Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **508,** 3551–3568 (2021)
Advance Access publication 2021 October 5

https://doi.org/10.1093/mnras/stab2855

Cleaning foregrounds from single-dish 21 cm intensity maps with Kernel principal component analysis

Melis O. Irfan[1,2]★ and Philip Bull[1,2]

[1]*Department of Physics and Astronomy, University of Western Cape, Cape Town 7535, South Africa*
[2]*Astronomy Unit, Queen Mary University of London, Mile End Road, London E1 4NS, UK*

Accepted 2021 September 29. Received 2021 September 28; in original form 2021 July 5

ABSTRACT

The high dynamic range between contaminating foreground emission and the fluctuating 21 cm brightness temperature field is one of the most problematic characteristics of 21 cm intensity mapping data. While these components would ordinarily have distinctive frequency spectra, making it relatively easy to separate them, instrumental effects and calibration errors further complicate matters by modulating and mixing them together. A popular class of foreground cleaning method are unsupervised techniques related to principal component analysis (PCA), which exploit the different shapes and amplitudes of each component's contribution to the covariance of the data in order to segregate the signals. These methods have been shown to be effective at removing foregrounds, while also unavoidably filtering out some of the 21 cm signal too. In this paper we examine, for the first time in the context of 21 cm intensity mapping, a generalized method called Kernel PCA, which instead operates on the covariance of non-linear transformations of the data. This allows more flexible functional bases to be constructed, in principle allowing a cleaner separation between foregrounds and the 21 cm signal to be found. We show that Kernel PCA is effective when applied to simulated single-dish (auto-correlation) 21 cm data under a variety of assumptions about foregrounds models, instrumental effects etc. It presents a different set of behaviours to PCA, e.g. in terms of sensitivity to the data resolution and smoothing scale, outperforming it on intermediate to large scales in most scenarios.

Key words: methods: data analysis – methods: statistical – large-scale structure of Universe – cosmology: observations – radio lines: galaxies.

1 INTRODUCTION

Forthcoming 21 cm intensity mapping surveys promise to survey extremely large volumes of the Universe over a broad span of redshifts with a comparatively high survey speed. Modern radio telescopes have sufficient raw sensitivity to enable this, but a firm detection of the cosmological 21 cm signal remains elusive – while definitive detections have been made in cross-correlation with optical galaxy surveys (Chang et al. 2010; Wolz et al. 2017a; Anderson et al. 2018; Wolz et al. 2021), it has not yet been possible to obtain a reliable measurement of the 21 cm auto-power spectrum, or the baryon acoustic oscillation feature that is their principal scientific target (Chang et al. 2008; Bull et al. 2015). The main reason for this boils down to the very large dynamic range of the data, with foreground contamination around 3–4 orders of magnitude larger than the cosmological signal.

While foregrounds are in principle distinguishable from the rapidly spatially and spectrally varying 21 cm fluctuations by their smooth frequency spectra, approaching a power-law behaviour characteristic of Galactic and extragalactic synchrotron emission. This is significantly complicated by instrumental effects and calibration errors however, which modulate and scatter the spectrally smooth foregrounds into other regions of Fourier space, thus contaminating what would otherwise be 'clean' signal-dominated modes. The detailed behaviours of these effects can depend heavily on the particular instrument in question, but some reasonably generic issues can be identified. For example, outside the main lobe of the primary beam pattern of the telescope, near zero-crossings (nulls) tend to arise, with their exact angular distance from the beam centre depending on frequency. Bright sources close to the null at one frequency may therefore pass through it at another frequency, strongly modulating the otherwise smooth spectrum of the source (Ghosh et al. 2011). Similarly, the leakage of polarized foreground emission into the unpolarized channel due to beam and receiver imperfections can also generate additional spectral structure; polarized emission is inherently more spectrally complex due to Faraday rotation and other line-of-sight effects, and this can be exacerbated by the additional spectral structure of the beam imperfections themselves (Alonso, Ferreira & Santos 2014; Shaw et al. 2015).

The most common approach to dealing with the foreground contamination issue is to apply a so-called 'blind' foreground subtraction algorithm to the data (Wolz et al. 2014; Alonso et al. 2015; Olivari, Remazeilles & Dickinson 2016; Carucci, Irfan & Bobin 2020; Cunnington et al. 2021). These methods do not require a model of the data to be specified; instead, they perform transforms of various kinds in order to try and segregate the foreground information

★ E-mail: melis.irfan@qmul.ac.uk

© 2021 The Author(s)
Published by Oxford University Press on behalf of Royal Astronomical Society

Downloaded from https://academic.oup.com/mnras/article/508/3/3551/6381718 by guest on 16 April 2024

from the cosmological 21 cm signal (and noise, which is also rapidly fluctuating). A broad class of such algorithms is based on principal component analysis (PCA), which calculates the empirical frequency-frequency covariance matrix of the data (i.e. by averaging over angular pixels), and then performs an eigendecomposition to arrive at a set of distinct modes that can, in principle, be segregated according to their signal-to-noise ratio. Since the foreground emission is vastly brighter than the 21 cm signal, this should allow the foregrounds to be filtered out simply by discarding the modes with the largest eigenvalues (i.e. the biggest signal to noise ratio (SNR)). A number of other unsupervised foreground removal methods use a similar filtering approach to PCA while using alternative means to arrive at a set of appropriate basis functions. For example, Independent Component Analysis (e.g. ICA; Wolz et al. 2014), uses a cost function that maximizes the non-Gaussianity of the recovered modes in addition to a eigenbasis decomposition. In principle, this allows the non-Gaussian-distributed foregrounds to be more cleanly separated from the approximately Gaussian 21 cm signal and noise. Conversely, Gaussian Progress Regression (GPR; Mertens, Ghosh & Koopmans 2018; Ghosh et al. 2020; Kern & Liu 2021; Soares et al. 2021a), which is a non-parametric Bayesian method, chooses to encapsulate the 21 cm contaminants as a combination of Gaussian variables so as to capture stochastic contaminates, such as calibration errors and contributions from the ionosphere. Other methods perform the component separation in different bases in which a cleaner separation of signal and foreground is possible, such as Generalized Morphological Component Analysis (GMCA; Bobin et al. 2008; Chapman et al. 2013; Carucci et al. 2020), which uses sparsity to enforce the separation of the different components, and Generalized Needlet Internal Linear Combination (GNILC; Remazeilles, Delabrouille & Cardoso 2011; Olivari et al. 2016), which performs the eigendecomposition in the needlet domain.

These methods are largely successful at removing the bulk of the foreground emission, but at a cost. The high-SNR foreground modes tend to have a similar spectral structure to the longest-wavelength radial Fourier modes (i.e. those at lowest $k_{\parallel}$) in the 21 cm signal. The signal modes are therefore filtered out as well, leading to suppression or even total loss of the cosmological 21 cm signal on some range of scales (Chang et al. 2010; Alonso et al. 2015). Similar over-fitting also affects the noise. Because instrumental and other systematic effects scatter the foregrounds across a broader region of Fourier space, the foreground filters must use an increasing number of modes or more complex basis functions to be able to remove the foregrounds well, leading to progressively larger signal loss that affects an even wider range of scales. In realistic applications, some methods may require tens of modes to be subtracted from data with only a couple of hundred frequency channels, resulting in substantial signal loss even at relatively high k values (e.g. Wolz et al. 2017a; Wolz et al. 2021). This is often corrected for by performing simulated signal injection on the data to identify which (signal) Fourier modes are being suppressed and to what level. This information can then be used to construct a 'transfer function' that corrects the measured power spectrum (Masui et al. 2013; Switzer et al. 2013; Wolz et al. 2021). The effectiveness of this procedure depends to some extent on the realism of the injected simulations and how well they reproduce instrumental filtering effects.

For cross-correlation analyses, somewhat less aggressive foreground filtering can be applied. Since the foregrounds are in principle uncorrelated with the non-21 cm survey, e.g. an optical galaxy survey, they will tend to average out in the cross-correlation, and so a moderate foreground residual can be left in the 21 cm data, reducing the amount of signal loss due to filtering (Pourtsidou et al. 2016; Cunnington et al. 2019). Leaving too large a residual will impact the noise properties of the cross-correlation however, which goes approximately like the product of the square roots of the noise variances of the individual surveys (e.g. Villaescusa-Navarro et al. 2015), where in this case the foreground residual would be included as part of the 21 cm survey's noise variance.

In this paper, we study an extension of PCA filtering to a broader class of algorithms called Kernel PCA (KPCA; Schölkopf, Smola & Müller 1997). These allow arbitrary sets of non-linear transformations to be applied to the data, making it possible to construct more complicated decompositions of the frequency-frequency covariance and therefore – in principle – provide a cleaner separation of the foregrounds and 21 cm signal. KPCA methods are commonly used in other domains for segmentation and classification, where 'decision boundaries' between different sub-populations of the data are not well-described by linear functions of the data dimensions. The choice of kernel function determines which kinds of nonlinear transformation can be applied, and once a few hyperparameters have been specified, no further human intervention is needed, making this a class of 'blind' methods also.

We consider whether KPCA can provide a better separation between the 21 cm and foregrounds in the presence of instrumental effects than linear PCA-type methods. We use simulated data for a MeerKAT-like 21 cm survey that has been subjected to semi-realistic beam convolutions, and which includes realistic data-driven foreground models with complex spectral structure and masking effects, comparing our results with the well-known PCA and ICA methods. We also consider if improvements can be gained by smoothing the data, at each, frequency to a common resolution. We present our results in terms of the recovered (residual) power spectrum of the data, and also the recovered cross-power spectrum with a fictitious galaxy survey at the same redshifts.

This paper is structured as follows. In Section 2, we review the PCA and Kernel PCA methods, and the choices that can be made when configuring Kernel PCA, such as the choice of kernel and a possible pre-cleaning step. In Section 3, we describe the properties of the simulated 21 cm data cubes that we use in this study, including our models for the matter, 21 cm, and galaxy density fields, the instrumental beam, and a selection of foreground models. In Section 4, we present our results, comparing the Kernel PCA method with PCA and ICA in a number of different scenarios. We conclude in Section 5. Throughout this paper, we assume a ΛCDM cosmology with parameters $\Omega_{\rm cdm} = 0.25$, $\Omega_b = 0.05$, $h = 0.7$, $n_s = 0.95$, and $\sigma_8 = 0.8$.

2 FOREGROUND REMOVAL WITH KERNEL PCA

In this section, we introduce Kernel PCA as a foreground removal technique for 21 cm intensity mapping data. We first review the use of standard PCA as a foreground filter, before describing how Kernel PCA works and demonstrating the effect of different choices of kernel.

2.1 Principal Component Analysis

PCA is an unsupervised method that can be used to construct a suitable basis of frequency-dependent functions that can then be used to project the foregrounds out from the data. It works by finding the eigenbasis decomposition of the empirical frequency-frequency covariance matrix of the (mean-subtracted) data, i.e. averaging over

Downloaded from https://academic.oup.com/mnras/article/508/3/3551/6381718 by guest on 16 April 2024

the available angular pixels,

$$C(\nu, \nu') = \frac{1}{N_{\rm pix}} \sum_{j}^{N_{\rm pix}} \delta T_j(\nu) \delta T_j(\nu'). \quad (1)$$

The modes with the largest eigenvalues correspond to the largest sources of variance in the data. Since the foregrounds are orders of magnitude brighter than the 21 cm signal, they are expected to be the dominant source of variance and therefore mostly confined to these large-eigenvalue modes. We can then subtract the projection of the data in each pixel on to these modes, leaving behind a lower-variance residual that should in principle contain only 21 cm signal plus noise. Because the foregrounds tend to have smooth (power law-like) frequency spectra, these modes will also tend to be smooth functions of frequency.

In reality, there is not a perfect split between the foregrounds and the 21 cm signal; while substantially different in terms of their variance and spectral structure, they are not completely orthogonal. Some of the smoother radial modes in the 21 cm signal, corresponding to low values of $k_\parallel$ (and also low $k_\parallel$), can therefore also be projected out with the foreground modes, leading to loss of signal power. Depending on the number of PCA modes that are subtracted, this can substantially suppress the recovered 21 cm power spectrum over a range of scales. Typically, this signal loss is then corrected for by using the effect of the foreground filter on a known signal injected into the data to determine a signal loss transfer function, which can then be divided out from the measured power spectrum.

A further complication is that other spurious spectral structure can also be present in the data. The instrumental beam can modulate the sky signal in a way that mixes components and induces additional spectral structure for example (Matshawule et al. 2020), while polarization leakage can add non-smooth polarized foregrounds into temperature-only data. This requires more PCA modes to be subtracted, thus increasing the amount of signal loss. In the next section, we introduce Kernel PCA, a non-linear method that is capable of constructing more complicated ways of separating components.

2.2 Kernel PCA

Kernel PCA (Schölkopf et al. 1997) is an extension of PCA that operates on a non-linear transformation of the original data vectors. The transformed space (called the *feature space*) can have an arbitrarily higher dimension than the original data. In general, this makes it easier to find an effective segregation between different components of the signal, which can be achieved simply by using standard linear PCA in the higher-dimensional feature space.

As an example, consider a 21 cm data cube that we split up into a set of data vectors, one frequency spectrum (of dimension N_ν) per pixel on the sky, which we denote $\vec{x}_j$ for pixel j. A non-linear transformation is then applied, $\vec{y}_j = \Phi(\vec{x}_j)$, resulting in a transformed vector $\vec{y}_j$ of dimension $N_{\rm feature}$. Because the transformation Φ is non-linear, the feature space now has dimensions that are various non-linear combinations of the original dimensions of the data; for example, for a spectrum with $N_\nu = 2$ and original dimensions $\{\hat{x}_0, \hat{x}_1\}$, one could form a feature space with dimensions $\{\hat{x}_0, \hat{x}_1, \hat{x}_0\hat{x}_1, \hat{x}_0^2, \hat{x}_1^2, \ldots\}$, depending on the choice of Φ.[1] Linear combinations of these dimensions can then be found that decompose the feature-space covariance matrix,

$$\tilde{C} = \frac{1}{N_{\rm pix}} \sum_{j}^{N_{\rm pix}} \vec{y}_j \vec{y}_j^{\rm T} = \frac{1}{N_{\rm pix}} \sum_{j}^{N_{\rm pix}} \Phi(\vec{x}_j) \Phi^{\rm T}(\vec{x}_j), \quad (2)$$

into principal components. $\tilde{C}$ has dimensions $N_{\rm feature} \times N_{\rm feature}$, which can be very large. When these components are projected down on to the original lower-dimensional space of the data, the resulting functions can have more complicated functional forms than the simple linear forms that standard PCA is able to use, making this method more flexible in its ability to separate different components of a signal. To take a simple example, consider a set of data points that trace out a circle in the 2D space $\{\hat{x}_0, \hat{x}_1\}$. Linear PCA would not be able to provide a good description of this distribution, as it can only draw a straight line through it (i.e. a linear function of the data dimensions). The Kernel PCA method with a non-linear transformation Φ that generates dimensions $\hat{x}_0^2$ and $\hat{x}_1^2$ (amongst others) would easily be able to trace out a circle however.

