. The quality of the fit is quite insensitive to the value of λ up to at least a factor of 2, however.

2.2.2 GMCA

Blind source separation (BSS) uses a mixing model $\vec{x} = \mathbf{A}\vec{s} + \vec{n}$, where \vec{x} is the observed data \mathbf{A} is the mixing matrix, \vec{s} is the unmixed data components and \vec{n} is the noise. What defines a BSS problem is the need to estimate both \mathbf{A} and \vec{s} with no prior knowledge of either (note the difference to CCA, where a prior form for \mathbf{A} was assumed). Methods differ in their approach to this estimation with, for example, the independent component analysis technique FastICA (Hyvärinen (1999), Hyvärinen et al. (2001) and applied to EoR data by Chapman et al. (2012)) assuming statistical independence of the components \vec{s} . Here we utilise another BSS technique, Generalized Morphological Component Analysis (GMCA), which assumes morphological diversity and sparsity of the foregrounds in order to model them. This approach originated with Zibulevsky & Pearlmutter (2001) who suggested that one could find a basis set in which the components to be found would be sparsely represented, i.e. a basis set where only a few of the coefficients would be non-zero. With the components being unlikely to have the same few non-zero coefficients one could then use this sparsity to more easily separate the mixture. We attempt to recover the cosmological signal as a residual of the process, i.e. it is actually part of \vec{n} . We can expand the data \vec{x} in a

wavelet basis and seek an unmixing scheme, through the estimation of \mathbf{A} , which yields the sparsest components \vec{s} in the wavelet domain.

For more technical details about GMCA, we refer the interested reader to Bobin et al. (2007, 2008a,b, 2013), where it is shown that sparsity, as used in GMCA, allows for a more precise estimation of the mixing matrix \mathbf{A} and more robustness to noise than ICA-based techniques such as FastICA. For a previous application of GMCA to EoR data see Chapman et al. (2013).

This component separation method has been applied to the Planck PR1 data to estimate a low-foreground CMB map (Bobin et al. 2014). In this context, sparsity is well adapted to remove naturally non-Gaussian and heterogeneous components such as foregrounds.

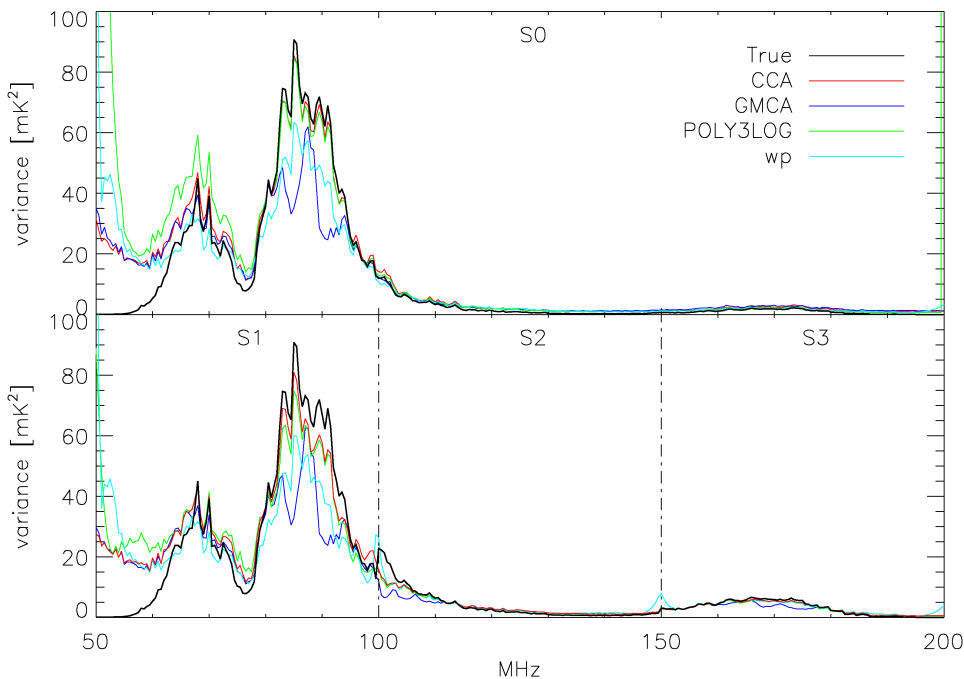


Figure 2: Variance of the simulated cosmological signal (black line) and of the reconstructed cosmological signals for the 4 foreground removal methods (coloured lines) for the S0 cube (top) and the S1, S2 and S3 cubes (bottom).

2.3 Foreground Avoidance

While the methods so far introduced have been focused on removing the foreground from the total signal, there has recently been discussion of avoiding the foregrounds instead. The coupling between the expected frequency smoothness of foregrounds and the unavoidable frequency dependent response of an interferometer leaves a characteristic footprint in k -space, separating an area over which foreground dominates, from a region which is virtually foreground free — a so-called ‘EoR window’. The boundaries of this EoR window have been discussed at length in the literature (Datta et al. 2010; Vedantham et al. 2012; Morales et al. 2012; Trott et al. 2012; Parsons et al. 2012; Liu et al. 2014a; Liu et al. 2014b) as well as seen in early results from low frequency observations

