



<b>Publication Year</b>	2021
<b>Acceptance in OA</b>	2025-03-04T11:33:04Z
<b>Title</b>	21cm foregrounds and polarization leakage: a user's guide on cleaning and mitigation strategies
<b>Authors</b>	Cunnington, Steven, Irfan, Melis O., CARUCCI, Isabella Paola, Pourtsidou, Alkistis, Bobin, Jérôme
<b>Publisher's version (DOI)</b>	10.1093/mnras/stab856
<b>Handle</b>	<a href="http://hdl.handle.net/20.500.12386/36402">http://hdl.handle.net/20.500.12386/36402</a>
<b>Journal</b>	MONTHLY NOTICES OF THE ROYAL ASTRONOMICAL SOCIETY
<b>Volume</b>	504

# 21-cm foregrounds and polarization leakage: cleaning and mitigation strategies

Steven Cunnington<sup>1</sup>,<sup>★</sup> Melis O. Irfan,<sup>2,3</sup> Isabella P. Carucci<sup>3</sup>, Alkistis Pourtsidou<sup>1,2</sup> and Jérôme Bobin<sup>3</sup>

<sup>1</sup>*School of Physics and Astronomy, Queen Mary University of London, Mile End Road, London E1 4NS, UK*

<sup>2</sup>*Department of Physics & Astronomy, University of the Western Cape, Cape Town 7535, South Africa*

<sup>3</sup>*AIM, CEA, CNRS, Université Paris-Saclay, Université Paris Diderot, Sorbonne Paris Cité, F-91191 Gif-sur-Yvette, France*

Accepted 2021 March 22. Received 2021 March 19; in original form 2020 October 19

## ABSTRACT

The success of H I intensity mapping is largely dependent on how well 21-cm foreground contamination can be controlled. In order to progress our understanding further, we present a range of simulated foreground data from two different  $\sim 3000$  deg<sup>2</sup> sky regions, with varying effects from polarization leakage. Combining these with cosmological H I simulations creates a range of intensity mapping test cases that require different foreground treatments. This allows us to conduct the most generalized study to date into 21-cm foregrounds and their cleaning techniques for the post-reionization era. We first provide a pedagogical review of the most commonly used blind foreground removal techniques [principal component analysis (PCA)/singular value decomposition (SVD), fast independent component analysis (FASTICA), and generalized morphological component analysis (GMCA)]. We also trial a non-blind parametric fitting technique and discuss potential hybridization of methods. We highlight the similarities and differences in these techniques finding that the blind methods produce near equivalent results, and we explain the fundamental reasons for this. Our results demonstrate that polarized foreground residuals should be generally subdominant to H I on small scales ( $k \gtrsim 0.1 h \text{ Mpc}^{-1}$ ). However, on larger scales, results are more case dependent. In some cases, aggressive cleans severely damp H I power but still leave dominant foreground residuals. We find a changing polarization fraction has little impact on results within a realistic range (0.5–2 per cent); however, a higher level of Faraday rotation does require more aggressive cleaning. We also demonstrate the gain from cross-correlations with optical galaxy surveys, where extreme levels of residual foregrounds can be circumvented. However, these residuals still contribute to errors and we discuss the optimal balance between overcleaning and undercleaning.

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

## 1 INTRODUCTION

Mapping the cosmic neutral hydrogen (H I) from the post-reionization era is as an excellent way to probe the large-scale structure of the Universe. By mapping the redshifted 21-cm signal from H I residing within galaxies, the underlying 3D matter density can be inferred and cosmological information can be extracted, in a similar fashion to optical galaxy surveys. A novel technique allowing to do this is intensity mapping (Bharadwaj et al. 2001; Battye, Davies & Weller 2004; Chang et al. 2008).

In this work, we focus on so-called *single-dish* intensity mapping (Battye et al. 2013), which uses the autocorrelation data of a telescope array [e.g. the Square Kilometer Array (SKA) – SKA Cosmology (SWG et al. 2020)], as opposed to the more traditional interferometric mode of operation. Unlike a conventional spectroscopic galaxy survey that has to resolve galaxies and conduct spectroscopy to infer a redshift with sufficient precision, intensity mapping does

not resolve galaxies but records the diffuse, unresolved H I. This has the advantages of being able to rapidly observe very large volumes of the Universe, and is not as susceptible to high levels of shot-noise. The resulting maps have a relatively low-angular resolution due to the radio telescope beam, which is related to the dish diameter for single-dish observations. This damps the H I power spectrum for modes perpendicular to the line of sight but despite this, many large cosmological scales can still be resolved. Furthermore, the spectroscopic resolution in these radio observations is excellent and thus modes can in principle be resolved to very small scales along the line of sight (Villaescusa-Navarro, Alonso & Viel 2017).

There are unique challenges to H I intensity mapping which conventional galaxy surveys largely avoid. Whilst intensity mapping is unlikely to be limited by shot noise, there is instrumental (thermal) noise. Assuming enough observation time, and a well controlled system temperature, this noise should be sub-dominant relative to the H I signal and well approximated as Gaussian white noise. As noted in Harper et al. (2018), complications from other systematics such as  $1/f$  noise pose a more complex challenge. However, analysis of recent H I intensity mapping data from MeerKAT suggest this

\* E-mail: s.cunnington@qmul.ac.uk

should be a controllable systematic (Li et al. 2020a). A further issue is contamination from human-made Radio Frequency Interference (RFI) such as global navigation satellites (Harper & Dickinson 2018).

Another major challenge, and the focus of this paper, is foreground contamination from astrophysical sources and how they interact with the telescope. The main source of 21-cm foreground signals comes from *Galactic synchrotron* (sourced by cosmic ray electrons accelerated by the Galactic magnetic field), *free-free emission* (sourced by free electrons scattering off ions largely within our Galaxy but weaker emission can also come from extragalactic sources), and *point sources* [extragalactic objects emitting strong radio signals e.g. active galactic nuclei (AGNs)].

Some of these foregrounds can be orders of magnitude more dominant than the H I signal, but their spectrum evolves slowly through frequency. This is in contrast to cosmic H I, which varies with redshift and thus oscillates to a near-Gaussian approximation with frequency. The fact that the raw foregrounds are smooth continuums through frequency means they can in principle be removed with modelling or source separation (Liu & Tegmark 2011; Wolz et al. 2014; Alonso et al. 2015; Shaw et al. 2015). However, large-scale foreground signals typically have some degeneracy with the cosmological H I, and require some form of treatment in order not to bias power spectra measurements and cosmological parameter estimation (Wolz et al. 2014; Cunnington, Camera & Pourtsidou 2020b; Soares et al. 2021).

In reality, the challenge of separating the H I signal from the foregrounds becomes even more complicated by the foreground's response to the instrument. Unless instrumental effects from spectral response and chromaticity from the beam are controlled, the spectral smoothness of the foregrounds can be degraded. The most potentially concerning instrumental effect is from polarization leakage (Jelic et al. 2008, 2010; Moore et al. 2013). Cosmological H I is unpolarized and thus attempts are made to sufficiently calibrate telescopes to avoid polarized signals (Liao et al. 2016). However, a sufficient level of calibration is not guaranteed and even a small amount of polarized synchrotron leaking into the observational data can dominate the H I signal. Furthermore, the Faraday rotation (FR) that interferes with the polarization state is expected to be frequency dependent, which means these leaked signals will not have such a smooth spectrum and will be harder to single out (Carucci, Irfan & Bobin 2020a).

Previous investigations into foreground cleaning generally involve introducing a single set of foreground simulations which cleaning techniques can then be tuned to. However, these rarely include instrumental response effects such as polarization leakage (although there are some exceptions e.g. Shaw et al. 2015; Carucci et al. 2020a). These idealized simulated foregrounds require much less aggressive cleans than what is usually needed in real data analyses from pathfinder intensity mapping experiments (Masui et al. 2013; Switzer et al. 2013; Wolz et al. 2017, 2021; Anderson et al. 2018).

In this work, we add an extra layer of complication. Whilst we cannot yet provide full end-to-end simulations that directly mimic a realistic experiment, we are able to present results where it is necessary to use aggressive foreground cleans akin to those employed in real data. By using two different sky regions, with varying polarization leakage effects, we create a variety of cases in which different levels of foreground cleaning are required and different problems arise. We introduce and apply a range of different foreground cleaning methods and compare the results. Since we are dealing with simulations, we have full control over the data and provide analysis into problems concerning damping of H I power as well as the biases and errors introduced from foreground residuals.

This paper is structured as follows. In Section 2, we present our foreground simulations, the sky regions we consider, and the different cases of polarization leakage. In Section 3, we present our method for producing the underlying cosmological H I intensity maps (and the accompanying optical galaxy data for cross-correlations) that we aim to recover. We provide a generalized review of foreground cleaning methods in Section 4 and identify the exact methods we apply to our simulated data. We then present our results in Section 5 and conclude in Section 6.

## 2 FOREGROUND SIMULATIONS

We begin by identifying two different regions on the sky. We desire each of our regions to have the same size which is dictated by the size of our  $1 h^{-3} \text{Gpc}^3$  H I simulation box (to be described in detail in Section 3). At the central redshift of this simulation ( $z = 0.39$ ) these dimensions are equivalent to a sky size of  $54.1 \times 54.1 \sim 2972 \text{deg}^2$ . This is similar to the sky area proposed in MeerKLASS (Santos et al. 2017), a wide area survey using the MeerKAT telescope, which is the pathfinder for the Square Kilometre Array (SKA)<sup>1</sup> (SKA Cosmology SWG et al. 2020). We choose a frequency range of 899–1184 MHz ( $0.2 < z < 0.58$ ), again consistent with a MeerKAT-like observation performed in the L band. The sky regions we investigate are:

- (i) **[1] Stripe 82:** A small, 300 deg<sup>2</sup> field imaged numerous times by galaxy surveys. To ensure consistency in sky sizes, this region is a 2927 deg<sup>2</sup> patch centred on the Stripe 82 field.
- (ii) **[2] South Celestial Pole (SCP):** Low declination region away from Galactic Plane where combined emission from all foregrounds is expected to be low.

These regions are outlined in Fig. 1 on the full sky (top-map) and individually shown in the flattened maps below. We assume the maps are sufficiently small in size that they can be projected on to a Cartesian grid with minimal distortion. For each of these regions, we simulate effects from polarization leakage, which we discuss further in Section 2.4.