The non-linear transformation Φ can result in an arbitrarily large feature space, making calculations unwieldy; explicit computation of the covariance matrix may be intractable for example. Fortunately, it is not necessary to work directly in the feature space. We can instead use the so-called *kernel trick* to rewrite the inner product of the transformed vectors as

$$k(\vec{x}, \vec{x}') = \Phi(\vec{x}) \cdot \Phi(\vec{x}'), \quad (3)$$

where k is the *kernel*. The following matrix can then be formed by using the kernel to calculate the inner product of all pairs of transformed data vectors,

$$K_{ij} = k(\vec{x}_i, \vec{x}_j) = \Phi(\vec{x}_i) \cdot \Phi(\vec{x}_j), \quad (4)$$

which has shape $N_{\rm pix} \times N_{\rm pix}$. (The transformed data vectors are typically also mean-centred before this step, i.e. $\Phi(\vec{x}_j) \to \Phi(\vec{x}_j) - (1/N_{\rm pix}) \sum_i \Phi(\vec{x}_i)$.) Only this object is needed to perform a PCA decomposition of the feature-space covariance matrix, and not the covariance matrix itself.[2] The resulting eigenvectors in the feature space do not need to be evaluated explicitly either; instead, we can project an arbitrary vector in the data space on to the higher-dimensional principal components, again using the matrix K_{ij} to evaluate the necessary inner products. This makes working with even infinite-dimensional non-linear transformations tractable – any calculations involving inner products in the feature space can be evaluated simply by evaluating the kernel function for pairs of vectors in the much lower-dimensional data space, rather than having to go into the feature space and explicitly sum the individual components of the inner product for $N_{\rm feature}$ dimensions. For more details, and simple proofs of these statements, see Schölkopf et al. (1997).

To implement a foreground filter with Kernel PCA, we can then perform a similar procedure to standard PCA and simply subtract off templates constructed from the first few eigenmodes of the feature-space covariance from each pixel. In this case, the modes used are now the *projections* of the feature-space eigenmodes on to the data. While the templates themselves are constructed from non-linear functions of the dimensions of the data, the subtraction step itself is a linear operation.

Finally, we recall that the size of the inner product matrix K_{ij} ($N_{\rm pix} \times N_{\rm pix}$) can be quite large for a typical data set, and so operations

[1] To clarify our notation, we denote basis vectors as $\hat{x}_n$ and their coefficients as x_n, for each $n = 0 \ldots N_\nu - 1$, so that $\vec{x}_j = x_{0j}\hat{x}_0 + x_{1j}\hat{x}_1 = (x_{0j}, x_{1j})$.

[2] We need not even define an explicit form for Φ; by choosing a form for the kernel k, a particular Φ is implicitly defined.

Downloaded from https://academic.oup.com/mnras/article/508/3/3551/6381718 by guest on 16 April 2024

on this matrix tend to dominate the computational cost of the Kernel PCA algorithm. It should therefore be noted that Kernel PCA can have a substantially higher computational cost than standard PCA on the same data set.

2.3 Choice of kernel

Different choices of kernel generate different feature spaces, which may be more or less suited to segregating particular mixtures of signal depending on the functional forms that they allow. The kernels themselves also have hyperparameters that can be tuned to produce better results, e.g. by promoting more or less smoothness of the functional forms, fixing a typical length-scale and so on. Common choices of kernel include polynomials, radial basis functions (Gaussians), and sigmoid (tanh) functions, amongst others.

In this paper, we compare these three common choices of kernel to make an empirical determination of suitability to the 21 cm foreground removal problem. We do not seek to find optimal choices of kernel or hyperparameters, leaving this to future work.

We use the following kernels, as implemented in the `decomposition.KernelPCA` function of `scikit-learn` (Pedregosa et al. 2011):

$$k(\vec{p},\vec{q}) = \left(\gamma\,\vec{p}^{\mathrm{T}}\vec{q} + c\right)^{d} \text{ (Polynomial)} \tag{5}$$

$$\exp\left(-\gamma||\vec{p}-\vec{q}||^2\right) \text{ (Gaussian/RBF)} \tag{6}$$

$$\tanh\left(\gamma\,\vec{p}^{\mathrm{T}}\vec{q} + c\right) \text{ (Sigmoid).} \tag{7}$$

The hyperparameters γ, c, and d control the scale of the kernel, an offset, and the maximum order of the polynomial basis respectively. The kernel scale (γ) controls the curve of the decision boundary, which is the hypersurface that separates out the different components within feature space (Shadeed, Alwan & Abd 2020). The way that each of these parameters changes the detailed behaviour of the kernel is kernel-dependent, and the best choice of kernel for a particular problem is typically selected through some combination of trial and error and hyperparameter optimization. Zhang, Zhou & Chen (2006) present an unsupervised method of selecting the optimum kernel for use by first transforming the data into feature matrices, using numerous different kernels, and then finding the kernel combination that maximizes the Frobenius norm of the column-wise and row-wise elements of the feature matrices. Although the data themselves determine the optimum kernel and hyperparameter choices, these choices are not physically intuitive. We do not explore this issue further here.

2.4 Pre-cleaning

Even the most basic foreground removal methods are broadly successful at reconstructing the true 21 cm signal, producing residuals of order the size of the cosmological signal or less despite the foregrounds themselves being several orders of magnitude larger. The bigger challenge for foreground removal methods are subtler features in the data with more complex spectral signatures, such as those generated by the interaction between foreground emission and frequency-dependent beam sidelobes.

We are primarily interested in how Kernel PCA performs in the presence of these features. As such, we focus our analysis on the data *after* the bulk of the foreground structure has already been removed, or 'pre-cleaned', using a rough initial pass of another foreground cleaning method. Ideally, the pre-cleaning technique would reduce the variance of the foreground fluctuations from four orders of magnitude larger than the 21 cm signal fluctuations to around the same order of magnitude (e.g. perhaps a few times larger), while leaving the 21 cm signal as intact as possible at this stage, i.e. minimizing the signal loss. The desired behaviour can be verified by eye, to check that the pre-cleaned maps display temperature fluctuations larger than a few times the expected level of the 21 cm signal for example. We use a simple linear PCA method that removes only two eigenmodes to do the pre-cleaning, but any technique capable of reducing the foreground variance by several orders of magnitude without over-fitting would be suitable.

Unlike some other methods, we note that this pre-cleaning step appears to be necessary for KPCA; without it, substantially larger residuals are obtained, even when the number of subtracted modes is increased and different kernels and hyperparameters are used. While we do not have a detailed explanation for this behaviour, we anticipate that it is caused at least in part by the sensitivity of non-linear decompositions to signals separated by a large dynamic range. Pre-cleaning improves and stabilizes the behaviour substantially, suggesting that KPCA's strengths are in separating complicated signals of a similar size rather than in reducing the dynamic range of the data.

3 SIMULATIONS OF AUTO-CORRELATION SURVEYS

In this section, we describe a set of simple simulated 21 cm auto-correlation survey data cubes that we use to test Kernel PCA. We simulate a 3600 degree square region and cover the frequency range of 950.6–1234.7 MHz; replicating a cosmological box of dimensions $1.25 \times 1.25 \times 1.25$ Gpc in the redshift range $z = 0.150 - 0.494$. The simulated 21 cm field is based on Gaussian realizations of the matter power spectrum with appropriate linear bias factors and a lognormal transform applied. We also add two different sets of foreground models, frequency-dependent Gaussian instrumental noise, and convolve the fields with a realistic frequency-dependent beam model, based on the specifications of the MeerKAT array. We choose to focus on the MeerKAT experimental set-up due to the recently proposed 21 cm intensity mapping survey: MeerKLASS (Santos et al. 2017), which is currently underway (Wang et al. 2021).

3.1 Neutral hydrogen distribution

On large scales, beyond the non-linear regime, the matter distribution is expected to be approximately Gaussian. Apart from at the lowest redshifts, the beam smoothing for MeerKAT and SKAO in auto-correlation mode is sufficiently large that the density field can only be recovered on large linear scales anyway, and so non-linearities are only a minor concern. As such, we would expect Gaussian density fields to be a good approximation for essentially all of the scenarios that we consider in this paper. This allows us to use a particularly simple method to generate simulated density fields:

(i) Draw unit Gaussian random numbers in each cell of a cubic grid and perform a Fast Fourier Transform (FFT).

(ii) In Fourier space, multiply the value in each grid cell by the square root of the matter power spectrum, $P(k, z)$.

(iii) After multiplying by an appropriate normalization, perform the inverse FFT to recover a Gaussian random realization of the matter density field.

Downloaded from https://academic.oup.com/mnras/article/508/3/3551/6381718 by guest on 16 April 2024

We use CCL[3] (Chisari et al. 2019) to calculate the matter power spectrum, and allow the grid to have different side lengths in each dimension.

Pathologies can arise in the Gaussian realizations when examining smaller scales however. For example, the value of the density contrast, δ_m, can drop below -1 in some grid cells if the grid cell size is small enough. One way of fixing this issue is to apply a lognormal transformation to the Gaussian density field (Coles & Jones 1991; Agrawal et al. 2017), which guarantees that the density contrast remains in the physical range, $\delta_m \in [-1, \infty]$. This also provides a simple model of the non-linear density field, yielding a configuration-space matter power spectrum that is accurate to within about 10 per cent at $k = 0.2\,h/\mathrm{Mpc}$ (Chuang et al. 2015). In all that follows, we use lognormal transformations to model the matter and 21 cm brightness temperature fields, as well as the galaxy number counts discussed in Section 3.4.

For the 21 cm brightness temperature field, we adopt a simple linear prescription to map the matter density contrast to brightness temperature fluctuations, such that

$$\Delta T_b(\vec{x}, z) = \overline{T}_b(z)\, b_{\rm HI}(z)\, \delta_m(\vec{x}, z). \tag{8}$$

The mean brightness temperature, H I bias, and fractional H I density are given by the following fitting formulas, calculated from the model described in Bull et al. (2015):

$$\begin{aligned}
\overline{T}_b(z) &= 0.055919 + 0.23242\, z - 0.02414\, z^2 \\
b_{\rm HI}(z) &= \frac{b_{\rm HI,0}}{0.677105}\left(0.66655 + 0.17765\, z + 0.050223\, z^2\right) \\
\Omega_{\rm HI}(z) &= \frac{\Omega_{\rm HI,0}}{4.86}\left(4.8304 + 3.8856\, z - 0.65119\, z^2\right),
\end{aligned} \tag{9}$$

where we set $\Omega_{\rm H\,I,\,0} = 4.86 \times 10^{-4}$ and $b_{\rm H\,I,\,0} = 0.677\,105$. We incorporate redshift-space distortions by shifting each voxel of the lognormal transformed field to a new radial position,

$$s_\parallel = x_\parallel - \frac{\mathrm{v}_\parallel + \mathrm{v}_{\rm NL}}{H(z_c)}, \tag{10}$$

where $x_\parallel$ is the original radial coordinate of the voxel centre, $\mathrm{v}_\parallel$ is the velocity in the radial direction (calculated from the matter density field using the linear continuity equation), the non-linear velocity $\mathrm{v}_{\rm NL}$ in each voxel is drawn from an uncorrelated Gaussian random distribution with standard deviation $\sigma_{\rm NL} = 120\,\mathrm{km\,s^{-1}}$, and the expansion rate $H(z)$ is evaluated at the central redshift of the simulation box, z_c. After applying this shift, the voxels are interpolated back on to a regular grid. The expression in equation (8) can then be updated to its redshift-space equivalent,

$$\Delta T_b(\vec{s}, z) = \overline{T}_b(z_c)\, \delta_s(\vec{s}, z_c), \tag{11}$$

where δ_s is the biased lognormal transform of the H I density contrast in redshift space.

With the redshift-space brightness temperature distribution in hand, we then project the simulated box into observational co-ordinates, i.e. redshifts/frequencies and angles. We employ the distant observer approximation, evaluating the relevant cosmological quantities at the central redshift of the box, and ignoring their evolution within the box. Using this approximation, we also neglect curved sky effects, such that each frequency slice occupies the plane perpendicular to the line of sight connecting the observer to the centre of the simulation box. Since the solid angle subtended by the simulation volume ($60^\circ \times 60^\circ$) is quite large and the central redshift of $z_c \approx 0.32$ (corresponding to a comoving distance to the centre of the box of 1.28 Gpc) is reasonably small, this approximation is expected to be limited in accuracy on the largest scales we consider. We do not expect this to significantly impact our conclusions however.

3.2 Instrumental beam

The angular and spectral structure of the observed foregrounds is strongly affected by the behaviour of the instrumental beam that the sky signal is convolved with. At its most basic, the effect of the beam is to introduce a frequency-dependent angular smoothing, of characteristic scale $\theta \sim \lambda/D_{\rm dish}$, where λ is the wavelength of observation, and $D_{\rm dish}$ is the diameter of each dish ($\sim$13.5 m for MeerKAT). This has the effect of mixing together the signals from foreground regions that may have different spectral indices etc., resulting in a slightly more complex spectral behaviour while smoothing away any small-scale structure.