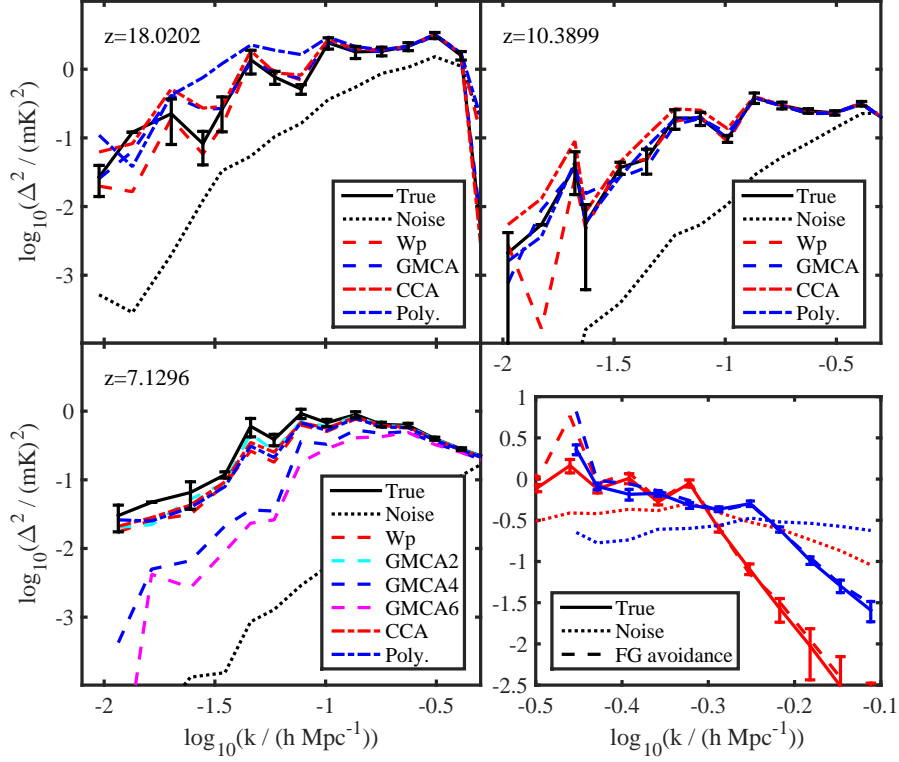


Figure 3: Spherically averaged power spectrum of the simulated cosmological signal, reconstructed cosmological signal, and noise. The top left, top right and bottom left panels show, respectively, power spectra computed in an 8 MHz slice in the centre of S1, S2 and S3. The central redshift of each slice is shown in the panel. Different line styles and colours show the true (input) signal, the noise, and the reconstructed signal for the four different foreground removal methods, as described in the legend. The axes are the same in each of these panels. Note that we show two extra GMCA lines according to foreground models with 2 and 6 components in the bottom left panel. We do not show the highest k points since they are strongly affected by the point spread function. The bottom right panel shows the application of foreground avoidance. Here, the different line styles show the true signal, the noise and the recovered signal, with the different colours used for the data at different redshifts (red for S2, blue for S3), with cutoffs in k_{10s} as described in the text. The result for S1 is poorer, and is not shown in order to avoid compressing the scale of the plot; note that the scale in this panel is different from the other three. In all cases, the power spectra are computed after applying a Hanning taper in the frequency direction to avoid ringing: this is particularly important for foreground avoidance, since it mitigates the aliasing of unsubtracted foreground power to high k_{10s} .

(Bernardi et al. 2013; Dillon et al. 2014; Pober et al. 2013). The largest extent of the EoR window happens to be at low k_{perp} where the frequency dependent response of the instrument is smooth, excluding the low k_{10s} mode that are likely to be dominated by foregrounds. When constructing other statistics such as a spherical power spectrum, one can then simply ignore all the foreground-dominated modes and use the k modes inside the EoR window alone. While this will result also in the loss of any cosmological signal outside the EoR window, the remaining region should be free of foreground contamination, if the foregrounds are indeed so well-defined even in the face of instrumental calibration errors. In comparison, while foreground removal allows the possibility of recovering the cosmological signal in all modes, there is also the possibility of foreground fitting

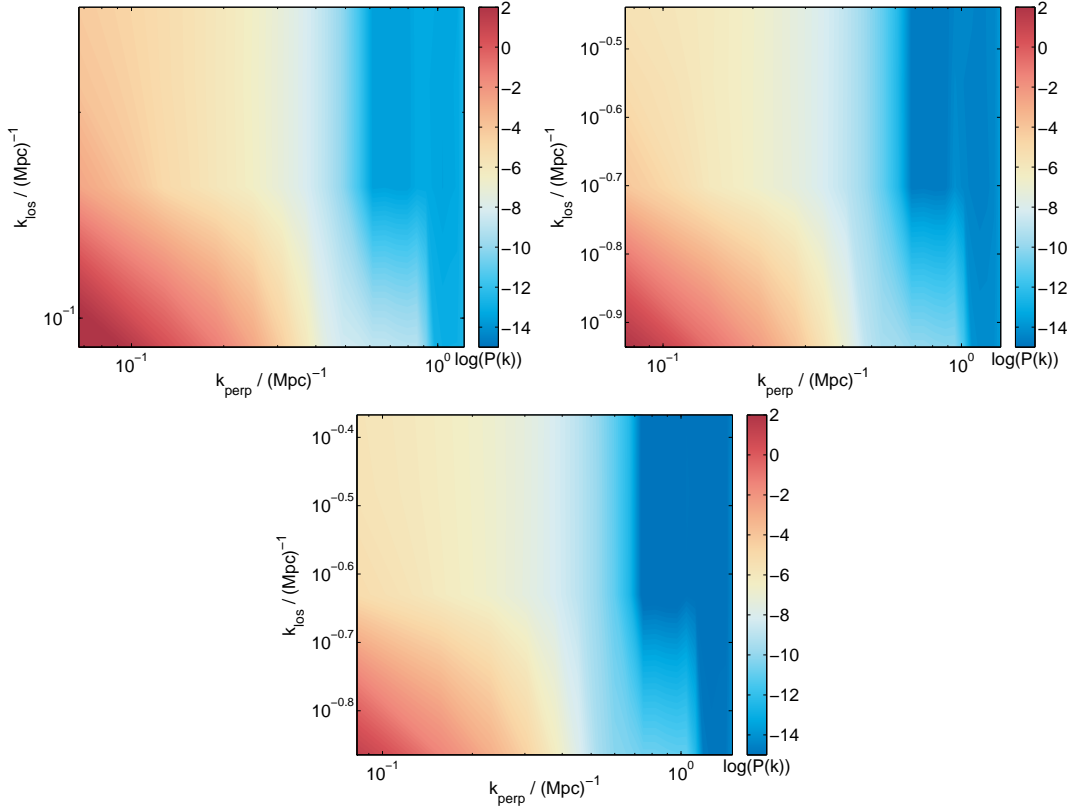


Figure 4: Cylindrical power spectrum of the S0 cube at 75 MHz, 125 MHz and 175 MHz (in reading order). The foreground contamination can be clearly seen for $k_{\text{los}} < 10^{-0.93}$, $k_{\text{los}} < 10^{-0.7}$ and $k_{\text{los}} < 10^{-0.63} \text{ Mpc}^{-1}$ respectively, along with the action of the PSF at high k_{perp} .

bias being introduced on all modes.