The total observed temperature data are a combination of the cosmological H I signal, the foregrounds, and instrumental noise, all binned into pixels whose position is defined by  $\theta$  at each frequency channel  $\nu$

$$\delta T_{\text{obs}}(\nu, \theta) = \delta T_{\text{HI}}(\nu, \theta) + \delta T_{\text{FG}}(\nu, \theta) + \delta T_{\text{noise}}(\nu, \theta). \quad (1)$$

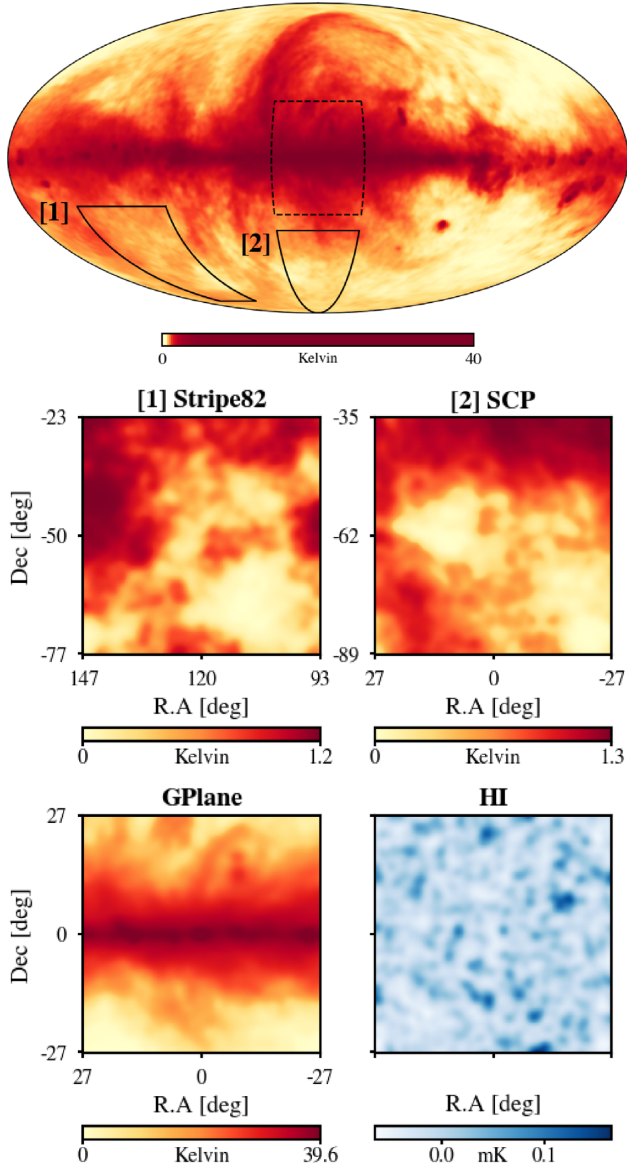
In this work, we will vary  $\delta T_{\text{FG}}$  between each region whilst keeping the other components fixed. The foreground signal can be further decomposed into the contributions from the different sources of foregrounds i.e. Galactic synchrotron emission, Galactic free-free emission, extragalactic point sources, and polarization leakage:

$$\delta T_{\text{FG}} = \delta T_{\text{syn}} + \delta T_{\text{free}} + \delta T_{\text{point}} + \delta T_{\text{pol}}. \quad (2)$$

We introduce our simulation approach for each of these components in this section (except the H I contribution that is discussed in Section 3). A full-sky realization of  $\delta T_{\text{FG}}$  is openly available in Carucci, Irfan & Bobin (2020b).

We chose the frequency range 899–1184 MHz and separate this range into 285 measurement bands. The effective resolution is determined by the beam size of the instrument which is dependent on the frequency  $\nu$  and therefore each channel is smoothed by a different amount. However, foreground removal algorithms (discussed in Section 4) perform better on data with a common resolution. We

<sup>1</sup>skatelescope.org



**Figure 1.** (Top map): Full sky simulated synchrotron, free–free emission, and point sources with an 80 arcmin resolution. The labelled black boards indicate the position of both sky regions we investigate. (Middle two maps): Both regions interpolated over a  $256^2$  pixel grid. For comparison, the bottom-left map shows the Galactic plane (position marked by the dashed outline in the top map) which is orders of magnitude higher in emission. Bottom-right is the cosmological H I (introduced in Section 3.1) which we overlay on to both regions [1] and [2] and attempt to recover. All maps are at 1050 MHz and include the effects from a 1.67 deg telescope beam.

therefore smooth the intensity maps to a constant beam size given by the minimum frequency  $\nu_{\min} = 899$  MHz. The full width half-maximum (FWHM) of the beam is given by

$$\theta_{\text{FWHM}} = \frac{1.18c}{\nu D_{\text{dish}}}. \quad (3)$$

where  $c$  is the speed of light and we assume a dish size of  $D_{\text{dish}} = 13.5$  m, which is the size of MeerKAT’s dishes and is approximately equivalent to the SKA-MID’s dishes too. The factor of 1.18 can vary depending on the beam pattern but we chose this value to be consistent with the recent MeerKAT investigations in Matshawule et al. (2020). From equation (3), we therefore get an

effective resolution for our maps of  $\theta_{\text{FWHM}} = 1.67$  deg. We chose to create our simulations at  $N_{\text{side}} = 2048$  and then to interpolate our  $54.1 \times 54.1$  patches on to  $256 \times 256$  pixel arrays.

We make use of the Planck Legacy Archive<sup>2</sup> FFP10 simulations within our simulations and, as they are given in  $T_{\text{CMB}}$ , the following conversion to the Rayleigh–Jeans regime is used:

$$T_{\text{RJ}} = \frac{x^2 e^x}{(e^x - 1)^2} T_{\text{CMB}}, \quad (4)$$

where  $x = h\nu/k_{\text{B}} T_{\text{CMB}}$ , with  $h$  the Planck constant and  $k_{\text{B}}$  the Boltzmann constant.

## 2.1 Simulated synchrotron emission

We use the FFP10 simulations of synchrotron emission at 217 and 353 GHz for our purposes as these maps are provided at  $N_{\text{side}} = 2048$ . These maps are formed from the source-subtracted and destriped 0.408 GHz map (Remazeilles et al. 2015). Despite the 0.408 GHz survey data having a resolution of 56 arcmin, Remazeilles et al. (2015) provide an  $N_{\text{side}} = 2048$  version of the data by filling in the higher resolution detail with a Gaussian random field.

These 217 and 353 GHz synchrotron maps can be used to determine the synchrotron spectral index map at  $N_{\text{side}} = 2048$ . The spectral index map used by FFP10 is the ‘Model 4’ synchrotron spectral index map of Miville-Deschênes et al. (2008), which has a resolution of  $\sim 5$  degrees. This map was formed from 0.408 GHz intensity data and 23 GHz polarization data. However, as we are simply trying to determine the accuracy of our foreground mitigation strategies, the accuracy of the synchrotron spectral index map does not come into consideration.

We will however, need a higher resolution view of the synchrotron spectral index than 5 deg and so we also choose to fill in the higher resolution detail with a Gaussian random field. Taking the synchrotron multipole scaling relation from Santos, Cooray & Knox (2005), our  $N_{\text{side}} = 2048$  synchrotron spectral index map is constructed as

$$\beta_{\text{sy}} = \beta_{\text{model4}} + \beta_{\text{ss}}, \quad (5)$$

where

$$C_{\ell}^{\beta_{\text{ss}}} = 7 \times 10^{-6} \left( \frac{1000}{\ell} \right)^{2.4} \left( \frac{\nu_r^2}{\nu_1 \nu_2} \right)^{2.8} \exp \left( \frac{-\log(\nu_1/\nu_2)^2}{2 \times 4^2} \right), \quad (6)$$

where  $\nu_r$  is 130 MHz,  $\nu_1$  is 580 MHz, and  $\nu_2$  is 1000 MHz. Our Gaussian random field is identical to that found in Santos et al. (2005) with the exception of the amplitude, which we alter to suit the magnitude of the synchrotron spectral index as opposed to the emission amplitude. We then smooth  $\beta_{\text{sy}}$  to 1.67 deg in order to match the desired resolution of our total simulation maps.

## 2.2 Simulated free-free emission

We take our simulated free–free amplitude ( $a_{\text{ff}}$ ) from the FFP10 217 GHz free–free simulation at  $N_{\text{side}} = 2048$ . This map is a composite of the Dickinson, Davies & Davis (2003) free–free template and the WMAP MEM free–free templates; the details of which can be found in (Miville-Deschênes et al. 2008). Our free–free emission is modelled by a power law

$$T_{\text{ff}}(\nu, \theta) = a_{\text{ff}}(\theta) \left( \frac{\nu}{\nu_0} \right)^{\beta_{\text{ff}}}, \quad (7)$$

<sup>2</sup>pla.esac.esa.int/pla

where the free-free spectral index is  $\beta_{\text{ff}} = -2.13$  and constant across all map pixels.

### 2.3 Simulated point sources

We use the empirical model of Battye et al. (2013), which fits a polynomial to a selection of radio source counts at 1.4 GHz. The specific details of assembling this model of the Poisson and clustering contributions at 1.4 GHz can be found in Olivari (2018). Following the method of Olivari et al. (2018), we then scale the 1.4 GHz point source map to our frequencies using a power law where the spectral index varies following a Gaussian distribution centred at  $-2.7$ , with a standard deviation of 0.2.

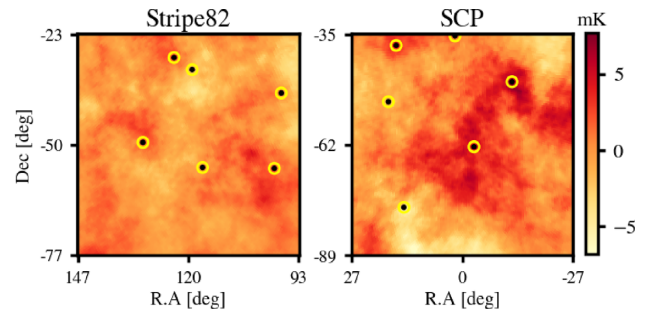
Battye et al. (2013) expect point sources over 10 mJy to be bright enough to be identified within the National Radio Astronomy Observatory Very Large Array Sky Survey (Lacy et al. 2020) and so, removed. In this work, we consider a 100 mJy upper bound on source extraction.