The existence of frequency-dependent sidelobes (secondary, lower-level peaks in sensitivity) and nulls (zeros or low-sensitivity regions) in the beam, at angles far from the pointing centre, further complicates the spectral structure by modulating the otherwise smooth spectra of sources that fall within these regions. While the amplitude of the sidelobes is typically much lower than the centre of the beam, they can cover a large solid angle, and so contribute a non-negligible fraction of the total detected emission after integrating over the entire beam pattern. Bright sources that move in and out of sidelobe/null regions as a function of frequency can also contribute substantially to the total signal, with a frequency spectrum that will be strongly modulated. While other, related, sources of foreground chromaticity also arise, such as gain errors and polarization leakage, we will only include beam chromaticity in this paper.

For our beam model, we adopt the L-band, Stokes I polarization beam pattern from the KATBEAM package (Mauch et al. 2020),[4] which provides a simplified model of the MeerKAT beam with substantial sidelobe structure and a small frequency ripple in the beamwidth, as described in Asad et al. (2021). The beam model is convolved with the simulated sky separately in each frequency channel using a 2D FFT, which preserves the mild asymmetry of the beam while assuming periodicity of the survey area in each frequency channel. We do not attempt to model a scanning strategy for the survey, which would result in the measured intensity in each pixel deriving from a mixture of beam convolutions at different angles, so the beam convolution we have applied is also simplified in this sense. The beam model as a function of frequency is shown in Fig. 1.

3.3 Foreground models

Bright foreground contamination – largely from diffuse Galactic synchrotron emission and extragalactic radio sources – is the primary systematic effect in 21 cm surveys. While the foreground emission itself is expected to have a smooth frequency spectrum in total intensity, additional spectral structure is imparted by calibration errors and the structure of the instrumental beam. Polarized emission, with a more complex spectral structure due to effects such as Faraday rotation, can also leak into the total intensity channel, further complicating the foreground signal.

We consider two different models of foreground contamination in this work: the Global Sky Model (GSM) and the *Planck* Sky Model

[3] https://github.com/LSSTDESC/CCL

[4] https://github.com/ska-sa/katbeam

Downloaded from https://academic.oup.com/mnras/article/508/3/3551/6381718 by guest on 16 April 2024

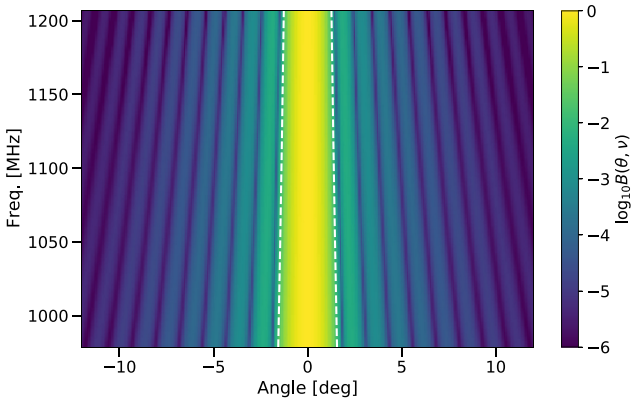

Figure 1. A slice through the beam model for the MeerKAT L-band (Stokes I polarization) from KATBEAM, as a function of angle and frequency. The white dashed lines show the approximate beamwidth as a function of frequency, $\Delta\theta \approx 1.2\lambda/D_{\rm dish}$.

Table 1. Components within our implementation of the PSM foreground model and their parameters.

FG component	Value	Reference
Free–Free amp.	varies	Planck Collaboration XII (2016)
Free–Free β	-2.10	Bennett et al. (1992)
Synchrotron amp.	varies	Planck Collaboration XII (2016)
Synchrotron β	varies	Miville-Deschênes (2008)
PS cut-off flux	0.1 Jy	Olivari et al. (2016)
PS β	-2.70 ± 0.2	Olivari et al. (2016)

(PSM). These models are constructed in quite different ways, and so by comparing them we can get a fair picture of the performance of various foreground removal methods, without needing to worry that they might be over-optimized for a particular foreground model, e.g. due to exploiting unrealistic structures.

The GSM (Zheng et al. 2017) is a publicly available algorithm[5] capable of producing all-sky HEALPix maps of diffuse Galactic emission at any frequency between 10 MHz and 5 THz. The GSM combines information from 29 different sky surveys, each covering different regions of the sky and frequencies, as well as having different resolutions, and uses PCA to pick out the major sky components and interpolate them across the full frequency range and sky area. Below 10 GHz, GSM is capable of producing maps at a resolution of either 5 degrees or 56 arcmin; we use the 56 arcmin simulations.

The PSM, as implemented in the *Planck* full sky simulations (Planck Collaboration XII 2016), is also publicly available in the form of the 'full focal plane' simulations on the *Planck* Legacy Archive.[6] While the GSM simulates the total diffuse sky emission present at each frequency, the PSM provides separate component maps for (e.g.) free–free and synchrotron emission. We assemble a set of foreground maps from the PSM using the method detailed in Carucci et al. (2020) and Cunnington et al. (2021). The pertinent details of our PSM model are summarized in Table 1, and the individual components are shown in Fig. 2.

[5]https://github.com/telegraphic/pygdsm
[6]https://pla.esac.esa.int/#maps

Our foreground simulations cover a 3600 square degree region centred at high Galactic latitude: $(l, b) = (120^\circ, -59^\circ)$. We project the simulations on to the flat sky using the HEALPix Cartesian projection. Given the size of the survey area, using the flat-sky approximation does result in a level of angular distortion of the foreground maps, which would need to be corrected in an observational study. For our simulated study, however, we neglect this correction, instead choosing to consider the distorted projected maps as the 'true' foreground model here. While the PSM simulations are just the summation of various components that each follow a power law, the GSM can be seen to display more complex frequency behaviour. Fig. 3 shows the mean emission temperatures from our region of interest at each frequency, multiplied by frequency squared and normalised to 1 at 956.7 MHz, from the GSM and PSM foreground simulations. The PSM simulation can be seen to closely follow the form of a power law, while the GSM simulation displays a different spectral dependence. In Fig. 4, we show corner plots for three frequencies of the total foregrounds within our region of interest plus 21 cm signal data, convolved with the beam, for the cases of both PSM and GSM foregrounds. Both cases show a strong positive, linear correlation between frequency channels, but the GSM-based simulated data exhibit more distinct deviations from this linear correlation. Each GSM frequency map is accompanied with an error budget of 5 per cent to 15 per cent, which would account for these deviations. This provides an interesting test case, as the GSM foreground model can be used as way to include analogous effects to spectral calibration errors alongside the standard perfect power-law representations of both the Galactic and extragalactic foregrounds in the PSM model.

In this work, we use a 3600 square degree region centred at high Galactic latitude: $(l, b) = (120^\circ, -59^\circ)$. As discussed in the previous section, the simulated data cubes, including foregrounds, are convolved with a realistic anisotropic beam model with sidelobes and non-trivial spectral structure. While we do not explicitly include a model for gain calibration errors, we use the GSM to provide the total foreground emission at each frequency and these maps are accompanied with a 5 per cent to 15 per cent error budget that is analogous to gain calibration errors. However, we do not consider additional spectral structure due to polarization leakage in this paper.

3.4 Mock galaxy distribution

In addition to the 21 cm field, we also generate a mock galaxy catalogue from our simulated density field to allow us to study the impact of foreground cleaning on the galaxy-21 cm cross-correlation signal too. To generate the catalogue, we use a lognormal transform of the linear matter density with a simple linear bias term $b(z)$ that depends only on redshift, so that the galaxy density contrast δ_g in each voxel is (Agrawal et al. 2017; Hand et al. 2018)

$$1 + \delta_g(\vec{x}) = \frac{\exp\left(b(z_c)\delta_m(\vec{x})\right)}{\langle b(z_c)\delta_m(\vec{x})\rangle}, \tag{12}$$

where z_c is the central redshift of the simulation box and in this case the angle brackets $\langle \dots \rangle$ denote a spatial average, necessary to obtain the correct normalization. The galaxy counts are realized using a Poisson random draw in each voxel, with a rate parameter (expected number of galaxies in the voxel)

$$\bar{N}(\vec{x}) = \delta V\, \bar{n}(z_c)\,(1 + b(z_c)\delta_m), \tag{13}$$

where δV is the comoving voxel volume and $\bar{n}$ is the expected comoving number density of galaxies. This simplified treatment is sufficient to provide a reasonable model of galaxy-21 cm cross-

Downloaded from https://academic.oup.com/mnras/article/508/3/3551/6381718 by guest on 16 April 2024

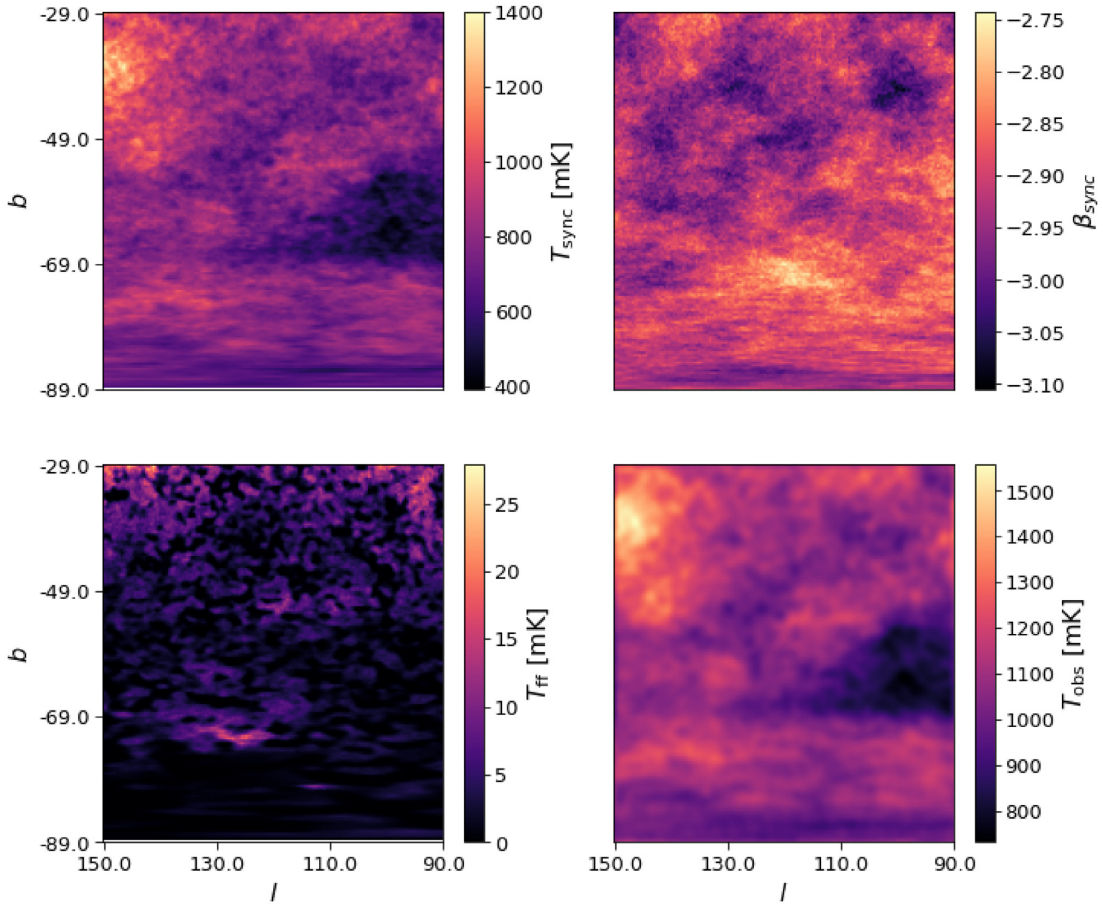

Figure 2. The PSM foregrounds in Galactic coordinates. (Upper left): synchrotron emission amplitude at 950.6 MHz. (Upper right): synchrotron spectral index. (Lower left): free–free emission amplitude at 950.6 MHz. (Lower right): combined diffuse plus point source contributions at 950.6 MHz, after convolution with the beam.

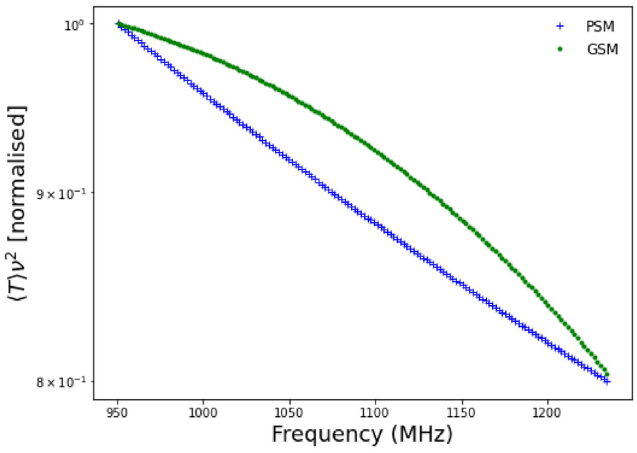

Figure 3. The mean emission temperature at each frequency from the GSM and PSM foreground simulations within the high Galactic latitude region investigated in this work; $(l, b) = (120°, -59°)$. The temperature curves are multiplied by frequency squared to emphasis any spectral discrepancies between the two models, and are normalized to unity at the lowest frequency, 950.6 MHz.

correlations for an auto-correlation survey; the linear bias model is expected to be accurate on large scales, and small-scale features of galaxy clustering that we have neglected, such as non-linear velocity dispersion, scale-dependent bias, and decorrelation between the 21 cm and galaxy fields on small scales, are relatively less important due to the suppression of the 21 cm field at high k due to the instrumental beam. Note that there are exceptions to this reasoning for auto-correlation experiments with large dishes operating at low redshifts, such as the Parkes survey, which have sufficient angular resolution for small-scale galaxy clustering to be an important consideration (Wolz, Blake & Wyithe 2017b; Anderson et al. 2018).

Rather than seeking to model a particular spectroscopic galaxy survey, we make simplistic choices of $b(z) = \sqrt{1+z}$ and $n(z) = 2 \times 10^{-3}\,\mathrm{Mpc}^{-3}$. This ensures that shot noise is unimportant, so the dominant effect on the cross-correlation signature will be systematic effects on the 21 cm signal, such as the instrumental beam and residual foregrounds.