3. SKA Phase 1 Simulations

3.1 Cosmological Signal

We simulate the cosmological signal using the semi-numerical reionization code SIMFAST21 (Santos et al. 2010).

We create initial conditions boxes on a grid of 1024^3 cells before producing ionization and brightness temperature boxes on a grid of 256^3 . The boxes are a constant 1 Gpc in side length and are output between redshifts 6 and 28 at separations of $dz = 0.1$. As the aim of this chapter is to assess the effectiveness of foreground removal methods, as opposed to studying a particular model of reionization, we have simply used the default options for SIMFAST21, for example a minimum halo mass of $1.0e8 M_{\odot}$. When calculating the brightness temperature, T_b , it is usual to assume that $T_s \gg T_{\text{cmb}}$ where T_s is the spin temperature and T_{cmb} is the CMB temperature. However, this assumption breaks down at high redshift due to the increasing effect of the gas temperature (T_K) and Lyman-alpha coupling on the spin temperature. We calculate the full spin temperature for all

redshifts above 10. We plot the evolution of the neutral hydrogen fraction, x_{HI} , and T_s, T_{cmb}, T_K and T_b in Fig. 1.

From the real space T_b boxes we create an observation cube, or ‘light cone’, with the LOS axis evolving in redshift and a constant 5 degree field of view.

3.2 Foregrounds

Though there have been foreground observations at frequencies relevant to LOFAR using WSRT (Bernardi et al. 2009; Bernardi et al. 2010) the foreground contamination at the frequencies and resolution of LOFAR remains poorly constrained. As a result, foreground models directly relevant for this paper rely on using constraints from observations at different frequency and resolution ranges. These constraints are used to normalize the necessary extrapolations made from observations to create a model relevant for LOFAR-EoR observations.

In general, the foreground components are modelled as power laws in 3+1 dimensions (i.e. three spatial and frequency) such that $T_b \propto \nu^\beta$.

The foreground simulations used in this paper are obtained using the foreground models described in Jelić et al. (2008); Jelić et al. (2010). The foreground contributions considered in these simulations are Galactic synchrotron emission, Galactic free-free emission and extragalactic foregrounds.

3.3 Instrumental Effects

The instrumental effects were modelled using the OSKAR simulator¹ and a list of preliminary station positions for the SKA1-LOW.

The images of the SKA-low PSF were produced by assuming full correlation between all 866 core stations, where the maximum baseline length is 5.29 km. The telescope was located at 52.7 degrees latitude. Baseline coordinates were generated for a 12-hour synthesis observation, with a phase centre on the sky at apparent equatorial coordinates $(\alpha, \delta) = (218, 34.5)$ degrees. A 5-minute sampling interval was used to give 144 snapshots each of 374545 baselines. PSF images were then generated in CASA across a 5-degree field-of-view using 256 w-projection planes. The noise is then normalized according to a 1000 hour integration time using the prescription described in e.g. Thompson et al. (2001).

3.4 Cubes for Analysis

In order to simulate an observation, one normally constructs a ‘dirty’ cube whereby the cosmological signal and foregrounds are convolved with the same frequency-dependent PSF used to construct the instrumental noise. However, the standard foreground removal methods require the channels to have common resolution in order to work optimally. In which case we use five cubes, all with channels separated by 0.5 MHz, in our analysis:

- S0: A cube running from 50–200 MHz consisting of the clean foregrounds and cosmological signal convolved with the PSF at 50 MHz and the instrumental noise constructed with a sampling equivalent to the 50 MHz sampling in each channel.

¹<http://www.oerc.ox.ac.uk/ska/oskar>

- S1: A cube running from 50–99.5 MHz consisting of the clean foregrounds and cosmological signal convolved with the PSF at 50 MHz and the instrumental noise constructed with a sampling equivalent to the 50 MHz sampling in each channel.
- S2: A cube running from 100–149.5 MHz consisting of the clean foregrounds and cosmological signal convolved with the PSF at 100 MHz and the instrumental noise constructed with a sampling equivalent to the 100 MHz sampling in each channel.
- S3: A cube running from 150–200 MHz consisting of the clean foregrounds and cosmological signal convolved with the PSF at 150 MHz and the instrumental noise constructed with a sampling equivalent to the 150 MHz sampling in each channel.
- R2: We multiply each channel of the clean foreground cube by a random number drawn from a Gaussian distribution with standard deviation of 0.05, simulating a 5% random wiggle along the line of sight. This foreground cube is then convolved and used to construct a cube as described in S2.

The adjustment of the entire frequency range to a common resolution in S_0 results in the loss of a lot of high resolution information at high frequency. There is a balance to be made between the amount of information lost and the amount of data the methods need to provide an optimal foreground estimate. We therefore test S1, S2 and S3 in order to assess whether the methods are able to make good signal recoveries of the higher resolution information at higher frequencies, despite a third of the data being available to constrain the foregrounds.

We construct R2 (R - rough) in order to test the reliance of the methods on the smoothness of the foregrounds. We expect methods with strong constraints on the smoothness such as polynomial, CCA and, to some degree, W_p smoothing to be more affected than GMCA which places no explicit prior on the smoothness. This roughness could be interpreted as an inherent roughness of the foregrounds themselves or as a simple approximation of an instrumental calibration error such a leakage of polarized foregrounds.

4. Results

4.1 Variance

We computed the variance of the true input cosmological signal and of the reconstructed cosmological signals for each of the 4 foreground removal methods. The variance has been computed for a pixel size of 2.3 arcmins. Given the low noise, the results are stable for changes of the pixel size. The results are shown in Fig. 2 for both the S_0 cube and the collated S1, S2 and S3 cubes.