### 2.4 Simulated polarization leakage

The magnetic fields within our Galaxy’s interstellar medium can cause FR effects which change the polarization angles of light. If this were a consistent effect, it would not be hugely problematic for foreground classification. However, FR is a frequency-dependent effect as demonstrated by Jelic et al. (2010) and Moore et al. (2013). If any spectrally fluctuating polarization intensity is leaked into the total intensity it would be difficult to subtract without large loss to the unpolarized H1 cosmological modes. Depending on both the instrument and the data reduction scheme implemented, there will be some percentage of leakage of Stokes Q and U synchrotron emission into Stokes I. FR alters the true polarization angle of the Stokes Q/U signal such that this leakage will not remain constant across all the observational channels.

We simulate this instrumental effect with the use of the CRIME<sup>3</sup> software, which provides maps of Stokes Q emission at each frequency with a choice for the polarization leakage fraction, typically between 0.5 and 1 per cent (Liao et al. 2016). Further observational data analysis would be needed to constrain a reasonable choice for this fraction so we therefore explore a range of polarization leakage fractions. Details for the rotation calculation of the Stokes Q synchrotron emission from Faraday depth measurements (Oppermann et al. 2012) are given in Alonso, Ferreira & Santos (2014).

We highlight that the polarization leakage model, we use has many limitations. For example, it assumes that all the leakage comes from just the Q Stokes component and not U, reducing possible mixtures; it uses rotation measures of extragalactic sources (Oppermann et al. 2012), corresponding to the maximum rotation from a single source; it does not account for multiple Faraday rotating components along a single line of sight. These issues could give rise to a more complex structure for this systematic. However, at present, the community lacks a better model than what has been proposed by Alonso et al. (2014). Shaw et al. (2015) also assembled a polarization leakage model using the rotation measure map of Oppermann et al. (2012), however, showing a qualitatively different structure in pixel space. The two polarization leakage models differ in their choices of polarization fraction magnitude and correlation length in frequency space. Both models are valid within the current knowledge of intensity mapping polarization leakage. The lack



**Figure 2.** Polarization leakage maps from CRIME for the both regions averaged along the line of sight. The yellow-black markers indicate the positions of the random pixels used for each of the  $\delta T_{\text{pol}}$  line-of-sight spectra in the below Fig. 3.

of smoothness in frequency is what makes polarization leakage a challenging component to separate. Therefore, to be conservative, we decide to make use of our simulated polarization leakage both in its *milder* regions and the most troublesome one, where a higher FR causes increased decorrelation (or a lack of smoothness). For each of the two sky regions we look at, we therefore provide two polarization leakage cases; one predicted by the CRIME model for that region, and a second from the CRIME output for the Galactic plane (see dashed boarder in Fig. 1). The latter stronger model of polarization leakage, we refer to as our high-FR case.

Fig. 2 shows the output maps for the polarization leakage component from CRIME for our two regions. The SCP regions contain slightly higher amplitudes than Stripe82, however, as discussed above, the main problem that polarization leakage introduces is frequency decoherence. In Fig. 3, we show the spectra for some random lines of sight in the polarized maps (coloured-solid lines). We find the Stripe82 region contains more oscillating spectra in this polarization model and we therefore expect this to be the more troublesome regions in our foreground tests. We also include in Fig. 3 the spectra from random lines of sight in the high FR case we use (data from the Galactic plane), shown as coloured-dashed lines. We can see for this case, the oscillations are more extreme and when these are incorporated into the total observed signal they will cause more frequency decoherence and represent the most challenging case for foreground cleaning.

### 2.5 Simulated noise

In this work, only Gaussian noise is considered, with a zero mean and standard deviation of

$$\sigma(v) = T_{\text{sys}}(v) \left( \delta_\nu t_{\text{tot}} \frac{\Omega_p}{\Omega_a} N_{\text{dish}} \right)^{-1/2}, \quad (8)$$

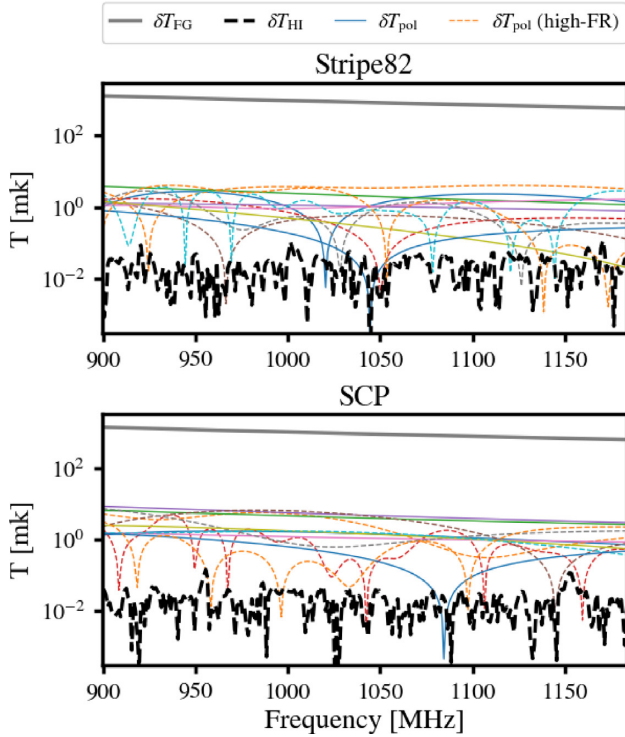
where  $\delta_\nu$  is the width of each frequency band (Hz),  $t_{\text{tot}}$  is the total survey time (s),  $N_{\text{dish}}$  is the number of dishes, and  $\Omega_{p/a}$  are the pixel and survey solid angle, respectively (Alonso et al. 2014). For the pixel solid angle only the beam FWHM expressed in radians is required

$$\Omega_p = 1.13 \theta_{\text{FWHM}}^2, \quad (9)$$

while for the survey solid angle the fraction of the sky covered is needed. If the angular area of the observed sky ( $A_{\text{sky}}$ ) is given in square degrees, we have

$$\Omega_a = 4\pi \frac{A_{\text{sky}}}{41253}. \quad (10)$$

<sup>3</sup>intensitymapping.physics.ox.ac.uk/CRIME.html



**Figure 3.** Spectra for the different components of the the observed signal. Grey-solid line shows the total foreground component (equation 2) for one random line-of-sight and black-dashed line shows HI-only. Smoothness of the foregrounds and their large amplitude relative to the HI is clear and are the distinguishing features utilized in a foreground clean. The solid-coloured-thin lines show six random lines of sight for the polarized contribution. The positions of these lines of sight in angular space are marked in Fig. 2. The polarized components cause a decoherence to the spectra creating difficulties in a foreground clean. We also show the polarized data taken from the Galactic plane (dashed-coloured-thin lines) which we refer to as a high FR case.

**Table 1.** The assumed receiver and survey properties for observation with bandwidth  $899 < \nu < 1184$  MHz ( $0.2 < z < 0.58$ ).

Quantity:	$\delta\nu$	$t_{\text{tot}}$	$N_{\text{dish}}$	$T_{\text{rec}}$	$A_{\text{sky}}$
Value:	1 MHz	1000 h	64	25 K	2927 deg <sup>2</sup>

The system temperature in each band ( $T_{\text{sys}}(\nu)$ ) is a combination of the receiver noise temperature ( $T_{\text{rec}}$ ) and the sky temperature (Santos et al. 2015):

$$T_{\text{sys}}(\nu) = 1.1 \times 60 \left( \frac{300}{\nu[\text{MHz}]} \right)^{2.55} + T_{\text{rec}}. \quad (11)$$

The specific receiver and survey properties used here are based on a MeerKLASS-like survey and are summarized in Table 1.

### 3 COSMOLOGICAL SIMULATIONS

We use the same simulated cosmological HI signal data for each sky region. Specifically, we make use of the MULTIDARK-PLANCK cosmological  $N$ -body simulation (Klypin et al. 2016), which evolved  $3840^3$  dark-matter particles in a  $1000^3 h^{-3} \text{Mpc}^3$  volume with the adopted cosmology complying with PLANCK15 (Planck Collaboration XIII 2016). The cosmological parameters used are therefore  $\Omega_{\text{M}} = 0.307$ ,  $\Omega_{\text{b}} = 0.048$ ,  $\Omega_{\Lambda} = 0.693$ ,  $\sigma_8 = 0.823$ ,  $n_s = 0.96$ , and Hubble parameter  $h = 0.678$ . This data have been processed

into the MULTIDARK-GALAXIES data (Knebe et al. 2018), which are galaxy catalogues publicly available from the Skies and Universes web page.<sup>4</sup> It is from these catalogues that we build the simulated HI intensity maps and an overlapping map of resolved optical galaxies.

Each snapshot from the MULTIDARK-GALAXIES simulation represents a different redshift and evolved state of the cosmological density field and the galaxies therein. We opt to use the catalogues at  $z = 0.39$  and take this as the effective redshift ( $z_{\text{eff}}$ ) for our data. This is analogous to real surveys assuming a central effective redshift provided that the width of the bin is small enough so that cosmological quantities can be assumed constant within it.

We still need to assume some redshift range, however, since we require a frequency range from which to produce the foregrounds. We therefore assume our data has redshift range of  $0.2 < z < 0.58$ , which for the HI intensity maps with  $\nu = 1420 \text{MHz}/(1+z)$ , will convert to a frequency range of  $899 < \nu < 1184$  MHz. This frequency range is probed with the  $L$  band from the MeerKAT telescope and is thus representative of a near-term intensity mapping survey (Santos et al. 2017).

The MULTIDARK data we use are for a Cartesian box with galaxy coordinates in physical distances. We thus work in this Cartesian regime throughout this investigation. This is common practice in large-scale structure surveys, where either a small enough sky is surveyed that a flat-sky approximation is valid, or where curved sky effects are accounted for (Blake, Carter & Koda 2018; Castorina & White 2018). At the effective redshift  $z_{\text{eff}} = 0.39$ , the redshift range of  $0.2 < z < 0.58$  we assume for our data converts to a physical distance of  $925 h^{-1} \text{Mpc}$ . We therefore trim the MULTIDARK data cube to this distance along one dimension, keeping the others the same. This results in a data cube with physical size  $L_x, L_y, L_z = 1000, 1000, 925 h^{-1} \text{Mpc}$  where we use the convention that  $x$  and  $y$  are the angular dimensions perpendicular to the line of sight and  $z$  is parallel to the line of sight. We use the plane-parallel approximation throughout. The data cube is gridded into volume-pixels (voxels), with  $n_x = n_y = 256$  along the angular dimensions and  $n_z = 285$  along the radial dimension. The choice of radial binning allows the  $899 < \nu < 1184$  MHz frequency range, we assume to have a frequency resolution of  $\delta\nu = 1$  MHz. As we have already mentioned, the approximate sky coverage of our data is just under  $3000 \text{deg}^2$ , which is fairly representative of proposed intensity mapping surveys like MeerKLASS (Santos et al. 2017).