4 RESULTS

In this section, we present the results for the Kernel PCA foreground removal method applied to our total emission data cube. This data cube consists of the 21 cm signal plus foreground components, which

Downloaded from https://academic.oup.com/mnras/article/508/3/3551/6381718 by guest on 16 April 2024

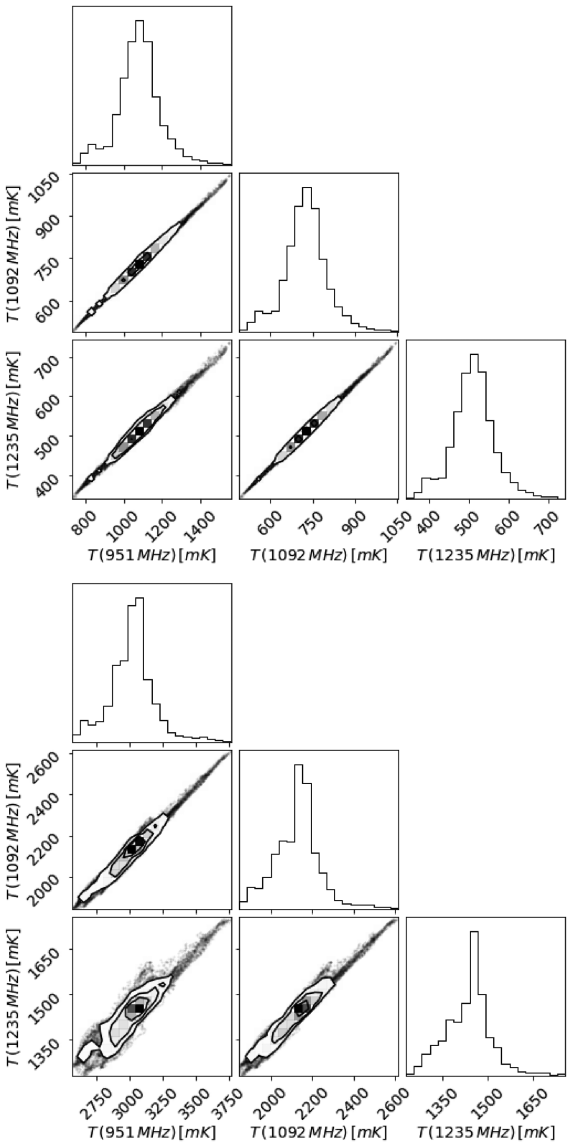

Figure 4. Corner plots to show the correlation across frequency for the 'observed' data from the *top:* PSM simulations and *bottom:* GSM simulations.

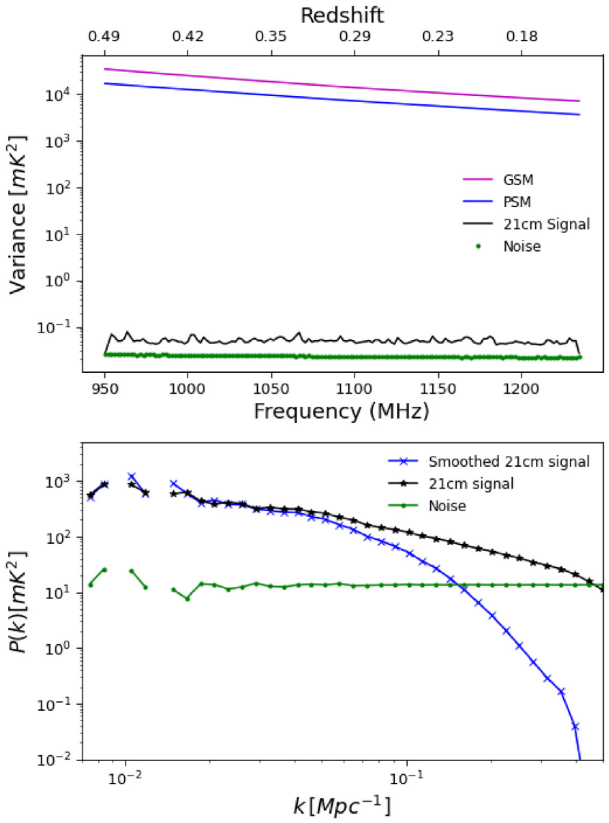

Figure 5. Top: the 21 cm, PSM, and GSM foreground and noise variances as a function of frequency/redshift. Bottom: the radial power spectrum for the 21 cm signal, before and after convolution with the MeerKAT beam, and instrumental noise.

are convolved with the MeerKAT beam and then combined with Gaussian instrumental noise. The top panel of Fig. 5 shows the variance of each of these components, before convolution with the MeerKAT beam, as a function of frequency. The effect of the beam on the 21 cm radial power spectra can be seen in the bottom panel, where it causes the 21 cm signal to dip below the noise level at large k values. We restrict our foreground cleaning analysis to k ranges lower than 0.1 Mpc^{-1} so as to ensure that the thermal noise level has negligible effects on our results.

Kernel PCA is used in a number of different simulated scenarios, and for a range of different hyperparameters. In all cases, we compare with PCA and ICA filters, using what we have found to be their optimal configurations in terms of recovery of the 1D signal power spectrum from the simulated data. For PCA and ICA, this means a choice of $n = 3$ foreground modes were selected for removal; this number produced the foreground-cleaned power spectra closest to the simulation 21 cm power spectra convolved with the beam. For all of the analyses shown here, ICA behaves identically to PCA,[7] and so we only show the results from PCA in most figures. For all plots and tables where ICA is not explicitly mentioned, the term 'PCA' should be taken to refer to 'PCA and ICA'.

Kernel PCA requires three choices: (1) which kernel to use; (2) the number of modes, n, to be removed; and (3) the values of the kernel hyperparameters. Additionally, the method and degree of pre-cleaning must also be chosen. In this work, we find the Sigmoid kernel to work best (see Section 4.1) and so must further choose n and the values for the Sigmoid kernel's two hyperparamters: the kernel width γ and the kernel offset c. As the kernel width is typically set to some scaling of n, we consider the true hyperparameter choice to be this scaling factor, ϵ, where $\gamma = 1/(\epsilon\, n)$.

For our standard set of simulations, we choose a resolution of 150 by 150 pixels to capture the 60 by 60 degree field of view. In Section 4.2, we shall investigate the effect resolution choice has on our analysis. For our standard (150 by 150 pixels) simulations, we set the number of modes to be removed by KPCA to be $n = 5$ for the PSM foreground simulations and 7 for the GSM simulations. The fact that a larger number of modes was required to clean the GSM simulations reflects the additional spectral features present in the

[7]Very similar results from PCA/SVD and ICA have also been obtained in past simulated analyses (e.g. Alonso et al. 2015). When applied to real data, the results can be quite different however (e.g. Wolz et al. 2017a). This discrepancy may result from systematics present in the real data, or perhaps more strongly non-Gaussian foregrounds than are included in the simulations.

Downloaded from https://academic.oup.com/mnras/article/508/3/3551/6381718 by guest on 16 April 2024

GSM foreground simulations. The optimal hyperparameter values will be discussed in more detail in Section 4.2, but for now we shall just state the configuration used for our standard simulations: $\epsilon = 10$ and $c = 0.9$.

For our analysis we compare the recovered 21 cm emission from each cleaning method with the true 21 cm signal convolved with the frequency-dependent beam. The ratio between the recovered PCA/ICA/KPCA-filtered 1D power spectrum and the true 21 cm beam-convolved 1D power spectrum is used as the figure of merit throughout. As the ideal cleaning method would produce a 1D power spectrum ratio of 1, the ratio $= 1 \pm 0.05$ region is highlighted on each of the ratio plots.

Fig. 6 shows the ratio of the recovered to true 1D power spectrum for each investigated technique. The top panel demonstrates the role of pre-cleaning for the PSM foregrounds, with the power spectrum ratio for the residual produced by PCA pre-cleaning shown in green. This residual is roughly double the level of the 21 cm signal. The brown curve shows the result of applying KPCA to this residual, as opposed to the total data cube. The middle panel shows the same brown curve, but this time plotted alongside the result from a PCA (3-mode) clean on the total data (cyan) and the residual from before (blue). Using PCA first with two modes, then with three, is no different from using PCA with five modes. Unlike Kernel PCA, ICA and linear PCA end up over-cleaning the total data and removing a substantial fraction of the 21 cm signal when given the pre-cleaned data cube. KPCA, on the other hand, makes use of the this pre-cleaning step, and can be seen to outperform both ICA and PCA after the data have been pre-cleaned. The bottom panel of Fig. 6 also shows the ratio of the recovered to true 1D power spectrum but for the case of GSM foregrounds, with similar conclusions.

The advantage of using KPCA over PCA is far more pronounced at low k/larger scales. In fact, it can seen for the GSM foreground simulations that PCA outperforms KPCA at high k values. Comparing both the middle and bottom plot of Fig. 6 reveals that gain calibration-like errors, as modelled by the additional spectral structure in the GSM simulations, result in larger inaccuracies for PCA, ICA, and KPCA at small scales (high k values). For k values lower than 6×10^{-2} Mpc^{-1}, KPCA is seen to outperform PCA, but for the smallest scales PCA would remain the preferred method. The inability to completely recover the 21 cm signal at large angular scales/low k values has been shown to be a feature of blind cleaning techniques (Switzer et al. 2015); the number of spurious correlations between foregrounds and the 21 cm signal is inversely proportional to the square root of the number of independent signal modes (which decreases at lower k values). Any improvements that can be made over standard PCA at large angular scales are therefore worth pursuing.

For both KPCA and PCA, the eigenvectors with the largest eigenvalues correspond to the components that the algorithm has, effectively, identified as foregrounds. We use KPCA on the pre-cleaned observation cube, which has the first two largest PCA components removed from it already, and so it is constructive to compare what both the KPCA and PCA determine to be a 'foreground' *after* this initial pre-cleaning step. Figs 7 and 8 compare these two post-pre-cleaning estimates of the 'foreground' component for both the PSM and GSM foreground cases, respectively. The plots show the foreground estimate for the first frequency channel, while the power spectra take into account the full data cube and so all available frequencies.

Fig. 7 displays point-like emission contributions whilst Fig. 8 does not; this is because the GSM is a diffuse sky model, which does

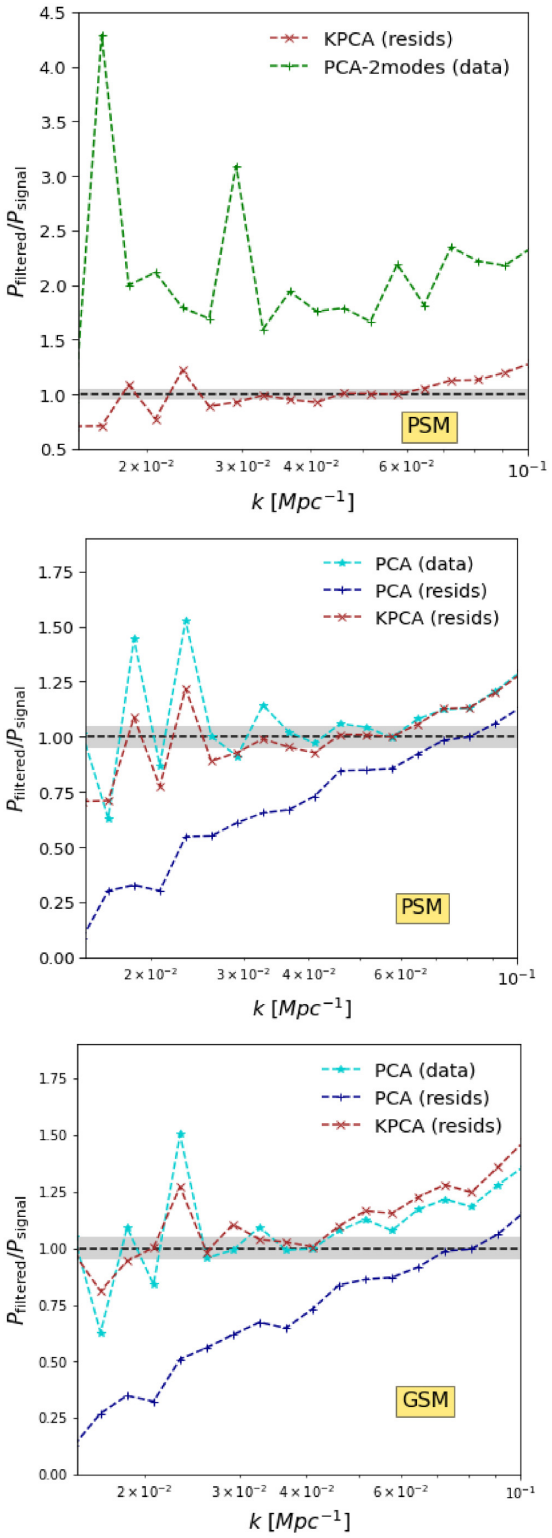

Figure 6. Ratio of recovered 1D power spectrum after foreground-filtering the data, $P_{\rm filtered}(k)$, to the input 21 cm signal power spectrum, $P_{\rm signal}(k)$, for three different foreground filtering algorithms. The top two figures apply to the case of PSM foregrounds while the bottom figure is for the case of GSM foregrounds. The grey shaded area denotes the ± 5 per cent region around perfect signal recovery (a ratio of unity, denoted by the horizontal dashed line). The parentheses indicate whether the foreground cleaning method has been applied to the 'observed' data cube ('data') or the pre-cleaned observation cube ('resids', i.e. following a 2-mode PCA removal).