All methods show an excess variance at $\nu < 60$ MHz which is due to foreground residuals. The results are good for the other frequencies. GMCA and W_p smoothing somewhat underestimate the variance in the 80–90 MHz frequency range. For all methods the variance is correctly recovered over a broad frequency range.

There is not an apparent major disadvantage to any of the methods by splitting the cube into three segments and so we pursue analysis of only the S1, S2 and S3 cubes in the following, in

order to retain as much spatial information as possible while conforming to the common resolution requirements of the methods.

One might wonder at the two peaks in variance below 100 MHz. While the clean signal does not show such clear peaks, the action of the realistic SKA PSF on the clean cosmological signal induces this effect.

4.2 Power Spectrum

The spherically-averaged 21-cm power spectrum is one of the quantities most readily computed from theory and is a rich source of information on cosmology and on the nature of high-redshift sources (e.g. Barkana & Loeb 2005, and many subsequent works). It is expected that the SKA will be able measure it with high signal-to-noise across a broad range of scales. Moreover, k -space might be considered a natural space in which to study the 21-cm signal, since interferometers natively measure Fourier modes in the plane of the sky, while smooth foregrounds are likely to be more isolated in k -space (residing mainly at low k_{los}) than in real space. The quality of power spectrum recovery has therefore become a popular metric when comparing foreground removal methods as well as instruments.

Fig. 3 shows the spherically averaged power spectrum for a slice in frequency in the centre of S1, S2 and S3. In a given frequency range, this is computed simply by finding the power (magnitude squared of the visibility) in each cell in Fourier space, and then binning this power in spherical shells, i.e. shells with a given $k = \sqrt{k_{\text{perp}}^2 + k_{\text{los}}^2}$. Following convention, we plot $\Delta^2(k) = k^3 P(k) / 2\pi^2$, which traces the contribution to the variance of a unit interval in $\log(k)$ at k .

In each case, the noise power spectrum has been subtracted from the residual power to recover an estimate of the cosmological signal power. Foreground removal mostly yields a reasonable estimate of the cosmological signal, with the possible exception of GMCA in the lowest-redshift slice (in S3). It is not immediately clear why GMCA as opposed to any other method would not perform as well in this frequency segment. The fiducial four-component model of GMCA was chosen rather arbitrarily and ideally all methods should undergo a full Bayesian model selection in order to select the various input parameters (such as the smoothing parameter in Wp smoothing and number of components in the GMCA foreground model). In the S3 panel we have also plotted the power spectrum for GMCA with two and six components in the foreground model. We can see that by changing the number of components in the GMCA foreground model we can achieve a similar fit to the other methods which is perhaps indicative that a more robust method of model selection is needed. It is also worth remembering that the CCA results have undergone an in-built residual minimization which allow it to perform much better than if the minimization were not carried out. This minimization could equally be applied to any method.

We also include in Fig. 3 the result of implementing foreground avoidance. By constructing a cylindrical power spectrum in $k_{\text{perp}}, k_{\text{los}}$ we can define a k_{los} bound, below which modes are considered contaminated by foregrounds and above which lies what is termed the ‘EoR window’. We can then construct a spherical power spectrum as described above but ignoring all modes below this bound. This is the ‘foreground avoidance’ line. We find this bound to be at $k_{\text{los}} < 10^{-0.93}$, $k_{\text{los}} < 10^{-0.7}$ and $k_{\text{los}} < 10^{-0.63} \text{ Mpc}^{-1}$ for the data at 75, 125 and 175 MHz respectively (see Fig. 4). As one can see, this severely limits the range of scales which can be recovered; however, there are