#### 3.1 HI intensity maps

To produce the intensity maps from the MULTIDARK data we utilize the catalogue produced from applying the SAGE (Croton et al. 2016) semi-analytical model to the data. We summarize our method below for how we produce the intensity maps from the MULTIDARK-SAGE catalogue. For a more complete description of this process, we refer the reader to Cunnington et al. (2020a), where an identical methodology was employed. The MULTIDARK-SAGE catalogue is fully outlined in (Knebe et al. 2018).

First, the cold gas mass for each galaxy is converted into an HI mass, which is then binned into the relevant voxel according to the galaxy’s coordinates. This gridded HI mass is then converted to an HI brightness temperature  $T_{\text{HI}}(\mathbf{x})$ . Since intensity mapping surveys will detect signal down to the very faintest of emitters, it is common in simulations to rescale the  $T_{\text{HI}}$  temperature of the field up to a realistic (expected) value. This is required because simulations have

<sup>4</sup>[www.skiesanduniverses.org](http://www.skiesanduniverses.org)

finite capabilities and often do not resolve haloes down to masses of  $\sim 10^8 h^{-1} M_\odot$  where HI is still predicted to reside (Villaescusa-Navarro et al. 2018; Spinelli et al. 2020). To determine this value, we utilize the results of the GBT-WiggleZ cross-correlation analysis (Masui et al. 2013), where it was found that the HI abundance is  $\Omega_{\text{HI}} b_{\text{HI}} r = [4.3 \pm 1.1] \times 10^{-4}$ , and assume it is constant with redshift. We also take the cross-correlation coefficient to be  $r = 1$  and use an HI bias fit from Villaescusa-Navarro et al. (2018). For our effective redshift this equates to  $b_{\text{HI}}(z_{\text{eff}}) = 1.105$ .

Lastly, in order to emulate the effects from the radio telescope beam, we smooth each channel using  $\theta_{\text{FWHM}} = 1.67$  deg (as discussed in Section 2). The observable field for intensity mapping is the overtemperature field defined as

$$\delta T_{\text{HI}}(z) = T_{\text{HI}}(z) - \langle T_{\text{HI}} \rangle = \langle T_{\text{HI}} \rangle b_{\text{HI}}(z) \delta_{\text{M}}(z), \quad (12)$$

where  $\delta_{\text{M}}(z)$  is the underlying matter density. We have shown an example HI intensity map in the bottom-right panel of Fig. 1 for one frequency channel at 1050 MHz.

### 3.2 Overlapping optical galaxy data

We also utilize the MULTIDARK-GALAXIES for creating an overlapping optical spectroscopic catalogue, which we will use for investigating cross-correlating techniques between HI intensity maps and optical galaxy data. For this purpose, we use the SAG (Cora 2006) semi-analytical model; that is because this catalogue has magnitude outputs for each of the *SDSS ugriz* broad-bands, which can be utilized to construct a realistic optical galaxy data set.

Whilst the MULTIDARK-SAG catalogue does also possess the cold gas mass outputs, and therefore could have also been used to produce the intensity maps, it has fewer galaxies ( $\sim 3.8 \times 10^7$ ) compared with the MULTIDARK-SAGE catalogue ( $\sim 7 \times 10^7$ ) at the snapshot redshift of  $z = 0.39$ . We prefer using MULTIDARK-SAGE for the intensity maps because it has a higher number of galaxies. Since both SAGE and SAG catalogues are generated from the same underlying MULTIDARK simulated density field, they should still produce sufficient cross-correlation signals. The MULTIDARK-SAG catalogue is fully outlined in (Knebe et al. 2018).

Optical galaxy surveys generically operate by constructing a catalogue of resolved galaxies whose luminosity is above some threshold determined by the telescope’s sensitivity. As a rather crude emulation of this method, which is sufficient for this investigation, we use the sum of the magnitudes from the five *SDSS ugriz* bands and select the highest total magnitudes from the simulation until a target  $N(z)$  redshift distribution is achieved. Following Mandelbaum et al. (2011), we construct a realistic target distribution by assuming a double Gaussian where 77.6 per cent of the galaxies are in the first Gaussian and the remaining 22.4 per cent are in the second Gaussian. The Gaussian’s are centred at  $\langle z \rangle = 0.595$  and  $\langle z \rangle = 0.558$  with standard deviations of  $\sigma_z = 0.236$  and  $\sigma_z = 0.112$ , respectively. Since we are simulating a spectroscopic redshift galaxy sample, we assume all redshifts have been measured correctly to the precision required for correct binning into our Cartesian grid. Imposing the redshift bin limits for our simulated survey of  $z_{\text{min}} = 0.2$  and  $z_{\text{max}} = 0.58$  provides the redshift distribution. We then finally stipulate that  $2 \times 10^6$  galaxies will be detected in the optical survey. The overdensity field for the optical galaxies is given as

$$\delta_{\text{g}}(z) = \frac{n_{\text{g}}(z) - \langle n_{\text{g}} \rangle}{\langle n_{\text{g}} \rangle} = b_{\text{g}}(z) \delta_{\text{M}}(z), \quad (13)$$

where  $b_{\text{g}}$  is the linear bias for the optical galaxy field. For these simulated galaxy maps and the simulated HI intensity maps (presented in

Section 3.1) we checked that both measured power spectra, and their cross-correlation, are modelled well by commonly used anisotropic redshift space clustering models (see e.g. Soares et al. 2021), thus validating their use as our underlying cosmological data.

## 4 METHODS FOR FOREGROUND CLEANING

Here, we discuss some of the most popular and well-studied approaches to 21-cm foreground cleaning. Our focus in this work is on single-dish observations, in the context of cosmological analysis and we are therefore ultimately trying to optimize a power spectrum measurement. All foreground removal methods aim to utilize the fact that the foreground contributions are slowly varying with frequency (unlike cosmological HI) and are orders of magnitude more dominant than the HI. Thus, the general approach is identifying a set of smooth functions that represent the dominant foreground contributions and subtracting these from the data to leave the cosmic HI signal. The method for estimating this set of smooth functions is largely where the techniques diverge into the wide library of foreground removal options available today [see e.g. Liu & Shaw (2020) for a more detailed summary].

Blind component separation methods dominate the literature concerning foreground removal techniques, and we also use them in our analysis. Blind separation means little input information is needed and the process exploits the fact that relatively few dominant uncorrelated (or statistically independent) (or sparse) sources should contain the majority of the foreground emission in the observed signal. The advantage of such an approach is that it does not require a detailed understanding of the foreground signals, e.g. their precise amplitude through frequency, and how they respond to instrumental systematics. Given that we are a long way from fully understanding sky emission at the  $\sim 21$ -cm wavelengths and that the intensity mapping technique is still in its infancy (meaning instrumental response and systematics are poorly understood), it is sensible for blind methods to be the preferred choice (Masui et al. 2013; Wolz et al. 2017; Anderson et al. 2018).

The raw observed sky signal in intensity mapping<sup>5</sup> can be decomposed into contributions from the cosmological HI, the foregrounds, and the thermal noise from the instrument (as in equation 1). These observed data can be represented by a matrix  $\mathbf{X}_{\text{obs}}$  with dimensions  $N_\nu \times N_\theta$  where  $N_\nu$  is the number of frequency channels along the line of sight and  $N_\theta$  the number of pixels. In this approach the 2D ( $N_\theta^{\text{RA}}, N_\theta^{\text{Dec}}$ ) angular pixel space is turned into a  $N_\theta = N_\theta^{\text{RA}} \times N_\theta^{\text{Dec}}$  long 1D vector to make the foreground cleaning formalism more concise.

We make the assumption that the data matrix  $\mathbf{X}_{\text{obs}}$  can be represented as a linear system

$$\mathbf{X}_{\text{obs}} = \hat{\mathbf{A}} \mathbf{S} + \mathbf{R}, \quad (14)$$

where  $\hat{\mathbf{A}}$  represents the estimated set of  $N_{\text{FG}}$  smooth functions (often referred to as the mixing matrix) with shape  $[N_\nu, N_{\text{FG}}]$  that evolve the  $N_{\text{FG}}$  separable source maps  $\mathbf{S}$  through frequency. Generally, the sources can be identified by projecting the mixing matrix along the observed data<sup>6</sup>

$$\mathbf{S} = (\hat{\mathbf{A}}^T \hat{\mathbf{A}})^{-1} \hat{\mathbf{A}}^T \mathbf{X}_{\text{obs}}. \quad (15)$$

<sup>5</sup>Neglecting contributions from more complex systematics.

<sup>6</sup>For PCA and SVD, by construction, the set of functions identified for the mixing matrix are orthogonal and hence  $(\hat{\mathbf{A}}^T \hat{\mathbf{A}})^{-1} = \mathbf{I}$ ; thus, this factor is often neglected in equation (15) i.e.  $\mathbf{S} = \hat{\mathbf{A}}^T \mathbf{X}_{\text{obs}}$ .

$N_{\text{FG}}$  is the pre-selected number of separable sources which we expect our foreground emission to be contained within. The remaining signal from subtracting the smooth functions and sources from the data is in the residual term  $\mathbf{R}$  and this is used for the cleaned intensity map data:

$$\mathbf{X}_{\text{clean}} \equiv \mathbf{R} = \mathbf{X}_{\text{obs}} - \hat{\mathbf{A}}\mathbf{S}. \quad (16)$$

This will contain cosmological HI, noise, and typically some residual foreground emission. The resulting cleaned intensity maps can be summarized as

$$\delta T_{\text{clean}} = \delta T_{\text{obs}} - \widehat{\delta T}_{\text{FG}}(\nu, \boldsymbol{\theta}) = \delta T_{\text{obs}} - \sum_{n=1}^{N_{\text{FG}}} \hat{A}_n(\nu) S_n(\boldsymbol{\theta}). \quad (17)$$

As we will see in this investigation, the optimal choice of  $N_{\text{FG}}$  can vary considerably. Generally speaking, an  $N_{\text{FG}}$  that is too low will result in too much foreground signal remaining in the residual component, and an  $N_{\text{FG}}$  that is too high will result in too much cosmological HI leakage into the subtracted component causing a loss of true signal. Finding an optimal balance is the aim of a successful foreground clean, and a key focus in our investigation.