Downloaded from https://academic.oup.com/mnras/article/508/3/3551/6381718 by guest on 16 April 2024

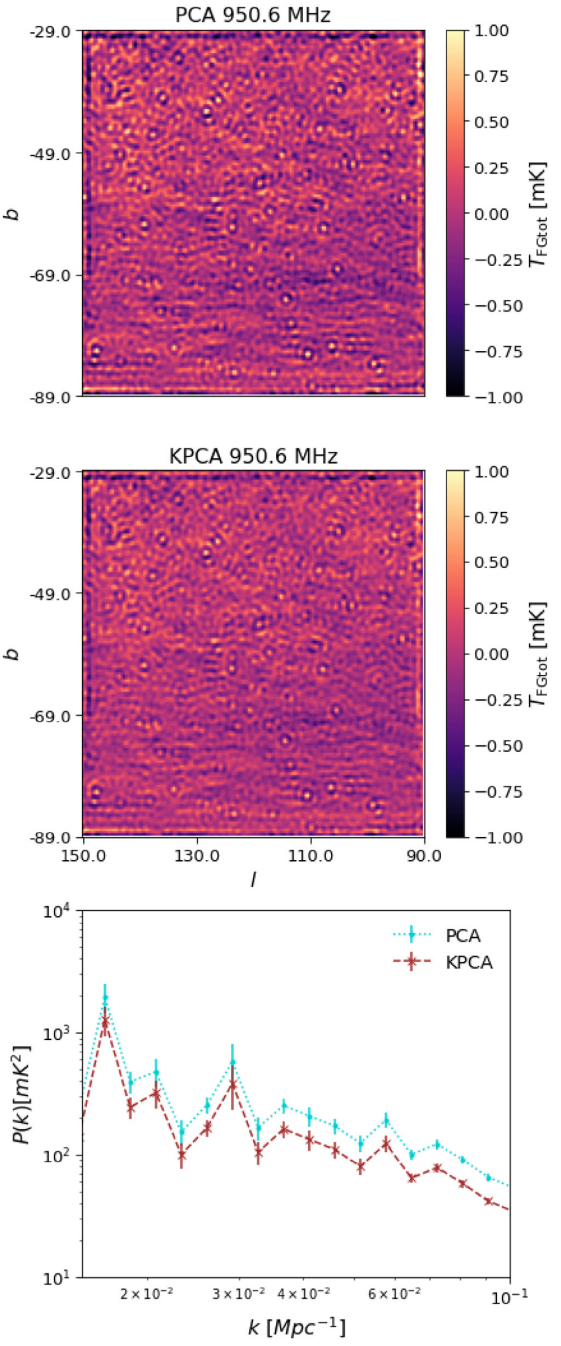

Figure 7. The foreground component after the initial pre-clean as identified by PCA (upper panel) and KPCA (middle panel) for the PSM foreground case. The maps are shown in Galactic coordinates. (Lower panel): the 1D power spectra for the KPCA and PCA foreground components, after the initial pre-clean.

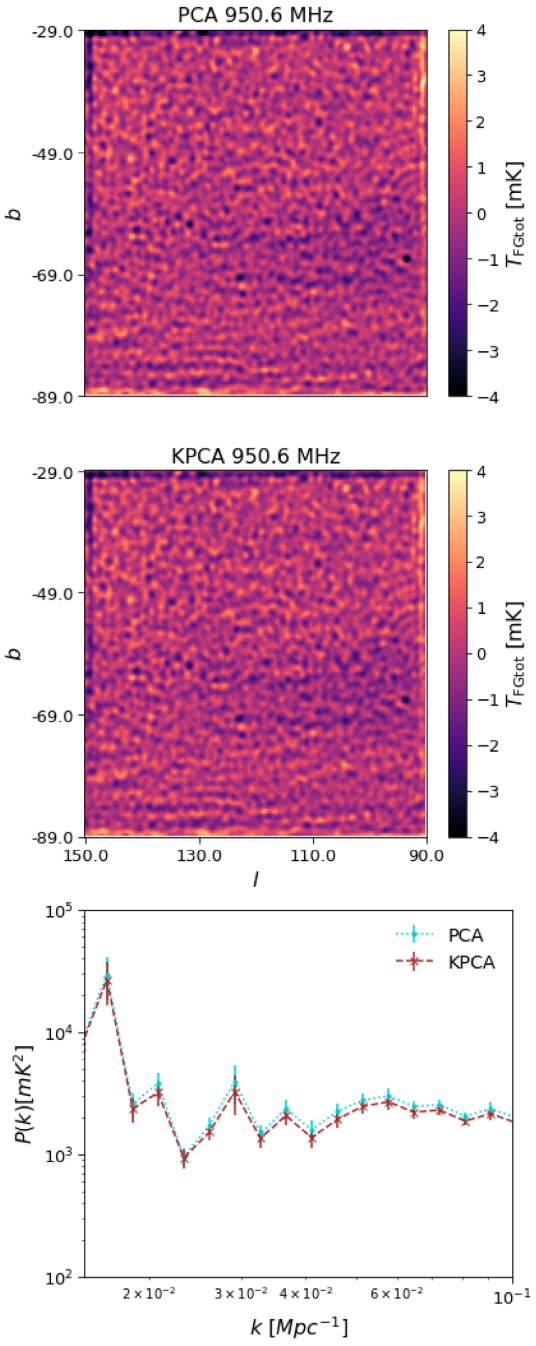

Figure 8. The same quantities as in Fig. 7, but now for the GSM foregrounds.

not directly include a resolvable point source contribution. There is little visible difference between the structure seen by the KPCA and PCA foreground estimate, after the pre-clean, in both the PSM and GSM cases. The most notable difference between the two in both foreground cases (though the difference is more pronounced for the PSM case) is the power level, which is always higher for PCA. KPCA, with the number of components set to 5, removes less power from the total emission maps than PCA with the number of components set to 3. A clear advantage to using KPCA over standard PCA is therefore the finer level of precision with which contaminants can be identified and removed. In some sense, this can be thought of as allowing fractional numbers of PCA components to be removed, e.g. subtracting the equivalent of more than 2 but less than 3 PCA components from the total observational cube. With appropriate optimization of the KPCA parameters, this should allow for reduced signal loss while still removing the bulk of the foreground emission.

4.1 Effect of different choices of kernel

The choice of which kernel/hyperparameters to use has the greatest effect on the successful application of KPCA. Similar discussions can

Downloaded from https://academic.oup.com/mnras/article/508/3/3551/6381718 by guest on 16 April 2024

be seen in the application of Gaussian Process Regression to 21 cm intensity mapping simulations (Mertens et al. 2018; Ghosh et al. 2020; Kern & Liu 2021; Soares et al. 2021a); siting GPR within a Bayesian framework, the optimum choice of kernel and hyperparameters can be found by comparing the log-marginal likelihood for each kernel/hyperparameter set under investigation. In other applications of KPCA, a cross-validation technique is often applied where the various kernel/hyperparameter options are tested on subsets of the data, and the kernel/hyperparameters that can produce the most consistent reproductions of the desired image (or otherwise) from these subsets are adopted. In order to compare the results from different kernel selections without knowing the true form of the image that KPCA is trying to recover, the kernel estimates must be mapped back from their feature space into the common input space (the image pixel space). This is known as the 'pre-imaging' problem (Alam & Fukumizu 2014).

In our application of KPCA, we determine the optimal kernel/hyperparameter set empirically, which is also what we do for PCA and ICA. Future applications of KPCA to observed 21 cm intensity mapping data could instead use simulations to inform the kernel/hyperparameter choice, as well as the technique of eigenvalue convergence which looks at the ratio between foreground eigenvalues and the total data eigenvalues (Cunnington et al. 2021). We do not investigate optimization methods further here however.

Our analysis mostly focuses on the Sigmoid kernel, the choice of which we justify in Fig. 9. For both the PSM and GSM foreground cases, the sigmoid kernel is optimal for our particular simulation set-up when compared with the polynomial and RBF kernels. Interestingly, for the PSM foregrounds the sigmoid kernel produces a different recovered 21 cm power spectrum shape as a function of k when compared to the polynomial and RBF kernels, while for the GSM foreground all three kernels can be seen to seen to have similar recovered 21 cm power spectrum forms.

For the sake of comparison, each kernel is given the same set of shared hyperparameters ($n = 5$, $\epsilon = 10$, $c = 0.9$). The polynomial kernel also has an additional parameter: the order of the polynomial, d. For the case of the PSM foreground, the optimal value for this parameter was $d = 1$, whilst the GSM foregrounds the optimal value was found to be $d = 4$. We examine the effect of changing hyperparameter values in the next section.

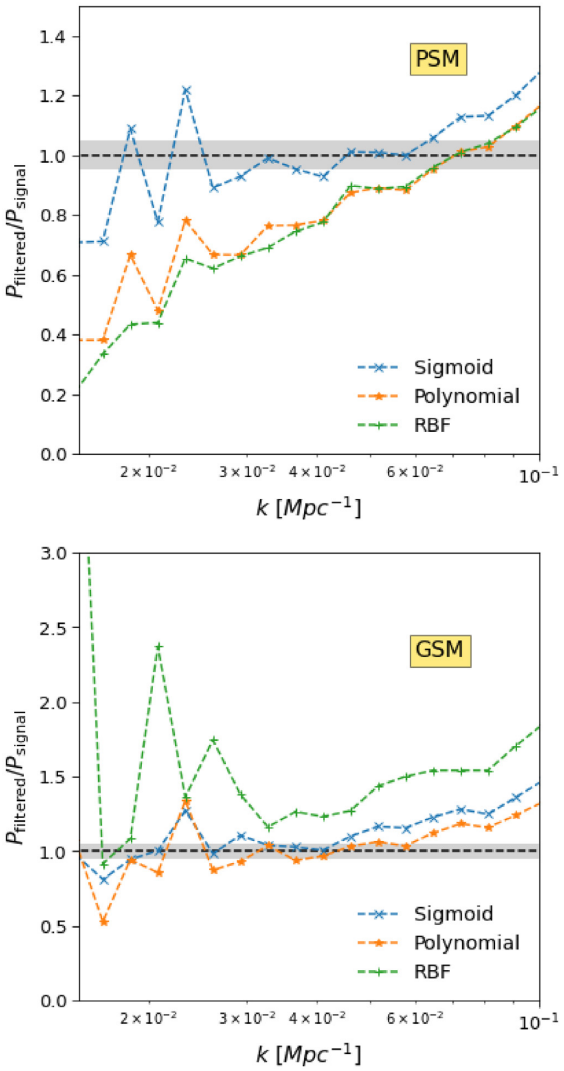

Figure 9. Ratio of recovered 1D power spectrum after foreground-filtering the data, $P_{\rm filtered}(k)$, to the input 21 cm signal power spectrum, $P_{\rm signal}(k)$, for Kernel PCA using different kernels. Results are shown for the PSM (upper panel) and GSM (lower panel).

4.2 Effect of the data cube resolution

The results from the previous section are for a 60 by 60 degree region of the sky projected on to a Cartesian grid of dimensions 150 by 150 pixels, which provides a pixel size of 0.4°. Wang et al. (2021) use a pixel size of ∼0.3° for the MeerKLASS survey, as this is a third of the average beam full width at half-maximum (FWHM). In this section, we investigate the effect of the pixel resolution on the foreground cleaning techniques. The simulations continue to be convolved with the MeerKAT L-band beam, but for these results the Cartesian grid is changed from 150 pixels squared to 100 (0.6 degree pixel size) and then 200 (0.3 degree pixel size). Note that the simulation resolution is specified in three dimensions, and so by changing the number of angular pixels we also change the number of frequency channels e.g. from 150 to 100 or 200. As such, we are really testing the dependence of the KPCA method on the spatial *and* spectral resolution here. It should also be noted that changing the frequency bandwidth also alters the level of thermal noise, increasing it for lower resolutions and decreasing it for higher resolutions. Even at our lowest resolution of 100 frequency channels, the thermal noise power level is still below the smoothed 21 cm power level for $k < 0.1$ Mpc^{-1} however.

Fig. 10 explores the effect of the ϵ and c hyperparameters on the power spectrum results for both the 100 and 150 pixel resolution simulations using the PSM foregrounds. ϵ values of 7, 9, and 11 (with $c = 0.9$) are shown alongside results where we fix $\epsilon = 7$ and vary c with values 0.8, 0.9, and 1.0. For the resolution 150 case, the choice of $\epsilon = 10$ is optimal, but for the lower resolution simulation, decreasing ϵ (increasing γ) improves the performance of Kernel PCA. We find the optimum ϵ values to be 5, 10, and 20 for the 100, 150, and 200 pixel resolutions respectively. This is somewhat understandable; as ϵ is increased, γ is decreased, leading to a broader decision region in feature space (Shadeed et al. 2020). As more pixels are present to describe the same spatial and spectral information in the higher resolution cases, one can imagine a broader hypersurface being drawn around these data points in feature space. The same response to increasing ϵ is seen for the GSM foregrounds as for the PSM foregrounds, so we do not present the additional plots for the GSM case.

Downloaded from https://academic.oup.com/mnras/article/508/3/3551/6381718 by guest on 16 April 2024

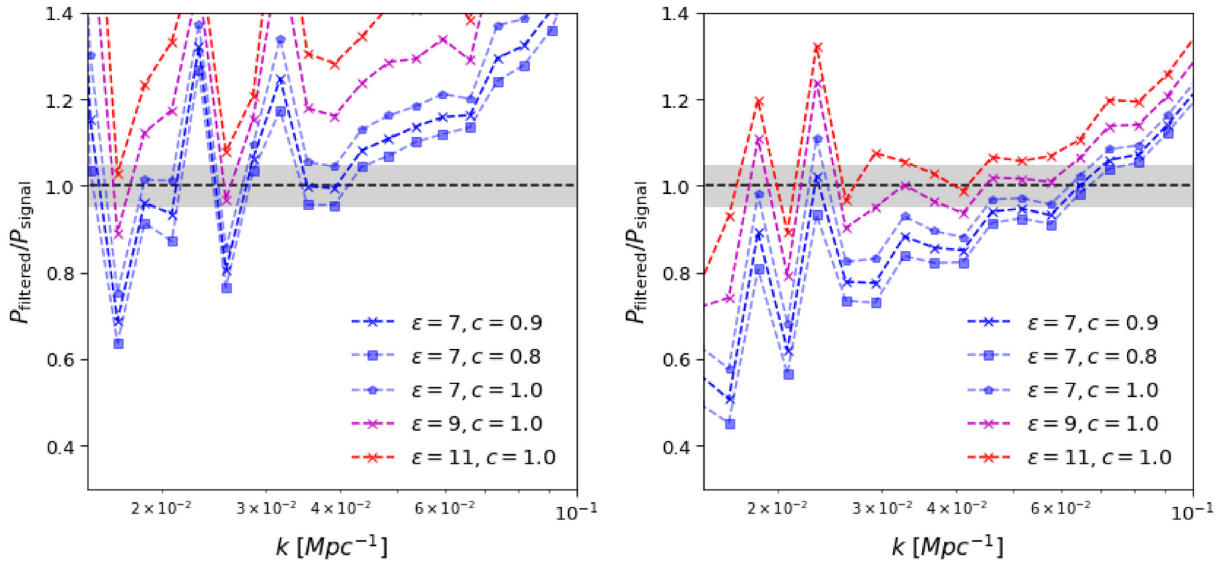

Figure 10. Ratio of recovered 1D power spectrum after foreground-filtering the data, $P_{\rm filtered}(k)$, to the input 21 cm signal power spectrum, $P_{\rm signal}(k)$, for Kernel PCA using different values of ϵ and c. The resolution 100 results are shown in the left-hand panel while the resolution 150 results are shown on the right. Both plots are for the case of PSM foregrounds.