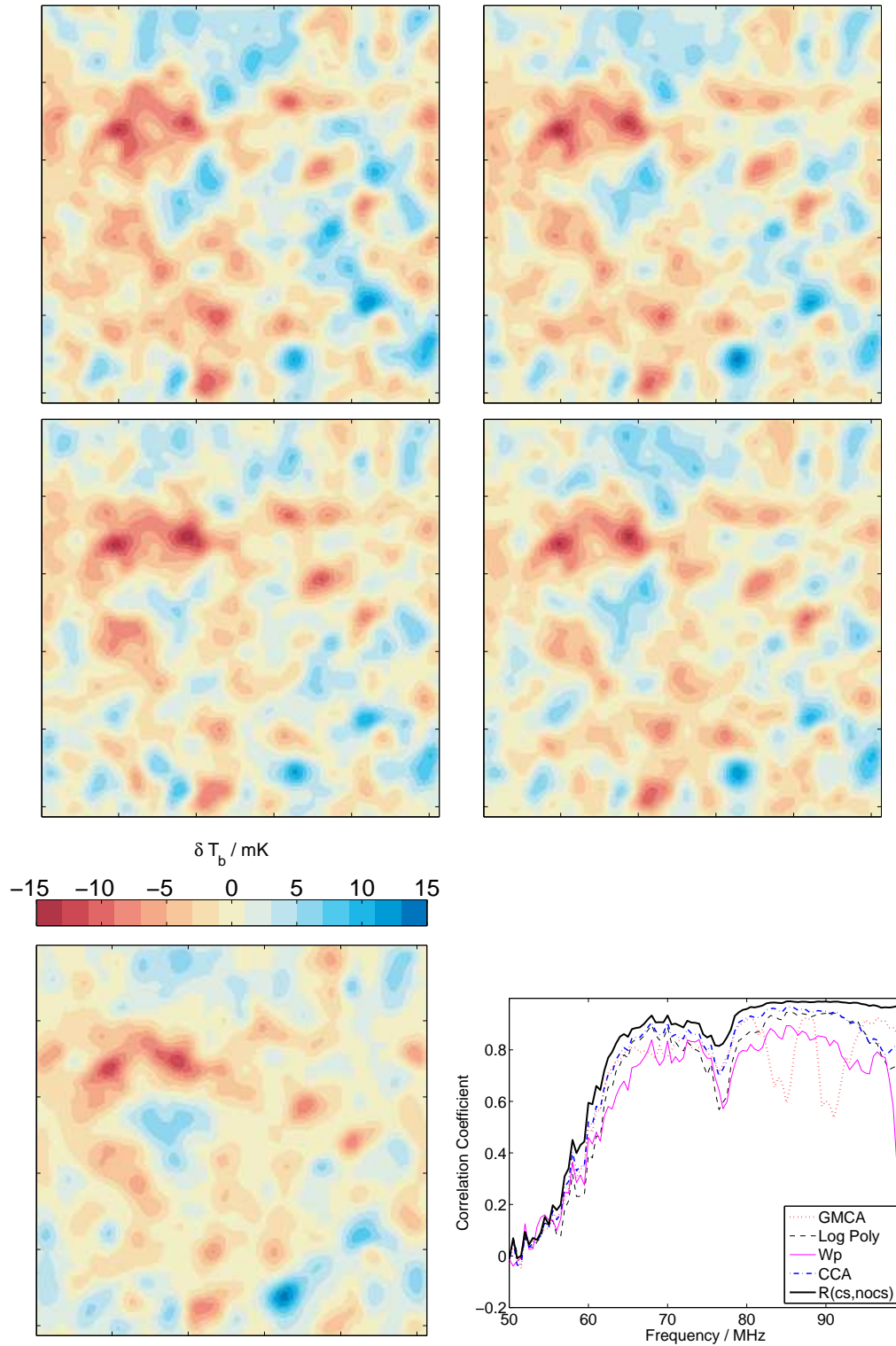


Figure 5: Top two rows, reading order: the smoothed residual maps of S1 at 75 MHz from the Polynomial, CCA, Wp and GMCA methods. Bottom row, left to right: The smoothed cosmological signal at 75 MHz and the correlation coefficient relating to residuals vs. cosmological signal. The best theoretical recovery possible, whereby foreground fitting errors are zero, is shown by the correlation between noise plus simulated cosmological signal vs. simulated cosmological signal in solid black.

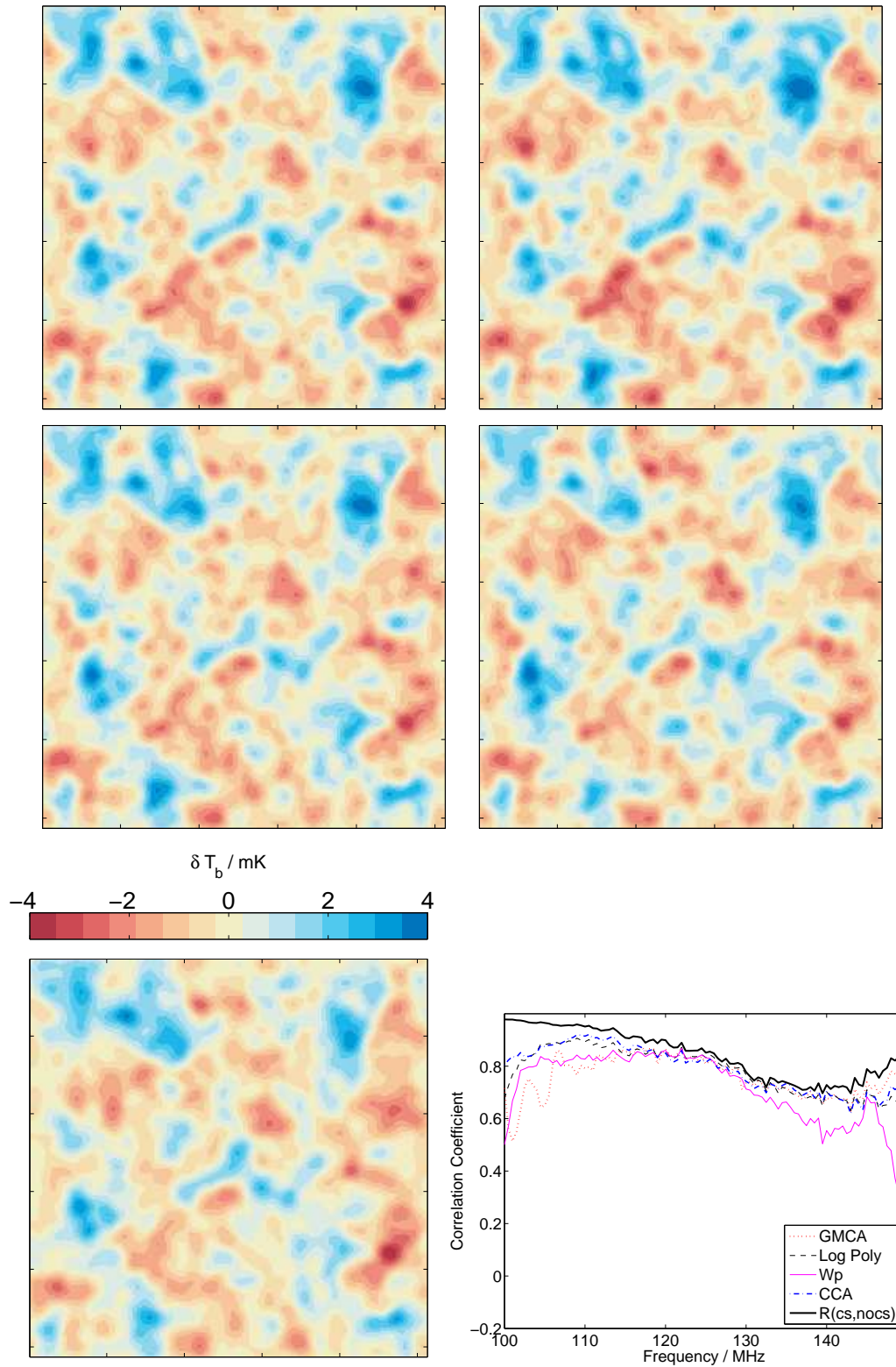


Figure 6: Top two rows, reading order: the smoothed residual maps of S2 at 125 MHz from the Polynomial, CCA, Wp and GMCA methods. Bottom row, left to right: The smoothed cosmological signal at 125 MHz and the correlation coefficient relating to residuals vs. cosmological signal.