There are many existing methods for estimating  $\hat{\mathbf{A}}$  for a given choice of  $N_{\text{FG}}$ , and we explore some in the remainder of this section. In this section, we aim to introduce some of the most popular blind source separation techniques, and highlight their similarities. We will use an SVD-based technique [or equivalently principal component analysis (PCA) – an equivalence we will explain] as our default foreground cleaning method that we introduce next. We then explore some related techniques with extended sophistication and test them on our simulated data.

We emphasize that the methods we outline are in no way an exhaustive list, and many more methods exist for foreground removal that could be applicable to 21-cm intensity mapping e.g. GNILC (Olivari, Remazeilles & Dickinson 2016), SMICA (Delabrouille, Cardoso & Patanchon 2003), RPCA (Zuo et al. 2018) etc. [see the list in the appendix of Leach et al. (2008) for more information]. One further notable approach is Gaussian process regression (GPR; Mertens, Ghosh & Koopmans 2018), which has been recently used on real data but for a higher redshift, epoch or reionization survey (Mertens et al. 2020). Investigating this method with low-redshift 21-cm intensity mapping data is very interesting and will be the focus of future work.

#### 4.1 PCA (and SVD)

PCA is a widely used technique in statistics, closely related to singular value decomposition (SVD). It provides a hierarchical coordinate system to represent high-dimensional correlated data by transforming it to a dimensional basis that maximizes the variance. These new basis vectors are the principal components. In the context of correlated foreground emission in 21-cm data, due to their large amplitude and highly correlated frequency structure, it is likely that the foreground signals can be reconstructed from just a few of these principal components. Hence, the first few  $N_{\text{FG}}$  dominant basis vectors found in this process represent the estimate for the set of smooth functions in equation (14), which can then be removed from the observational data.

The steps for performing PCA to construct an estimate of the foreground contamination  $\hat{\mathbf{X}}_{\text{FG}}$ , which is then removed from the data, can be concisely outlined as follows:

(i) The data are mean-centred, i.e. the mean at each frequency is subtracted from the data for each frequency channel.

(ii) The covariance matrix of the mean-centred data is calculated:  $\mathbf{C} = \mathbf{X}_{\text{obs}}^T \mathbf{X}_{\text{obs}} / (N_{\theta} - 1)$ .

(iii) The eigen-decomposition of the covariance matrix is computed:  $\mathbf{C}\mathbf{V} = \mathbf{V}\Lambda$ , where  $\Lambda$  is the diagonal matrix of eigenvalues ordered by descending magnitude.

(iv) The first  $N_{\text{FG}}$  columns from the eigenvector matrix  $\mathbf{V}$  represent the set of smooth functions to construct the mixing matrix, i.e.  $\hat{\mathbf{A}} = \mathbf{V}\mathbf{B}$  where  $\mathbf{B}$  is a  $[N_{\nu}, N_{\text{FG}}]$  selection matrix with 1 along the  $i = j$  elements and 0 elsewhere.

(v) The projection of the selected eigenvectors along the mean-centred data provides the eigen-sources,  $\mathbf{S} = \mathbf{A}^T \mathbf{X}_{\text{obs}}$ , which are combined with the mixing matrix to provide the reconstructed foreground estimation  $\hat{\mathbf{X}}_{\text{FG}} = \mathbf{A}\mathbf{S}$ .

##### 4.1.1 Singular value decomposition

The SVD is a unique matrix decomposition of the data (note that PCA and SVD are inherently related). The SVD of the observed data  $\mathbf{X}$  is given by<sup>7</sup>

$$\mathbf{X} = \mathbf{U}\Sigma\mathbf{V}^T, \quad (18)$$

where  $\mathbf{U}$  and  $\mathbf{V}$  are unitary matrices with orthonormal columns and  $\Sigma$  is a diagonal matrix whose entries represent the *singular values*. It can be demonstrated how closely related the SVD is to an eigenvalue decomposition. By considering equation (18), and given that  $\mathbf{X}^T = \mathbf{V}\Sigma\mathbf{U}^T$ , the covariance can be written as

$$\mathbf{C} \equiv \frac{\mathbf{X}^T\mathbf{X}}{N_{\theta} - 1} = \frac{\mathbf{V}\Sigma\mathbf{U}^T\mathbf{U}\Sigma\mathbf{V}^T}{N_{\theta} - 1}. \quad (19)$$

A fundamental property of SVD stipulates that  $\mathbf{U}^T\mathbf{U}$  is a unitary matrix ( $\mathbf{U}^T\mathbf{U} = \mathbf{I}$ ). With a little rearranging we get

$$\mathbf{C}\mathbf{V} = \frac{\mathbf{V}\Sigma^2}{N_{\theta} - 1}, \quad (20)$$

which we can recognize as an eigenvalue decomposition of the correlation matrix  $\mathbf{C}$  (as shown in step 3 in Section 4.1) where the columns of  $\mathbf{V}$  are the eigenvectors and the singular values  $\Sigma$  are proportional to the positive square roots of the eigenvalues.

In real intensity mapping data, in order to mitigate the high levels of thermal noise bias present in pathfinder experiments and decrease systematics, it is necessary to cross-correlate data from different observation runs e.g.  $\mathbf{X}_A \times \mathbf{X}_B$ . Whilst separating the data in this way decreases sensitivity and boosts thermal noise for each individual run, the noise should be uncorrelated for each run and thus thermal noise bias is mitigated in the cross-power. In this situation, the covariance matrix  $\mathbf{X}_A^T \mathbf{X}_B$  is no longer symmetric and an SVD is required where the left and right singular vectors in  $\mathbf{U}$  and  $\mathbf{V}$  are used to reconstruct foreground estimates in each run (Switzer et al. 2013). In this work, we do not explore such a situation and therefore the SVD and PCA can be seen as equivalent treatments.

A related, and essentially equivalent, method to PCA is polynomial fitting (Ansari et al. 2012). Although similarities exist, this is not to be confused with parametric fitting (see Section 4.4) and conventionally refers to a blind approach to foreground cleaning. The approach works by identifying a set of smooth fitted functions  $f_k$  where polynomials are used as basis functions e.g.

$$f_k(\log(\nu)) = [\log(\nu)]^{k-1}. \quad (21)$$

<sup>7</sup>The more general form for SVD is  $\mathbf{X} = \mathbf{U}\Sigma\mathbf{V}^*$ , however, in the context of 21-cm data we are always dealing with real-valued matrices where  $\mathbf{V}^* \equiv \mathbf{V}^T$ .

Then, by least-squares fitting these functions to each line of sight, the foreground contribution can be approximated. Since previous work has already demonstrated the theoretical equivalence this has with PCA [e.g. Alonso et al. (2015) that also provides simulation tests] we do not include this in our investigation.

#### 4.1.2 Truncation choice

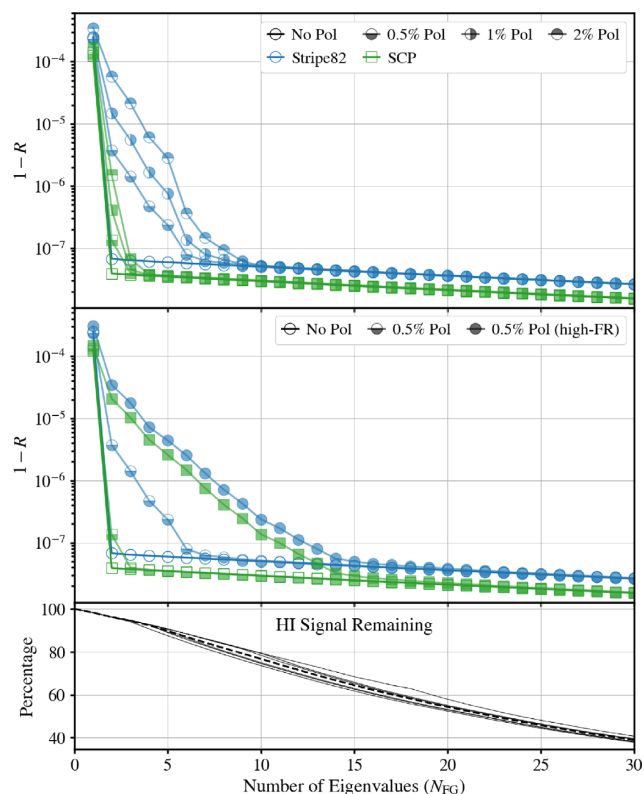
Deciding where to truncate to, i.e. the number of  $N_{\text{FG}}$  principal components to include in the foreground estimate (and hence remove) is key to an optimized blind foreground clean. By analysing the eigenvalues in  $\Lambda$  (or, equivalently, the singular values from the SVD), that estimate the amount of variance in the data captured in the corresponding principal components, an informed choice can be made. As discussed above, due to the nature of the foreground emission, most of the information is contained in a small sub-set of principal components where often  $N_{\text{FG}} \sim 3 \rightarrow 20$  (depending on the foreground emission and instrument response) can produce a reasonable reconstruction. We can quantitatively analyse this choice with

$$R = \frac{\sum_{i=1}^{N_{\text{FG}}} \lambda_i}{\sum_{i=1}^{N_v} \lambda_i}, \quad (22)$$

where  $\lambda_i$  are the eigenvalues in  $\Lambda$ , descending in magnitude and  $N_v$  is the number of frequency channels along the line of sight. Since a higher number for  $N_{\text{FG}}$  will remove more HI information, the aim for an optimal choice is to maximize  $R \rightarrow 1$  for a minimal value for  $N_{\text{FG}}$ .