Table 2. The mean absolute deviation from unity of $P_{\rm filtered}/P_{\rm signal}$ for each foreground cleaning method used in the cases of GSM and PSM foreground simulations. The average is taken over the k-bins in the range specified. The 'Benefit' column highlights whether or not any advantage over PCA was seen from the use of KPCA. Note that a 2-mode PCA pre-cleaning step has been applied to the data only for the KPCA case.

FG model	**Resolution**	***k* range (Mpc^{-1})**	PCA ($n = 3$)	ICA ($n = 3$)	KPCA	Sigmoid kernel params.	Benefit
PSM	100	(0.015, 0.058)	0.14	0.14	0.14	$n = 5, \epsilon = 5, c = 0.9$	=
		(0.058, 0.1)	0.20	0.20	0.18		✓
	150	(0.015, 0.058)	0.17	0.17	0.10	$n = 5, \epsilon = 10, c = 0.9$	✓
		(0.058, 0.1)	0.11	0.11	0.10		✓
	200	(0.015, 0.058)	0.07	0.07	0.05	$n = 5, \epsilon = 20, c = 0.9$	✓
		(0.058, 0.1)	0.07	0.07	0.09		✗
GSM	100	(0.015, 0.058)	0.14	0.14	0.44	$n = 7, \epsilon = 5, c = 0.9$	✗
		(0.058, 0.1)	0.30	0.30	0.87		✗
	150	(0.015, 0.058)	0.14	0.14	0.09	$n = 7, \epsilon = 10, c = 0.9$	✓
		(0.058, 0.1)	0.19	0.19	0.25		✗
	200	(0.015, 0.058)	0.08	0.08	0.05	$n = 7, \epsilon = 20, c = 0.9$	✓
		(0.058, 0.1)	0.11	0.11	0.17		✗

The combination of the effects of ϵ and c is particularly interesting, as it can be seen that through careful manipulation of both these parameters, an increasingly faithful representation of the 21 cm signal 1D power spectrum can be obtained. This is both an advantage and disadvantage of KPCA – the hyperparameters allow for a more sensitive tuning of the component separation than would be allowed by linear PCA for example, which enables KPCA to perform between 5 per cent and 10 per cent better than linear PCA (except at the smallest scales). Non-optimal hyperparameters lead to an equivalent performance between KPCA and linear PCA however, and so the ultimate effectiveness of the method depends on the tuning that has been applied.

Table 2 presents the mean absolute deviation (about the fiducial value of unity) of $P_{\rm filtered}/P_{\rm signal}$ for each foreground cleaning method, where the average is taken over k. The lower the deviation, the more reliably effective the method is at recovering the 21 cm signal on average. The deviation is calculated for two ranges of k to highlight the difference in behaviour as a function of scale. The hyperparameters used for each scenario are stated alongside the results.

For the PSM simulation, KPCA is seen to offer an improvement over PCA and ICA at all resolutions and angular scales except for the smallest scales at the highest resolution tested (200 pixels). For the GSM simulation however, KPCA is only seen to outperform linear PCA and ICA at large scales for the 150 and 200 pixel resolution simulations, and not across any scales for the 100 pixel resolution simulation.

Fig. 11 shows the cylindrically averaged 2D power spectrum ratio between the recovered and true signal for PCA (top panel) and KPCA (bottom panel) in the case of the GSM foregrounds at 150 pixel resolution. In the 2D power spectrum case, $k_\perp$ is equal to $\sqrt{k_x + k_y}$ and represents angular scales, while $k_\parallel$ corresponds to the

Downloaded from https://academic.oup.com/mnras/article/508/3/3551/6381718 by guest on 16 April 2024

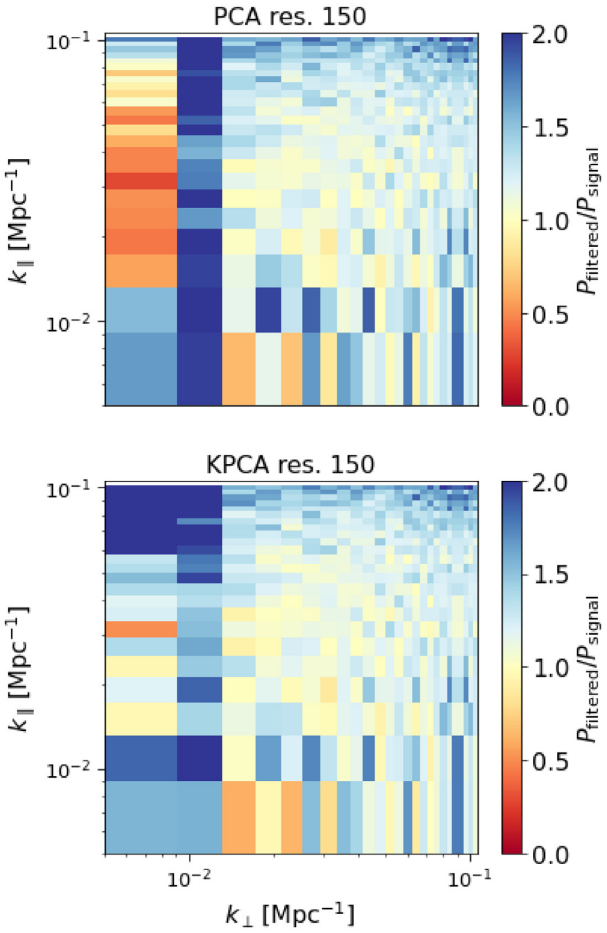

Figure 11. Ratio of recovered 2D power spectrum after foreground-filtering the data using KPCA or PCA, $P_{\rm filtered}(k_\perp, k_\parallel)$, to the input 21 cm power spectrum, $P_{\rm signal}(k_\perp, k_\parallel)$, for the resolution 150 case (GSM foregrounds). Blue regions denote an excess of power, implying the presence of residual foregrounds, while red regions imply over-subtraction has occurred.

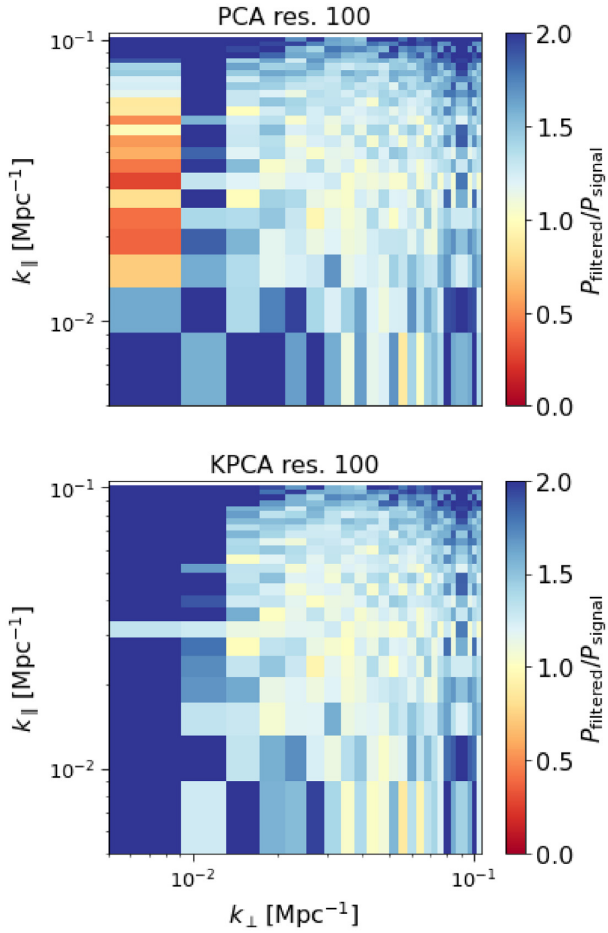

Figure 12. The same quantities as in Fig. 11, but now for resolution 100.

frequency/redshift direction. The 3D band powers have been binned into 25 equally spaced bins in $k_\perp$ between 0.005 and 0.11 Mpc^{-1}, and the contributing $k_\parallel$ powers have been limited to those between 0.005 and 0.11 Mpc^{-1}. Both methods can be seen to struggle to recover the 21 cm signal at the largest angular scales (smallest $k_\perp$ values) across all $k_\parallel$ values, with PCA showing the largest signal loss. However at 100 pixel resolution (Fig. 12), KPCA instead overestimates the 21 cm signal at the largest angular scales across all $k_\parallel$ values.

KPCA (with a Sigmoid kernel) does not appear to be well-equipped to cope with the additional spectral structure present in the GSM simulation when the data are under-sampled (a pixel size of 0.6 degrees for an average beam FWHM of 1 degree is only one pixel per beam), or for any resolution at small scales (high k values). Note, however, that the polynomial kernel case shown in the bottom panel of Fig. 9 does appear to perform slightly better on small scales for the GSM-based simulations, and so better performance on small scales can likely be attained with a different choice of kernel and hyperparameters.

4.3 Performance in the presence of a mask

Incomplete data coverage, e.g. due to point source masking and scanning strategies that lead to irregular survey areas, is an important consideration when performing foreground removal. In this section, we introduce a mask, shown in Fig. 13, to mimic the effects of masking bright point sources and a non-square scanning strategy. We construct the mask under the assumption that the Cartesian projection at each frequency is in fact the measured temperature map, i.e. ignoring the distortions to the mask that would be introduced by performing a projection from spherical coordinates. The mask is applied to each observed frequency, after which we reapply our foreground cleaning techniques. To soften the sharp features introduced in the data by masking, we in-paint the masked regions using the average values of their nearest-neighbour pixels.

In Fig. 14, we plot the change in the ratio of the recovered-to-true 1D power spectra from filtering the full observational data to filtering the masked data set. At the smallest k values, for the PSM foregrounds only, PCA can be seen to perform substantially worse in the presence of the mask, more so than KPCA. Masking and in-painting the data degrades the accuracy of the 21 cm estimates produced by all the techniques tested, but no single method is worse affected by the inclusion of a mask across all scales. The same conclusion can be drawn from both the PSM and GSM foreground simulations. Arguably a more sophisticated method of in-painting than just a nearest-neighbour average could have been used, but the only purpose of this test was to ensure that masking would not degrade the performance of KPCA more so than it would PCA/ICA.

Downloaded from https://academic.oup.com/mnras/article/508/3/3551/6381718 by guest on 16 April 2024

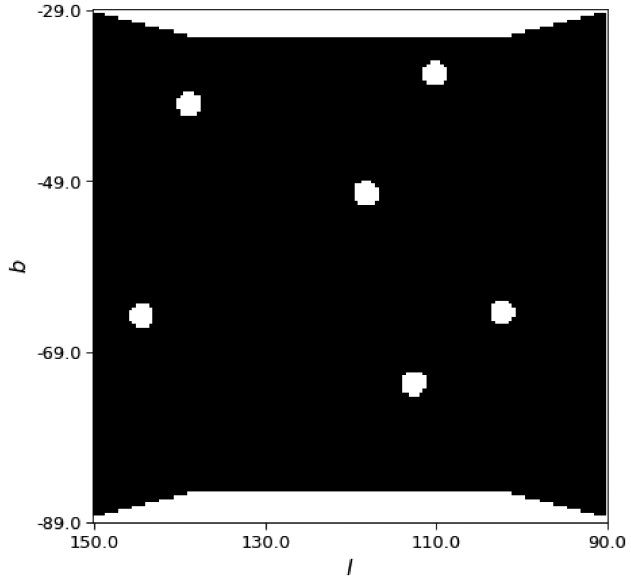

Figure 13. Point source and scan strategy mask applied at each frequency, shown in Galactic coordinates. The mask is intended to model the excision of a few bright point sources in the field (with masked regions of order the beamwidth), and irregular survey edges due to the choice of scan strategy.

4.4 Smoothing to a common resolution

A technique is sometimes used to mitigate the effects of the frequency-dependent beam, its sidelobe structure, and polarization leakage, and to smooth the data at all frequencies to a common angular resolution. For example, Switzer et al. (2013) used a common resolution of 1.4 times the largest beam size in their analysis. Since the beam is typically largest at the lowest frequencies, this degrades the angular resolution of the data at higher frequencies, losing some information from the 21 cm field in the transverse direction at lower redshifts. In some applications, this loss is acceptable given the additional mitigation of systematic effects, whereas in others (e.g. BAO recovery from the correlation function; Avila et al. 2021; Kennedy & Bull 2021), any amount of additional angular smoothing is undesirable.

To investigate the effect of smoothing on the performance of KPCA, we smoothed the simulated data to a common angular resolution of 2.25°, which corresponds to 1.4× the maximum beam size (the beam FWHM ranges from approximately 1.24−1.61° in our simulations). Practically, this entails taking the combined 21 cm and foreground datacubes that have already been convolved with the frequency-dependant KATBEAM beams, plus noise, and then convolving the map in each frequency channel with a Gaussian beam of FWHM $= \sqrt{(2.25^\circ)^2 - \theta^2_{\rm approx}}$, where $\theta_{\rm approx} \approx 1.2\lambda/(13.5{\rm m})$ is the FWHM at each frequency given by the Gaussian approximation for the MeerKAT dishes. This procedure accounts for the existing degree of smoothing at each frequency, and then applies whatever additional smoothing is necessary to reach the common angular resolution. The Gaussian approximation for the existing beam smoothing does not take into account features, such as sidelobes however, and so the resulting 'effective' beam after smoothing will not necessarily be completely frequency-independent.