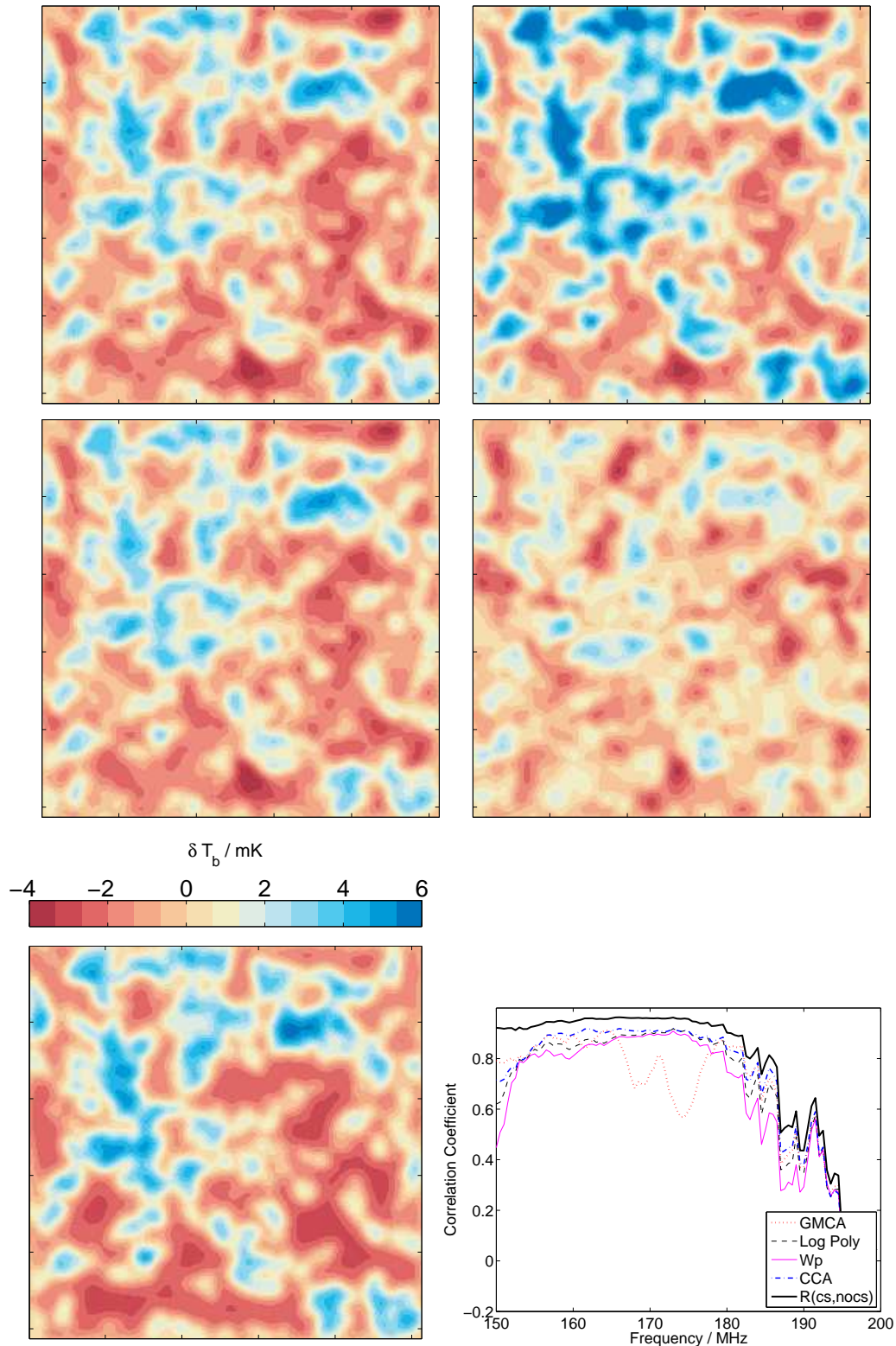


Figure 7: Top two rows, reading order: the smoothed residual maps of S3 at 175 MHz from the Polynomial, CCA, Wp and GMCA methods. Bottom row, left to right: The smoothed cosmological signal at 175 MHz and the correlation coefficient relating to residuals vs. cosmological signal.

values of z and k for which it performs very well. An optimal power spectrum estimation strategy will likely combine removal and avoidance in some way, with some attempts in this direction having been made by Liu et al. (2014b).

4.3 Images

One of the most exciting scientific outcomes of the SKA is the ability to image the EoR and Cosmic Dawn. We now review how the foreground removal methods affect this capability. We present slices from the residual cubes for three different scenarios. In Fig. 5 we take S1 and show the residual slices at 75 MHz; in Fig. 6 we take S2 and show the residual slices at 125 MHz and in Fig. 7 we take S3 and show the residual slices at 175 MHz. Note that, for all images shown, the residual cube has been smoothed with a Gaussian kernel of FWHM eight pixels, which is equivalent to 9.36 arcminutes, in order to mitigate the effect of the instrumental noise. In each figure we also show the correlation coefficient between the smoothed residual cube and the smoothed simulated cosmological signal for the different methods. As we will only have statistical knowledge of the noise, we would not be well motivated in correlating the reconstructed and simulated cosmological signal, and instead also plot an ‘envelope’ in the form of a correlation between the simulated cosmological signal and the simulated cosmological signal combined with the instrumental noise. This provides an upper bound for the best correlation we can expect to see in the data if zero foreground fitting errors were present. We see that an impressive image recovery is apparent for all methods, though the extent of that recovery is highly variable with frequency.