We show some values for  $R$  in Fig. 4 for the different sky regions in our simulated data, with differing polarization leakage cases; we plot  $1 - R$  to demonstrate the convergence. This means the closer to zero the particular combination of eigenvalues is, the better a representation of the full data that reconstruction will be since it is capturing more of the full data's variance, and its reconstruction will capture more of the foreground contamination which can be removed. Fig. 4 immediately shows how highly correlated the observed data are given that just one eigenvalue in all cases has  $R \sim 1$  meaning nearly 100 percent of the signal can be represented with just 1 principal component ( $N_{\text{FG}} = 1$ ). However, just a small amount of residual foreground, even at the sub-percent level, is enough to entirely dominate the HI signal. Therefore, in all cases  $N_{\text{FG}} = 1$  is not sufficient for a foreground clean. This plot gives an indication of how far one needs to go in the reconstruction. All cases eventually reach a plateau where including more eigenvalues barely contributes to the reconstructed signal and it is here where PCA has likely reached its efficiency limit and will not be able to remove much more foreground. The high FR cases will demand the most modes for a successful reconstruction, requiring  $N_{\text{FG}} \sim 15$  to converge to the small eigenvalue plateau. We note that since this effect is being simulated from the Galactic plane for both Stripe82 and SCP regions, we will expect our results to converge to a plateau at a similar number of eigenvalues for both regions in this high-FR case. We see this in the middle panel and also see this in later results. This limits comparisons between the two regions for this extreme case but, as discussed, this approach is necessary to provide simulated data which require an aggressive foreground clean with a high  $N_{\text{FG}}$ .

In contrast, the HI information cannot be compressed into a small number of principal components due to its Gaussian-like nature. This is the main principle behind the blind source separation approach. The highly correlated information containing the majority of foregrounds can be removed using a few  $N_{\text{FG}}$  modes, leaving the



**Figure 4.** Weighted contributions from increasing numbers of principal components for the frequency–frequency covariance matrix for different polarization cases in both sky regions.  $R$  (outlined by equation 22) is the sum of the first  $N_{\text{FG}}$  eigenvalues divided by the sum of all eigenvalues. Therefore, the closer to zero  $1 - R$  is, the more eigen information (or variance) is represented in those principal components. (Top-panel) the difference from an increasing polarized fraction percentage. (Middle-panel) the impact from higher FR. (Bottom-panel) an estimation for the amount of HI information along the line of sight that remains after  $N_{\text{FG}}$  principal components are removed. Thin lines are for each of the different cases in the above panels, and the thick-dashed line is their average.

bulk of the HI information that is mostly evenly distributed among the remaining components. However, it does mean a fine balance needs to be attained in a successful foreground clean. Being too aggressive and choosing too high values for  $N_{\text{FG}}$  will begin to remove HI information, typically large-scale line-of-sight modes.

In a simulation-based procedure, we can effectively analyse this problem since we have access to the separated *pure*-HI<sup>8</sup> and *pure*-foreground simulated data. We can therefore calculate the contributions from these components remaining in the residuals after a foreground clean. The separated residuals are calculated using the estimated mixing matrix  $\hat{\mathbf{A}}$ , and projecting the pure-HI (or pure-foreground) simulated data along this:

$$\mathbf{X}_{\text{residHI}} = \mathbf{X}_{\text{HI}} - \hat{\mathbf{A}}(\hat{\mathbf{A}}^T \hat{\mathbf{A}})^{-1} \hat{\mathbf{A}}^T \mathbf{X}_{\text{HI}}, \quad (23)$$

$$\mathbf{X}_{\text{residFG}} = \mathbf{X}_{\text{FG}} - \hat{\mathbf{A}}(\hat{\mathbf{A}}^T \hat{\mathbf{A}})^{-1} \hat{\mathbf{A}}^T \mathbf{X}_{\text{FG}}. \quad (24)$$

Note the  $(\hat{\mathbf{A}}^T \hat{\mathbf{A}})^{-1}$  factor is not needed for the PCA method since, by construction, the mixing matrix is orthogonal and this quantity will equal the identity matrix. However, the vectors in the mixing matrix for the fast independent component analysis (FASTICA) and

<sup>8</sup>We also include the contribution from thermal noise in this calculation.

generalized morphological component analysis (GMCA) approach we use are not orthogonal, thus this quantity is needed to obtain the correct projection.

We utilize these separated residual calculations extensively in our results to analyse the performance from various cleaning methods under different situations. We also use this concept for the bottom panel of Fig. 4 where we demonstrate an estimation for the amount of eigen information lost along the line of sight for each choice of  $N_{\text{FG}}$ . This is calculated by computing the eigendecomposition of the residual HI (equation 23) and summing the eigenvalues. Dividing this by the sum of all the eigenvalues in the original HI data gives a proxy for the amount of eigen information remaining after subtracting  $N_{\text{FG}}$  principal components, i.e. the portion of variance that is removed. This should further illustrate the challenge of foreground cleaning, a balance between removing foreground while trying to leave the HI signal intact – however, we always expect some signal loss. Since foreground dominated data are typically decomposed into dominant eigenmodes containing highly correlated information along the line-of-sight subtracting  $N_{\text{FG}}$  principal components generally removes large-scale modes along the line of sight in the HI power spectrum, i.e. small  $k_{\parallel}$  modes. Modelling this is non-trivial and a major challenge for precision radio cosmology.

## 4.2 FASTICA

FASTICA is another widely used method for foreground cleaning and has been tested on simulated data (Chapman et al. 2012; Wolz et al. 2014; Cunnington et al. 2019) and also on real data (Wolz et al. 2017). When we discuss FASTICA, we are referring to the method developed in Hyvärinen (1999) and we use the package in *Scikit-learn*<sup>9</sup> (Pedregosa et al. 2011).

While PCA is generalized for reducing dimensionality in data, FASTICA (and more generally independent component analysis) is more specifically used to separate mixed signals, and is therefore naturally suited to a blind source separation problem. FASTICA forms an estimate for the mixing matrix  $\hat{\mathbf{A}}$  by assuming the sources are statistically independent of each other. The method therefore aims to maximize statistical independence that can be assessed using the central limit theorem, which states that the greater the number of independent variables in a distribution, the more Gaussian that distribution will be (that is, the probability density function of several independent variables is always more Gaussian than that of a single variable). Hence, by maximizing any statistical quantity that measures non-Gaussianity, we can identify statistical independence.

Before assessing non-Gaussianity, FASTICA begins by mean-centering the data then carries out a *whitening* step that aims to achieve a covariance matrix equal to the identity matrix for this whitened data (i.e. the components will be uncorrelated and their variances normalized to unity). Since this whitening step can be achieved with a PCA analysis, FASTICA is essentially an extension of PCA, and hence in most cases in the context of foreground cleaning, will provide very similar results.

For maximizing non-Gaussianity, an approximation of the negentropy can be used.<sup>10</sup> We refer the reader to Hyvärinen & Oja (2000) for further detail on this aspect of the algorithm. In the context of 21-cm foreground cleaning, the approximation of negentropy uses a set of optimally chosen non-quadratic functions which are applied to the data and averaged over for all available pixels. The maximization

of negentropy by averaging over angular pixels means that for purely Gaussian sources, FASTICA will be unable to improve upon the initial PCA step carried out in the whitening step. This is because the Gaussian sources will have an equivalent zero negentropy. This explains the similarity in results often found between PCA and FASTICA when most of the simulated components are Gaussian fields (Alonso et al. 2015). It is in situations over very large skies, where the negentropy approximation will be more optimal and sufficient non-Gaussian structure exists in the foreground maps, where FASTICA will perhaps make discernible differences to the PCA-only performance.

To summarize, the components found using PCA are uncorrelated linear combinations of the data, which are identified by maximizing the variance. FASTICA extends on this by finding components that are also uncorrelated linear combinations of the data but identified by maximizing statistical independence, through estimates of non-Gaussianity in angular pixels.

## 4.3 GMCA

GMCA (Bobin et al. 2007), it is a blind source separation algorithm exploiting the idea that the different components contributing to the signal are morphologically different. To enhance the morphological differences, the signal is projected into an adapted domain where we expect the components to have a sparse representation, i.e. to be described by few non-zero coefficients. When we find such a domain, the contrast between components increases, easing the separation process. Here, we make use of wavelets, which has recently been shown to be optimal for this context (Carucci et al. 2020a). GMCA has already been optimized and used with astrophysical data sets [e.g. cosmic microwave background data (Bobin et al. 2013, 2014), high-redshift 21-cm interferometric data (Patil et al. 2017), X-ray images of supernova remnants (Picquenot et al. 2019)].

In practice, once the data  $\mathbf{X}_{\text{obs}}$  have been wavelet transformed to  $\mathbf{X}^{\text{wt}}$ , GMCA promotes sparsity in the requested  $N_{\text{FG}}$  sources  $\mathbf{S}^{\text{wt}}$  by solving iteratively the minimization problem given by

$$\{\hat{\mathbf{A}}, \hat{\mathbf{S}}\} = \min_{\mathbf{A}, \mathbf{S}^{\text{wt}}} \sum_{i=1}^{N_{\text{FG}}} \lambda_i \|\mathbf{S}_i^{\text{wt}}\|_1 + \|\mathbf{X}^{\text{wt}} - \mathbf{A}\mathbf{S}^{\text{wt}}\|_F^2, \quad (25)$$

where the first term is the  $\ell_1$  norm, i.e.  $\sum_{j,k} |\mathbf{S}_{j,k}^{\text{wt}}|$ : this constitutes a constraint for sparsity, mediated by the regularization coefficients  $\lambda_i$ . The latter act as sparsity thresholds that in our case should be tuned by the difference in intensity between the foregrounds and the cosmological signal; we first estimate them with the median absolute deviation (MAD) method and progressively decrease towards a final noise-related level. The second term in equation (25) is the standard Frobenius norm that assures data-fidelity step by step.

Once the mixing matrix  $\hat{\mathbf{A}}$  has been estimated, we project the initial data  $\mathbf{X}_{\text{obs}}$  in pixel space (following equations 15 and 16) to retrieve the GMCA-reconstructed data cubes. We refer the reader to Carucci et al. (2020a) for more details.

## 4.4 Non-blind parametric fitting

In non-blind methods an estimator  $\hat{\mathbf{A}}$  for the mixing matrix is constructed using astrophysical, as opposed to statistical, knowledge about the foreground sources and has been previously explored (Ansari et al. 2012; Bigot-Sazy et al. 2015). A single-frequency channel of a 21-cm intensity mapping experiment will consist of diffuse synchrotron emission, diffuse free-free emission, extragalactic point sources, the HI signal, instrumental noise, and any other instrumental

<sup>9</sup><https://scikit-learn.org/>

<sup>10</sup>Kurtosis can also be used as a measure of non-Gaussianity.

contributions (e.g. polarization leakage). Synchrotron and free–free emission are believed to be spectrally smooth with well-understood spectral forms that can both be expressed as power laws. While the synchrotron spectral index is known to change across pixels, diffuse synchrotron emission is the signal identified with the largest signal-to-noise ratio within 21-cm intensity mapping experiments giving it the largest probability of an accurate characterization. It should be noted, however, that as synchrotron emission is around four orders of magnitude larger than the 21-cm signal of interest it would need to be characterized to an accuracy of 0.01 per cent in order to no longer obscure the HI signal.