An $n = 2$ PCA pre-cleaning step was used for KPCA, as before, and the reference PCA and ICA methods continued to remove $n = 3$ modes. The hyperparameters of KPCA, however, needed to be readjusted to perform optimally in this new set-up. For the original resolution PSM and GSM simulation data, the best-performing hyperparameters of the Sigmoid kernel were $n = 5$, $\epsilon = 10$, and $c = 0.9$ and $n = 7$, $\epsilon = 10$, and $c = 0.9$, respectively. For the smoothed data, the hyperparameters were set to $n = 3$, $\epsilon = 5$, and $c = 0.9$ for the PSM foreground simulations. The reduction in the number of modes removed, n, and the kernel width scaling factor, ϵ, reflects the simpler spectral structure of the data due to the smoothing. For the GSM foregrounds, however, the Sigmoid kernel was no longer the optimal choice, and so we switched to a polynomial kernel with hyperparameters $n = 7$, $\epsilon = 10$, $c = 0.9$, and $d = 4$, which provided a better 21 cm power spectrum recovery.

Fig. 15 shows the 1D power spectrum recovered-to-true ratio for both the PSM and GSM foreground simulations where the 'observed' data have been smoothed to a common resolution of 2.25°. The power spectrum ratios for the original resolution data, previously shown in Fig. 6, have also been re-plotted for comparison.

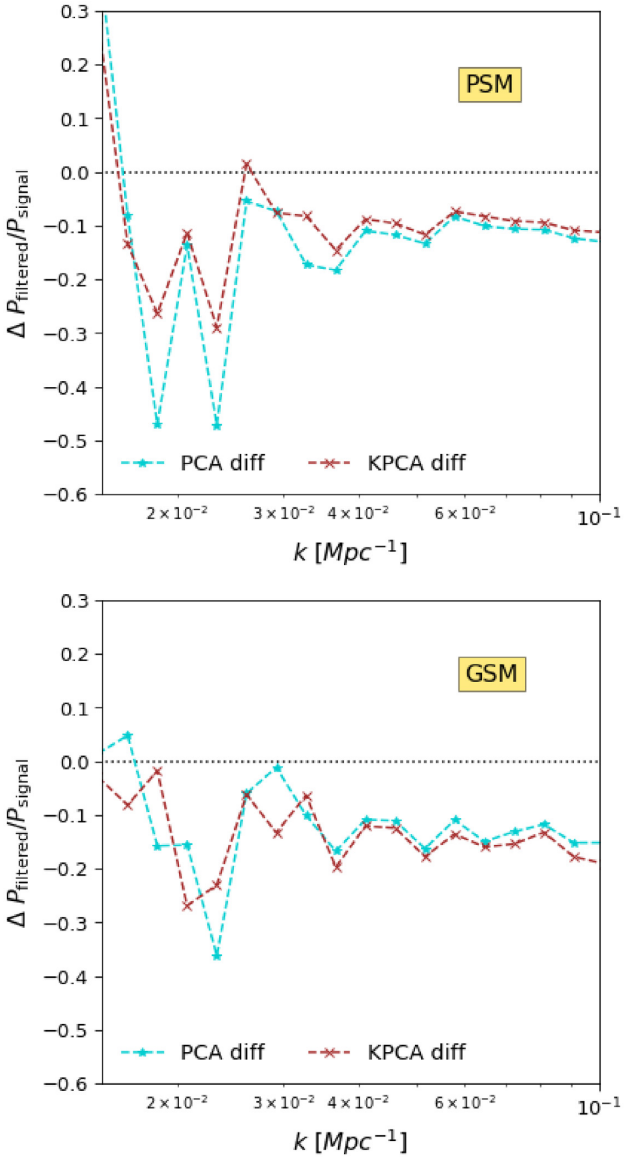

Figure 14. Change in the ratio of the recovered-to-true 1D power spectrum between foreground-filtering the masked data and the full data. Results are shown for the PSM (upper panel) and GSM (lower panel). A black dotted zero-line has been added to guide the eye.

Downloaded from https://academic.oup.com/mnras/article/508/3/3551/6381718 by guest on 16 April 2024

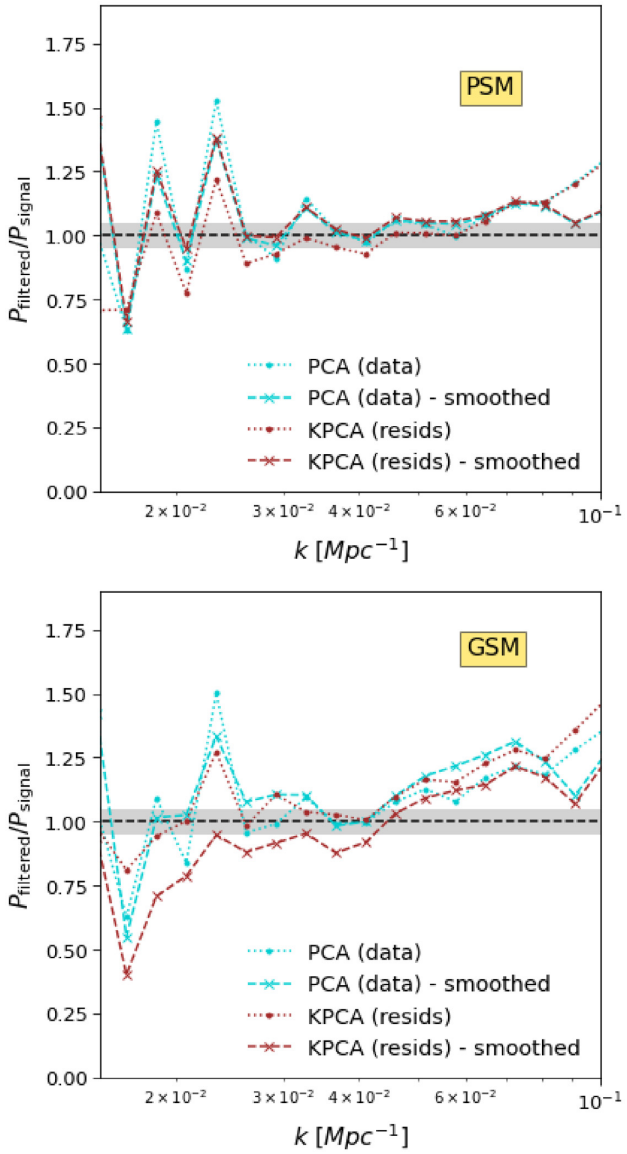

Figure 15. Ratio of recovered 1D power spectrum after foreground-filtering the data, $P_{\rm filtered}(k)$, to the input 21 cm signal power spectrum, $P_{\rm signal}(k)$ for the data at their original resolution and also in the case where the data have been smoothed to a common resolution. Results are shown for the PSM (upper panel) and GSM (lower panel) simulations.

For the PSM simulations, it can be seen that the performance of PCA/ICA improves slightly after smoothing, particularly on large scales, whereas the performance of KPCA degrades slightly, also mostly on large scales. In fact, the PCA/ICA and KPCA methods behave almost identically after smoothing, suggesting that whatever information KPCA was taking advantage of has been removed by the smoothing step. The differences are relatively minor however, in-keeping with results from previous sections that show that KPCA typically offers only moderate improvements over PCA.

The picture is slightly more nuanced in the case of the GSM foregrounds. PCA/ICA do slightly worse after smoothing, particularly on smaller scales, although there are some differences in behaviour on large scales too. Conversely, KPCA (now with a polynomial kernel) performs slightly better on small scales after smoothing, but worse on large scales. The PCA/ICA and KPCA results are not now almost identical after smoothing, as they were for the PSM foreground model, which may be due to either the additional spectral structure in the GSM model or the different choice of KPCA kernel.

In either case, it is clear that smoothing has an impact on the ability of KPCA to recover the 21 cm signal, tending to reduce its advantage compared with PCA/ICA, albeit in a scale-, kernel-, and foreground model-dependent manner. In contrast, smoothing can help improve the performance of PCA/ICA. This suggests that applications of these methods to real data should use unsmoothed data for KPCA and smoothed data for PCA/ICA. (Also recall our discussion of resolution effects in Section 4.2.)

4.5 Cross-correlations with galaxy surveys

Whilst the detection of the 21 cm signal from auto-power spectra relies heavily on the quality of the foreground cleaning, the use of cross-correlations with galaxy catalogues is a way to break this dependence (Cunnington et al. 2019; Padmanabhan, Refregier & Amara 2020). Optical galaxy surveys resolve individual galaxies, providing a high-resolution counterpart to intensity maps in overlapping volumes. To determine the impact of foreground cleaning on the 21 cm-galaxy cross-correlation power spectra, we also calculate the cross-power between the foreground-cleaned data cubes and the mock galaxy catalogue in the same simulated volume described in Section 3.4. We use the same kernel/hyperparameter combinations as in Fig. 6, with the PCA method applied directly to the data and the KPCA method applied to the pre-cleaning residuals.

The cross-correlation power spectra are computed between fluctuations in the 21 cm brightness temperature and the galaxy number counts. The top panels of Figs 16 and 17 show the 1D cross- and auto-correlation power spectra for our foreground cleaned maps for the cases of the PSM and GSM foregrounds. The 21 cm signal (convolved with the MeerKAT beam) auto-correlation spectrum is shown for reference, as is the 21 cm-galaxy cross-correlation spectrum and the auto-correlation power spectrum results first shown in Fig. 6. The bottom panels show the cross-correlation coefficient between the foreground cleaned maps and the galaxy sample data, defined as:

$$r = \frac{P_{\times}(k)}{\sqrt{P_{21\rm cm}(k)\,P_{\rm halo}(k)}}, \tag{14}$$

where $P_{\times}(k)$ is the cross-correlation power spectrum for the foreground cleaned map and the galaxy sample data, $P_{21\,\rm cm}(k)$ is the auto-correlation power spectrum for the 21 cm signal, and $P_{\rm halo}(k)$ is the auto-correlation power spectrum for the galaxy sample data. For the case of PSM foregrounds, the cross-correlation power spectra from the KPCA and PCA recovered 21 cm signal are quite similar; there is no clear advantage from choosing one method over the other, except for a slight reduction in signal loss from PCA on the largest scales. For the case of GSM foregrounds, PCA is also seen to sustain slightly less signal loss than KPCA at the smallest k values in cross-correlation. It is likely that the signal loss in the cross-power for the KPCA method could be reduced, e.g. by choosing hyperparameters that increase the residual foregrounds in the auto-power.

5 CONCLUSIONS

Foreground emission is the major source of systematic errors in 21 cm intensity maps, both as a contaminant in its own right, and through difficult-to-model interactions with the spectrally complex instrumental response. If foregrounds can be cleaned from the data efficiently enough, 21 cm auto-power spectrum analyses can be conducted in a manner analogous to spectroscopic galaxy surveys

Downloaded from https://academic.oup.com/mnras/article/508/3/3551/6381718 by guest on 16 April 2024

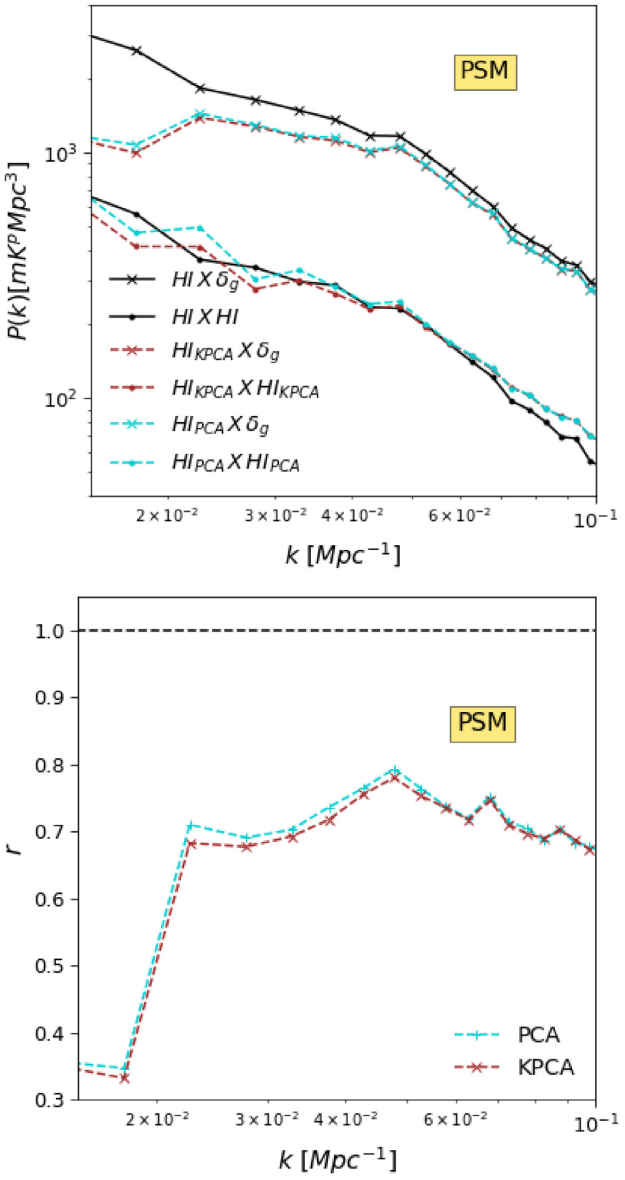

Figure 16. 1D cross- and auto-correlation power spectra (upper panel) and cross-correlation coefficient (lower panel) between the KPCA/PCA foreground-cleaned maps and the galaxy sample data for the PSM simulation. The y-axis units in the upper panel (mK^p) are $p = 1$ for the cross-correlation spectra and $p = 2$ for the auto-correlation spectra.