4.4 Relaxation of foreground smoothness

We now demonstrate how the methods fare when analysing a cube containing foregrounds which have a random 5% deviation from the smooth power law along the line of sight. We show the correlation coefficient between the residuals of cube R2 and the cosmological signal in the top-left panel of Fig. 8. It is interesting to see the failure of the polynomial method compared to the ability of CCA to recover from a similar failure using the correction method mentioned in Section 2.1.2. This correction method could be applied to all approaches and relies on the first order approximation (i.e. that the wiggle is superimposed on a power law) of the foreground being accurate enough. The crucial point to take away is that non-parametric methods, such as GMCA, which do not have a prior on the foreground smoothness, are able to model the foreground accurately with no extra modelling input by the user. In comparison, the parametric (and indeed non-parametric methods like Wp smoothing which require, as opposed to assume, smoothness of foregrounds) need some form of ‘extra modelling’ such as correction factors or tweaking of the fitting parameters. It is likely that Wp fitting would see a similar improvement to CCA with such correction factors, as the same first-order approximation of smoothness applies.

4.5 Other SKA configurations

We now look at how one of our results is affected in the case of an early-SKA implementation where the sensitivity is halved and an SKA2 implementation where the sensitivity is quadrupled. We do this for only one method, GMCA, for clarity and conciseness. We show the correlation coefficient between the recovered maps in the three remaining panels of Fig. 8.

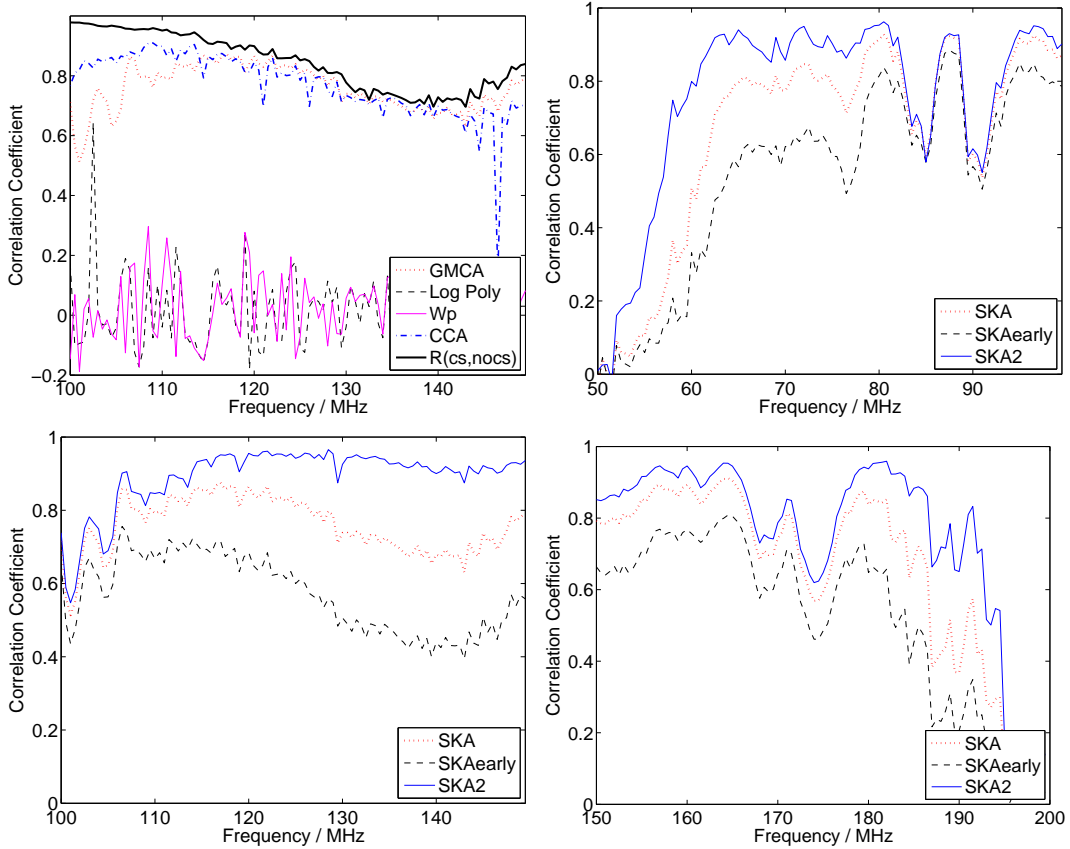


Figure 8: In reading order: The correlation coefficient relating to the R2 cube residuals and the cosmological signal. The correlation coefficient relating to the GMCA residuals for the different SKA scenarios vs. cosmological signal for the S1, S2 and S3 cubes.

As the instrumental noise level decreases, the recovery is greatly improved. The general shape of the curves remains similar, suggesting that there is significant contribution by foreground fitting errors which affect the correlation from slice to slice. There are indeed slices in the S1 cube, for example at 85 MHz, which produce the same correlation independent of noise level, suggesting foreground leakage is the dominant cause of error at those frequencies.

5. Conclusions

- We have applied a suite of foreground removal methods to a state-of-the-art simulation of SKA Phase 1 observations and analysed the recovered cosmological signal by using different statistics.
- The variance of the 21-cm signal is well recovered over a broad range of frequencies by all methods. While setting the entire bandwidth to a common resolution results in the loss of a lot of spatial information at high frequency, there is no obvious advantage in the accuracy of the recovered variance. We therefore opt for a compromise of setting several segments to common resolution.

- Foreground removal methods generally yield a reasonable estimate of the spherical power spectrum of the cosmological signal. Foreground avoidance severely limits the range of scales for which the power spectrum can be recovered, however in such a limited range the results are good.
- We obtain an impressive recovery of the images for all methods, the quality of which however varies with frequency.
- The relaxation of the hypothesis of foreground smoothness does not affect the performance of GMCA, which is non-parametric, while it affects polynomial fitting and W_p smoothing. However, as shown by the CCA, the quality of the results can be restored with some extra modelling, at least as long as the smooth model adopted is a reasonable approximation of the foreground spectrum.
- As the instrumental noise level decreases, the recovery of the signal is greatly improved. For some frequencies (85 MHz for example) the results are much more similar, which suggests that foreground subtraction errors in these cases are dominant.

6. Acknowledgments

EC would like to thank Jonathan Pritchard and the SKA-EoR working group for useful discussions. GH acknowledges funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007-2013) under REA Grant Agreement No 327999. AB acknowledges support from the European Research Council under the EC FP7 grant number 280127.