As such we propose a parametric fit which aims to parametrize the free–free and synchrotron foreground contributions explicitly. Diffuse synchrotron, diffuse free-emission and extragalactic point sources are strongly degenerate; free–free and synchrotron emission maps contain identical spatial features and all three spectra can be represented as power laws with similar spectral indices. Whilst we do not aim to explicitly fit for the extragalactic point sources we expect their contributions to be subsumed within the synchrotron and free–free emissions fits. We are essentially making the opposite assumption to ICA by relying upon the parameter degeneracy, if this assumption is correct then the residuals between our parametric fit and the total data should contain the HI plus any instrumental contributions.

Asorey et al. (2020) attempt a linear least-squares fitting to their data, modelling their combined foregrounds as a  $n$ th order polynomial. We also use the least-squares optimizer (equation 15) for the emission sources. However, in an attempt to capitalize on existing foreground information, we aim to provide the optimization with a realistic mixing matrix. We set up a mixing matrix with two components (to represent the combination of free–free, synchrotron emission, and point sources). For the first component, we use the assumption that the free–free spectral index is well-known and constant across pixels and so set the spectral form to the true value  $(\nu/\nu_0)^{-2.13}$ . For high Galactic latitudes such as the SCP and Stripe 82, the free–free emission is weak enough to be assumed negligible and thus we employ a least-squares fit assuming pure synchrotron emission. Close to the Galactic Plane, we find that actual component separation between free–free and synchrotron emission is required. As intensity mapping experiments will not typically target such regions we shall not discuss parametric component separation any further. However, the interested reader can refer to Bobin, Acero & Picquenot (2019) for a description of a novel semisupervised sparse component separation method, which has been used (by these authors and on the same total intensity simulations used in this work) to determine accurate synchrotron spectral indices in the presence of non-negligible free–free emission.

Before describing our least-squares fit, we point out that the data monopoles must be removed from each map; for our particular simulations that means the unresolved extragalactic point-source levels at each frequency. The spectral index for a particular emission is strongly tied to the monopole level of the maps and so the parametric fit we perform is tied to the zero-level of the observational data.

For our least-squares fit, which assumes that synchrotron emission dominates the total intensity maps, we limit the parameter space for  $\beta_{\text{sy}}$  to within  $\pm 10$  per cent of the total data spectral index and the parameter space for the synchrotron emission amplitude to within  $\pm 50$  per cent of the total temperature.

We also investigated the possibility of performing an MCMC fit [using EMCEE (Foreman-Mackey et al. 2013)], to see if this offered significant benefits over least-squares fitting. We imposed flat priors

on the synchrotron and free–free emission amplitudes and, following the methodology of Eriksen et al. (2008), the Jeffreys prior on the synchrotron spectral index. For the SCP region without polarization leakage, a marginal (on average around a tenth of a percent) improvement was seen for the estimation of the synchrotron spectral index. However, when polarization leakage is added the MCMC fit no longer outperforms the least-squares fit. This is unsurprising as the strength of an MCMC Bayesian fitting process is the ability to provide end-to-end error propagation. In our case, we are simply using the theoretical Gaussian noise level per frequency to form our noise estimates; we have no model for polarization leakage as part of the noise estimates. As we aim to test how well a parametric fit of intensity mapping data can perform using existing astrophysical information we stick to least-squares fitting. Over time, however, it should be possible to perform an instrument specific, iterative Bayesian fit to intensity mapping data such as the CMB data analysis performed in BeyondPlanck Collaboration I (2020).

Having used the least-squares fit to acquire the per-pixel synchrotron spectral index values we now have a complete mixing matrix which we use to calculate the diffuse Galactic emission amplitudes from the total temperature maps using equation (15). We can subtract our free–free emission and synchrotron emission estimates from the total temperature data to leave maps of HI plus instrumental contributions.

#### 4.5 Quantifying foreground removal effects

Despite the range of different foreground cleaning methods available, none are perfect and will inevitably remove some cosmological HI signal or leave behind foreground residuals. We discuss some methods for investigating this both with simulations and real data.

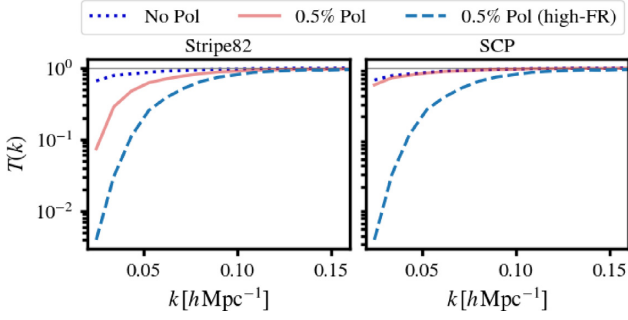
##### 4.5.1 Damping Cosmological HI

This usually occurs on large scales where the HI is most degenerate with the foregrounds. For idealized future surveys assuming excellent instrumental calibration, residual foregrounds should be well controlled and not exacerbated from effects such as polarization leakage. In these cases, the effects from a low- $N_{\text{FG}}$  foreground clean are relatively straightforward and can be potentially modelled as some damping to the power spectrum (see e.g. Cunnington et al. 2020a, b; Soares et al. 2021).

However, applying this to real data requires a high level of confidence in the modelling that builds upon a detailed understanding of the nature of foregrounds as well as systematic/instrumental effects, something we do not currently have. An alternative approach is to add the observed data itself to simulations (mocks), then apply a foreground clean and access the response the mock data had to this process. Signal loss can be quantified this way with a *foreground transfer function*, which is applied to the real data to compensate for these effects, although does not avoid the reduced sensitivity caused by the contamination. This has been the approach of several of the HI intensity mapping detections so far (Masui et al. 2013; Switzer et al. 2013; Anderson et al. 2018; Wolz et al. 2021).

Following Switzer et al. (2015), the transfer function can be constructed by adding mock data  $\mathbf{M}$  to the true observed data  $\mathbf{X}_{\text{obs}}$ , which includes foregrounds. This can then be cleaned to provide  $\mathbf{M}_{\text{cleaned}}$ , an estimate for the effects of removing the foregrounds on the mock map:

$$\mathbf{M}_{\text{cleaned}} = [\mathbf{M} + \mathbf{X}_{\text{obs}}]_{\text{PCA}} - [\mathbf{X}_{\text{obs}}]_{\text{PCA}}, \quad (26)$$



**Figure 5.** Foreground transfer functions for the different foreground cases in each sky region. These are produced using equation (27) and treating our MULTIDARK simulated intensity maps as ‘observed’ data and then using 100 lognormal sims as the input mocks. For the three foreground cases (no polarization leakage, 0.5 per cent polarization leakage, and high FR) we use a PCA clean with  $N_{\text{FG}} = \{3, 9, 15\}$  for Stripe82 and  $N_{\text{FG}} = \{3, 4, 15\}$  for SCP.

where the  $[\ ]_{\text{PCA}}$  notation represents performing a PCA clean, but in principle this could be done with any foreground cleaning method. Note that in equation (26), the cleaned data  $[\mathbf{X}_{\text{obs}}]_{\text{PCA}}$  has been subtracted. This is necessary to reduce the variance in this estimation since the unwanted data-HI component will serve as additional, unwanted noise. The transfer function is then given by

$$T(k) = \left\langle \frac{\mathcal{P}(\mathbf{M}_{\text{cleaned}}, \mathbf{M})}{\mathcal{P}(\mathbf{M}, \mathbf{M})} \right\rangle^2, \quad (27)$$

where  $\mathcal{P}()$  denotes an operator which measures the power spectrum in  $(k_{\perp}, k_{\parallel})$  space. The angled brackets denote an averaging over a large number of mocks. The power spectrum is then corrected for by dividing through by this transfer function. This can also be utilized in a cross-correlation measurement with the only difference being that the power of 2 is dropped from equation (27) because the effects of cleaning are only applied to the HI data. We employ the transfer function later in our analysis (Section 5.5) by constructing 100 lognormal mocks and using our MULTIDARK simulations as the ‘observed’ data. We show the transfer functions for the different data sets in Fig. 5. This shows the range in foreground contamination from the different foreground cases, largely driven by the choice of  $N_{\text{FG}}$  components to remove, which we discuss in detail in our results (Section 5).

The foreground transfer function is thus used as a data-driven way of compensating for signal loss in the foreground removal. Since the *real* data are used in its construction, it should incorporate the interplay between foreground and systematics (even unknown systematics). Some assumptions are inherently made regarding the degeneracy of the transfer function with cosmological parameters, but this will only be a problem for precision cosmology which will need more detailed investigation when HI intensity mapping reaches this level. This subtlety regarding degeneracies is discussed further in Cunnington et al. (2020a). Since  $T(k) \leq 1$ , the transfer function is not capable of addressing the issue caused by *additive* biases from foreground residuals, discussed in the following section.

#### 4.5.2 Foreground residuals

Whilst signal loss from overcleaning can be modelled or compensated for with a foreground transfer function, foreground residuals produced from undercleaning, which cause additive biases and boost errors, are more challenging to address. For near-future, pathfinder

surveys (e.g. MeerKAT; Santos et al. 2017) it is possible that the instrument response will not be sufficiently understood and polarization leakage effects could manifest, causing contamination from foreground residuals. Developing robust statistics that estimate the effects caused by these residuals will therefore be essential for future surveys (Switzer et al. 2015). There is not a large amount of research on this issue, since current data analysis usually has large thermal noise and unknown systematic effects (Switzer et al. 2013). Alternatively, detections have been made using cross-correlations with optical galaxy data (e.g. Masui et al. 2013; Anderson et al. 2018; Wolz et al. 2021). In cross-correlation, the residual foregrounds and survey-specific systematics do not correlate with the optical galaxy data and instead, simply boost errors (we will study this in detail in Section 5.5). As the intensity mapping technique matures and calibration and signal-to-noise capabilities of surveys improve, we will aim to conduct precision cosmology using autocorrelation measurements. Therefore, we need to develop a pipeline for quantifying the foreground residual contamination.