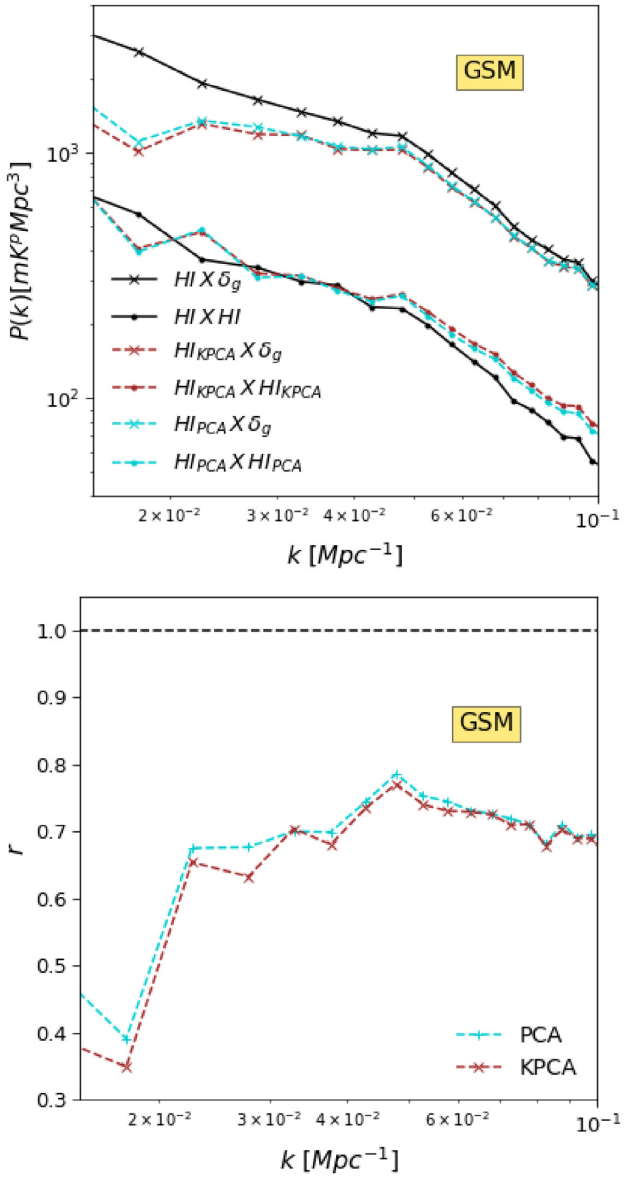

Figure 17. The same quantities as in Fig. 16, but now for the GSM foregrounds.

(e.g. Avila et al. 2021; Kennedy & Bull 2021; Soares et al. 2021b) for example, enabling a wide variety of cosmological applications for this tracer (e.g. Chang et al. 2008; Bull et al. 2015). Foregrounds are also important in 21 cm-galaxy cross-correlation analyses (e.g. Villaescusa-Navarro et al. 2015; Cunnington et al. 2019), as while in principle they are (mostly) uncorrelated with galaxies, the additional variance in the data contributed by the foregrounds significantly increase the noise on the cross-power spectrum if they are not at least partially removed.

A wide variety of methods exist to remove foreground emission, each using different ways to separate the 21 cm signal and the foregrounds, such as according to: signal-to-noise ratio, spectral smoothness, statistical independence, angular structure, sparsity, localization in other bases and so on. With all currently known methods, there are trade-offs to be made between the efficiency of foreground removal and the risk of over-subtracting, such that some of the cosmological signal is lost too. The range of (Fourier) scales over which the signal can be reliably recovered is the key metric for cosmological analyses, but the sensitivity of the method to prior assumptions about the structure of the data, instrumental effects etc. must also be taken into account. Methods that work extremely well on simulated data may struggle with inevitably more complex real-world data, particularly if those methods have been extensively optimized under particular simulated model assumptions.

It therefore behooves us to apply a range of different foreground removal methods to real-world data sets, in the hope that consistent answers are obtained from methodologically very different approaches, thus improving our confidence in the analysis. A lack of convergence can hint at the presence of unmitigated systematic effects instead (Switzer et al. 2013; Wolz et al. 2017a).

In this spirit, we have investigated a method that is new to this domain – Kernel PCA (KPCA). Unlike some other methods,

Downloaded from https://academic.oup.com/mnras/article/508/3/3551/6381718 by guest on 16 April 2024

KPCA seeks to represent the data in a *higher*-dimensional space than the data itself, using various non-linear combinations of the original dimensions of the data. This ultimately allows the data to be split into components with more complicated functional forms than allowed by purely linear decomposition methods like PCA, when said higher-dimensional functions are projected back down on to the original dimensions of the data. Thanks to the 'kernel trick', the computational expense of constructing the KPCA model is limited, as the full higher-dimensional representation of the data is never needed. The allowed functional forms and their weightings are determined by the choice of kernel and its hyperparameters, which can be selected to promote structure such as localization, radial symmetry and so forth. While KPCA is in principle a blind method, as with other methods (e.g. Gaussian Process Regression), the need to choose the kernel/hyperparameters means that some form of tuning can be applied that may benefit from knowledge of the data.

We have applied KPCA to simulations that incorporate a lognormal 21 cm signal model, noise, instrumental effects (e.g. beams with sidelobes), and a choice of two foreground models with substantially different model assumptions. We have studied the performance of KPCA in several scenarios, with a variety of choices of kernel and hyperparameters. In all cases, we have compared its efficacy to the more established (linear) PCA and ICA methods, which we have used as a benchmark. Our findings about KPCA as a foreground removal method are as follows:

(i) Kernel PCA typically performs better than PCA or ICA in terms of auto-power spectrum recovery on intermediate Fourier scales when an initial pre-cleaning step is performed (e.g. first using PCA with a small number of modes to reduce signal loss).

(ii) While it reliably improves on PCA for the PSM foregrounds, KPCA behaves more erratically when applied to the GSM foregrounds, likely due to difficulties in handling the additional spatial/spectral structure that is present. KPCA also does worse than PCA with under-sampled (lower-resolution) data.

(iii) KPCA does not display the same 21 cm signal loss as PCA at low values of $k_\perp$, though for the GSM foregrounds it leaves significant foreground contamination on these large angular scales.

(iv) The choice of kernel has a substantial impact on the effectiveness of the KPCA method. We found the Sigmoid kernel to have the best properties overall, e.g. reducing the degree of signal loss on large scales. The method also depends on the choice of hyperparameters, but in a relatively smooth and controlled way.

(v) KPCA is robust to missing data and masking; the presence of a mask slightly degrades every method we studied, but in broadly the same way for each of them.

(vi) Smoothing each frequency channel to a common beam resolution (e.g. to average out small-scale beam effects) is generally helpful for PCA, but slightly degrades the performance of KPCA. As with the dependence on resolution, this suggests that KPCA is (usefully) using information on small scales.

(vii) Turning to cross-correlations with galaxy surveys, the advantages KPCA has over PCA/ICA largely vanish. For the auto-power spectrum, KPCA is significantly better than PCA in several cases, but in cross-correlation they are both similar, with less signal loss observed for PCA on large scales.

In conclusion, we find that the KPCA method has substantially different behaviour from the (linear) PCA and ICA methods when applied to the recovery of the 21 cm auto-power spectrum. This makes it a useful alternative, e.g. for comparison studies used to verify the robustness of results to foreground removal. The KPCA method performs better than PCA and ICA in most (but not all) scenarios, specifically when the Sigmoid kernel is used. The kernel hyperparameters can be tuned using simulations to optimize performance, and the best choice of parameters will depend on the properties of the data (e.g. its resolution). The method is robust to masks and missing data, and does not require the data to be smoothed to a common resolution, as in fact it appears to benefit from the presence of small-scale information. KPCA performs almost equally, as well as PCA or ICA when applied to the recovery of the galaxy 21 cm cross-power spectrum, however, further underlining the expected robustness of cross-correlation analyses.

ACKNOWLEDGEMENTS

We are grateful to S. Paul, P. S. Soares, M. Spinelli, and E. R. Switzer for useful discussions. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 948764). M. I. acknowledges support from the South African Radio Astronomy Observatory and National Research Foundation (Grant No. 84156). We acknowledge use of the following software: CORNER (Foreman-Mackey 2016), HEALPIX (Gorski et al. 2005), matplotib (Hunter 2007), NBODYKIT (Hand et al. 2018), NUMPY (van der Walt, Colbert & Varoquaux 2011), SCIKIT-LEARN (Pedregosa et al. 2011), and SCIPY (Virtanen et al. 2020).

DATA AVAILABILITY

The PYTHON code used to produce the results in this paper is available from https://github.com/philbull/FastBox. The notebooks and simulated data used in this paper will be made available on reasonable request to the corresponding author.

REFERENCES

Agrawal A., Makiya R., Chiang C.-T., Jeong D., Saito S., Komatsu E., 2017, J. Cosmol. Astropart. Phys., 10, 003
Alam M. A., Fukumizu K., 2014, J. Comput. Sci., 10, 1139
Alonso D., Bull P., Ferreira P. G., Santos M. G., 2015, MNRAS, 447, 400
Alonso D., Ferreira P. G., Santos M. G., 2014, MNRAS, 444, 3183
Anderson C. J. et al., 2018, MNRAS, 476, 3382
Asad K. M. B. et al., 2021, MNRAS, 502, 2970
Avila S., Vos-Ginés B., Cunnington S., Stevens A. R. H., Yepes G., Knebe A., Chuang C.-H., 2021, preprint (arXiv:2105.10454)
Bennett C. L. et al., 1992, ApJ, 396, L7
Bobin J., Moudden Y., Starck J. L., Fadili J., Aghanim N., 2008, Stat. Methodol., 5, 307
Bull P., Ferreira P. G., Patel P., Santos M. G., 2015, ApJ, 803, 21
Carucci I. P., Irfan M. O., Bobin J., 2020, MNRAS, 499, 304
Chang T.-C., Pen U.-L., Bandura K., Peterson J. B., 2010, Nature, 466, 463
Chang T.-C., Pen U.-L., Peterson J. B., McDonald P., 2008, Phys. Rev. Lett., 100, 091303
Chapman E. et al., 2013, MNRAS, 429, 165
Chisari N. E. et al., 2019, ApJS, 242, 2
Chuang C.-H. et al., 2015, MNRAS, 452, 686
Coles P., Jones B., 1991, MNRAS, 248, 1
Cunnington S., Irfan M. O., Carucci I. P., Pourtsidou A., Bobin J., 2021, MNRAS, 504, 208
Cunnington S., Wolz L., Pourtsidou A., Bacon D., 2019, MNRAS, 488, 5452
Foreman-Mackey D., 2016, J. Open Sour. Softw., 1, 24
Ghosh A. et al., 2020, MNRAS, 495, 2813
Ghosh A., Bharadwaj S., Ali S. S., Chengalur J. N., 2011, MNRAS, 418, 2584
Gorski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelman M., 2005, ApJ, 622, 759

Downloaded from https://academic.oup.com/mnras/article/508/3/3551/6381718 by guest on 16 April 2024

Hand N., Feng Y., Beutler F., Li Y., Modi C., Seljak U., Slepian Z., 2018, AJ, 156, 160
Hunter J. D., 2007, Comput. Sci. Eng., 9, 90
Kennedy F., Bull P., 2021, MNRAS, 506, 2638
Kern N. S., Liu A., 2021, MNRAS, 501, 1463
Masui K. W. et al., 2013, ApJ, 763, L20
Matshawule S. D., Spinelli M., Santos M. G., Ngobese S., 2021, MNRAS, 506, 5075
Mauch T. et al. 2020, ApJ, 888, 61
Mertens F. G., Ghosh A., Koopmans L. V. E., 2018, MNRAS, 478, 3640
Miville-Deschênes M. A., Ysard N., Lavabre A., Ponthieu N., Macías-Pérez J. F., Aumont J., Bernard J. P., 2008, A&A, 490, 1093
Olivari L. C., Remazeilles M., Dickinson C., 2016, MNRAS, 456, 2749
Padmanabhan H., Refregier A., Amara A., 2020, MNRAS, 495, 3935
Pedregosa F. et al., 2011, J. Mach. Learn. Res., 12, 2825
Planck Collaboration XII, 2016, A&A, 594, A12
Pourtsidou A., Bacon D., Crittenden R., Metcalf R. B., 2016, MNRAS, 459, 863
Remazeilles M., Delabrouille J., Cardoso J.-F., 2011, MNRAS, 418, 467
Santos M. G. et al., 2017, preprint (arXiv:1709.06099)
Schölkopf B., Smola A., Müller K.-R., 1997, International Conference on Artificial Neural Networks. Springer-Verlag, Berlin, Heidelberg, p. 583
Shadeed I., Alwan J., Abd D., 2020, International Journal of Electrical and Computer Engineering (IJECE), Vol. 10, Institute of Advanced Engineering and Science, Yogyakarta, Indonesia, p. 5497
Shaw J. R., Sigurdson K., Sitwell M., Stebbins A., Pen U.-L., 2015, Phys. Rev. D, 91, 083514
Soares P. S., Cunnington S., Pourtsidou A., Blake C., 2021b, MNRAS, 502, 2549
Soares P. S., Watkinson C. A., Cunnington S., Pourtsidou A., 2021, MNRAS, Available at: https://doi.org/10.1093/mnras/stab2594
Switzer E. R. et al., 2013, MNRAS, 434, L46
Switzer E. R., Chang T. C., Masui K. W., Pen U. L., Voytek T. C., 2015, ApJ, 815, 51
van der Walt S., Colbert S. C., Varoquaux G., 2011, Comput. Sci. Eng., 13, 22
Villaescusa-Navarro F., Viel M., Alonso D., Datta K. K., Bull P., Santos M. G., 2015, J. Cosmol. Astropart. Phys., 03, 034
Virtanen P. et al., 2020, Nat. Methods, 17, 261
Wang J. et al., 2021, MNRAS, 505, 3698
Wolz L. et al., 2021, MNRAS, preprint (arXiv:2102.04946)
Wolz L., Abdalla F. B., Blake C., Shaw J. R., Chapman E., Rawlings S., 2014, MNRAS, 441, 3271
Wolz L. et al., 2017a, MNRAS, 464, 4938
Wolz L., Blake C., Wyithe J. S. B., 2017b, MNRAS, 470, 3220
Zhang D., Zhou Z.-h., Chen S., 2006, Sixth International Conference on Data Mining (ICDM'06). IEEE, Washington, DC, US, p. 1178
Zheng H. et al., 2017, MNRAS, 464, 3486

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

Downloaded from https://academic.oup.com/mnras/article/508/3/3551/6381718 by guest on 16 April 2024