References

- Barkana, R., & Loeb, A. 2005, *ApJL*, 624, L65
- Bernardi, G., de Bruyn, A. G., Brentjens, M. A., et al. 2009, *A&A*, 500, 965
- Bernardi, G., de Bruyn, A. G., Harker, G., et al. 2010, *A&A*, 522, A67
- Bernardi, G., Greenhill, L. J., Mitchell, D. A., et al. 2013, *ApJ*, 771, 105
- Bobin, J., Moudden, Y., Starck, J.-L., Fadili, J., & Aghanim, N. 2008a, *Statistical Methodology*, 5, 307
- Bobin, J., Starck, J.-L., Fadili, J., & Moudden, Y. 2007, *IEEE Transactions on Image Processing*, 16, 2662
- Bobin, J., Starck, J.-L., Moudden, Y., & Fadili, M. J. 2008b, in "Advances in Imaging and Electron Physics", 152, 221
- Bobin, J., Starck, J.-L., Sureau, F., & Basak, S. 2013, *A&A*, 550, A73
- Bobin, J., Sureau, F., Starck, J.-L., Rassat, A., & Paykari, P. 2014, *A&A*, 563
- Bonaldi, A., & Brown, M. L. 2015, *Monthly Notices of the Royal Astronomical Society*, 447, 1973
- Bonaldi, A., & Ricciardi, S. 2012, *Advances in Astronomy*
- Bonaldi, A., Ricciardi, S., Leach, S., et al. 2007, *MNRAS*, 382, 1791
- Bowman, J. D., Morales, M. F., & Hewitt, J. N. 2006, *ApJ*, 638, 20

- Chapman, E., Abdalla, F., Harker, G., et al. 2012, MNRAS, 423, 2518
- Chapman, E., Abdalla, F. B., Bobin, J., et al. 2013, MNRAS, 429, 165
- Chapman, E., Zaroubi, S., & Abdalla, F. 2014, preprint (astro-ph/1408.4695)
- Datta, A., Bowman, J. D., & Carilli, C. L. 2010, ApJ, 724, 526
- Di Matteo, T., Perna, R., Abel, T., & Rees, M. 2002, ApJ, 564, 576
- Di Matteo, T. and Ciardi, B. and Miniati, F. 2004, MNRAS, 355, 1053
- Dillon, J. S., Liu, A., Williams, C. L., et al. 2014, Phys. Rev. D, 89, 023002
- Gleser, L., Nusser, A., & Benson, A. J. 2008, MNRAS, 391, 383
- Gnedin, N. Y., & Shaver, P. A. 2004, ApJ, 608, 611
- Gu, J., Xu, H., Wang, J., An, T., & Chen, W. 2013, ApJ, 773, 38
- Harker, G., Zaroubi, S., Bernardi, G., et al. 2009, MNRAS, 397, 1138
- Hyvärinen, A. 1999, Neural Networks, IEEE Transactions on, 10, 626
- Hyvärinen, A., Karhunen, J., & Oja, E. 2001, *Independent Component Analysis* (John Wiley and Sons)
- Jelić, V., Zaroubi, S., Labropoulos, P., et al. 2010, MNRAS, 409, 1647
- Jelić, V., Zaroubi, S., Labropoulos, P., et al. 2008, MNRAS, 389, 1319
- Liu, A., Parsons, A. R., & Trott, C. M. 2014a, Phys. Rev. D, 90, 023018
- Liu, A., Parsons, A. R., & Trott, C. M. 2014b, Phys. Rev. D, 90, 023019
- Liu, A., & Tegmark, M. 2011, Phys. Rev. D, 83, 103006
- Liu, A., Tegmark, M., Bowman, J., Hewitt, J., & Zaldarriaga, M. 2009, MNRAS, 398, 401
- Mächler, M. 1989, PhD thesis, ETH Zürich
- . 1993, *Very smooth nonparametric curve estimation by penalizing change of curvature*, Research report 71, Seminar für Statistik ETH Zürich
- . 1995, Annals of Statistics, 23, 1496
- McQuinn, M., Zahn, O., Zaldarriaga, M., Hernquist, L., & Furlanetto, S. R. 2006, ApJ, 653, 815
- Morales, M. F., Bowman, J. D., & Hewitt, J. N. 2006, ApJ, 648, 767
- Morales, M. F., Hazelton, B., Sullivan, I., & Beardsley, A. 2012, ApJ, 752, 137
- Oh, S. P., & Mack, K. J. 2003, MNRAS, 346, 871
- Parsons, A. R., Pober, J. C., Aguirre, J. E., et al. 2012, ApJ, 756, 165
- Patil, A. H., Zaroubi, S., Chapman, E., et al. 2014, ArXiv e-prints
- Petrovic, N., & Oh, S. P. 2011, MNRAS, 413, 2103
- Planck Collaboration. 2013, A&A, 557, A53
- Pober, J. C., Parsons, A. R., Aguirre, J. E., et al. 2013, ApJL, 768, L36
- Ricciardi, S., Bonaldi, A., Natoli, P., et al. 2010, MNRAS, 406, 1644
- Santos, M. G., Cooray, A., & Knox, L. 2005, ApJ, 625, 575
- Santos, M. G., Ferramacho, L., Silva, M. B., Amblard, A., & Cooray, A. 2010, MNRAS, 406, 2421
- Thompson, A. R., Moran, J. M., & Swenson, Jr., G. W. 2001, *Interferometry and Synthesis in Radio Astronomy* (John Wiley and Sons)
- Trott, C. M., Wayth, R. B., & Tingay, S. J. 2012, ApJ, 757, 101
- Vedantham, H., Shankar, N. U., & Subrahmanyam, R. 2012, ApJ, 745, 176
- Wang, X., M. Tegmark, Santos, M. G., & Knox, L. 2006, ApJ, 650, 529
- Zaldarriaga, M., Furlanetto, S. R., & Hernquist, L. 2004, ApJ, 608, 622
- Zibulevsky, M., & Pearlmutter, B. A. 2001, Neural Computation, 13, 863