As discussed in Section 4.1.2, the residuals can be exactly calculated (see equation 23 and equation 24) because we are using simulations where the original decomposed HI and foreground contributions are known. A direct comparison between  $\mathbf{X}_{\text{residHI}}$  and  $\mathbf{X}_{\text{residFG}}$  is then extremely useful (and a topic we investigate) where one ideally desires a situation where  $\mathbf{X}_{\text{residHI}}$  dominates  $\mathbf{X}_{\text{residFG}}$ . A more dominant  $\mathbf{X}_{\text{residFG}}$  would increase additive biases due to the residuals correlating with each other. However, in real data, distinguishing the contribution between foreground residual and HI signal will be challenging. If one can develop a robust way of estimating the contribution from foreground residuals, then this can be effectively modelled in a similar way to the instrumental noise which causes additive power in autocorrelations along with boosting errors. So the HI autopower spectrum could be expressed as

$$P_{\text{HI}}(k) = \langle T_{\text{HI}} \rangle^2 b_{\text{HI}}^2 P_{\text{m}}(k) + P_{\text{N}}(k) + P_{\text{residFG}}(k), \quad (28)$$

where  $P_{\text{m}}$  is the matter power spectrum and  $P_{\text{N}}$  and  $P_{\text{residFG}}$  are the contributions from thermal noise and residual foregrounds. An estimation for the errors can then be analytically made with

$$\sigma_{\rho}(k) \sim \frac{P_{\text{HI}}(k) + P_{\text{N}}(k) + P_{\text{residFG}}(k)}{\sqrt{N_{\text{modes}}(k)}}, \quad (29)$$

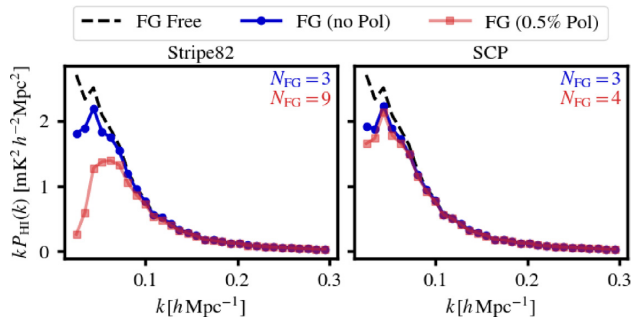
where  $N_{\text{modes}}$  is the number of unique modes in each  $k$ -bin, included to account for cosmic variance.

## 5 RESULTS

Here, we present our results from tests carried out on the simulated data sets and foreground removal methods outlined in the previous sections. To diagnose the performance of our foreground cleans we look at measurements of power spectra, both 1D  $P(k)$  and 2D  $P(k_{\perp}, k_{\parallel})$ , and compare these to equivalent foreground-free results where only the cosmological HI is being measured. The offset between the two then serves as a good indicator for how well the chosen method is performing. Following previous studies (e.g. Alonso et al. 2015; Carucci et al. 2020a), we define below the weighted difference between subtracted foreground and no foreground cases as a metric to help assess the success of the foreground removal under all the different scenarios:

$$\varepsilon(k) = \frac{P_{\text{SubFG}}(k) - P_{\text{NoFG}}(k)}{P_{\text{NoFG}}(k)}. \quad (30)$$

Here,  $P_{\text{SubFG}}(k)$  is the measured power spectrum for the simulated intensity maps with foregrounds included and then cleaned, while



**Figure 6.** Measured power spectra for both regions. We show the foreground free case (black-dashed), cleaned foregrounds without polarization leakage (blue-circle markers) and cleaned foregrounds with 0.5 per cent polarization leakage (red-square markers). For both regions where there is no polarization leakage, we use  $N_{\text{FG}} = 3$  as the number of removed principal components, but for the polarization leakage cases, a differing selection is needed (displayed in top-right of both panels). In all foreground cleaned cases, PCA is used.

$P_{\text{NoFG}}(k)$  is the measured power spectrum of the H I-only (foreground-free) intensity maps. We also analyse the 2D power spectrum and use an identical analysis in this basis where

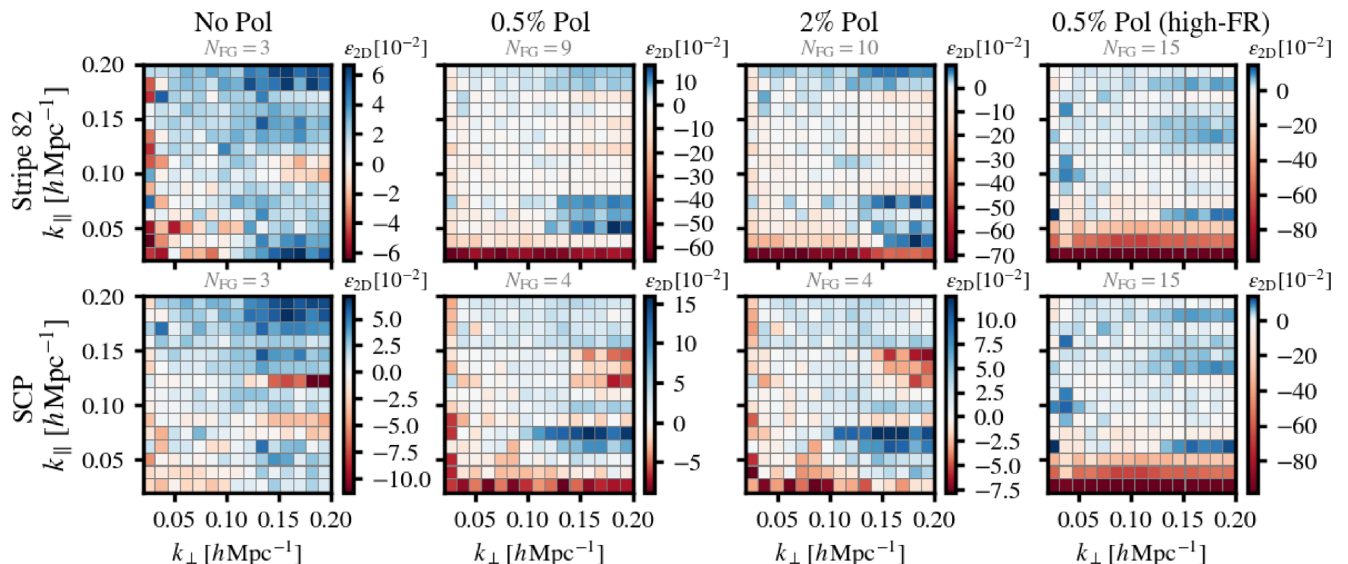
$$\varepsilon_{2\text{D}}(k_{\perp}, k_{\parallel}) = \frac{P_{\text{SubFG}}(k_{\perp}, k_{\parallel}) - P_{\text{NoFG}}(k_{\perp}, k_{\parallel})}{P_{\text{NoFG}}(k_{\perp}, k_{\parallel})}. \quad (31)$$

We begin by plotting the autopower spectra for both our chosen regions and perform a comparison between the foreground-free H I power spectrum (black dashed line) and the foreground-cleaned results using the PCA method. This demonstrates some differences between the regions and also some clear differences for the cases with polarization leakage (red squares) and without (blue circles). As expected most of the damping to the power comes from large scales (small- $k$ ). Because we use  $N_{\text{FG}} = 3$  in both regions where there is no polarization leakage, the damping across all regions is approximately equal. However, the foreground residuals could still differ in each region, with the most likely case being that residuals will be highest

in Stripe82 where foregrounds are slightly more dominant. Fig. 6 lastly shows that results can be extensively worse when including polarization leakage effects, except perhaps the SCP, the region least affected by polarization in the CRIME model. Results are much worse for Stripe82, where we opted to use  $N_{\text{FG}} = 9$  to remove more oscillating foreground contamination from polarization leakage. For example the largest mode (smallest- $k$ ) for Stripe82 is effectively damped to zero. We discuss in more detail the choice of  $N_{\text{FG}}$  later in Section 5.4.

We show similar results in Fig. 7 but here more information can be extracted on the nature of this contamination. This shows the weighted difference  $\varepsilon_{2\text{D}}$  between the H I-only 2D power spectrum and the foreground cleaned one. Again we plot both regions on different rows but now with a wider range of polarization leakage cases. This gives an illustration into the effects of foreground undercleaning and overcleaning and the delicate balance between the two. From equation (31) we can see that the blue regions are indicating modes that have higher power in the foreground-cleaned maps compared to the foreground-free ones. Thus, blue areas indicate undercleaned  $k$ -space. Conversely, red areas have lower power in the foreground-cleaned maps, indicating overcleaning.

Fig. 7 therefore shows the effects of over-cleaning tend to manifest in low- $k$  modes, as expected. This is particularly evident when going to high  $N_{\text{FG}}$  as is required in the high FR polarization leakage cases (far-right column), where we see significant damping to low  $k_{\parallel}$  modes, again as expected. This is because, in order to control contamination from polarization leakage, we are removing more principal components, each with different oscillating modes due to the instrumental response, but all will still have largely frequency correlated spectra. This inevitably removes the modes in H I which are also highly correlated in frequency i.e. low- $k_{\parallel}$  modes. This conclusion is quite general and not just specific to a PCA-based method. Generally, any method that utilizes the highly correlated nature of foreground signals will struggle to disentangle foregrounds and large H I modes parallel to the line of sight.



**Figure 7.** The impact of a PCA foreground clean on the 2D power spectra. Plotted is the weighted difference  $\varepsilon_{2\text{D}}(k_{\perp}, k_{\parallel})$  between the foreground free and cleaned 2D power spectra (outlined in equation 31). Each panel represents a different polarization leakage case for both Stripe82 (top-row) and SCP (bottom-row) sky regions. We show 0, 0.5, and 2 per cent leakage along with the special high-FR case. The results can be loosely interpreted as blue positive-pixels representing *undercleaned* modes (modes with dominant foreground residual), and red negative-pixels representing *overcleaned* modes. The values for  $N_{\text{FG}}$  are displayed in grey above each panel.

For the unpolarized cases, where a lower  $N_{\text{FG}}$  is used, it is interesting to see that small- $k_{\perp}$  modes are damped to a similar level as the small- $k_{\parallel}$  modes. This will be because the foregrounds generally have quite large angular structures (see maps in Fig. 1), therefore when removing the dominant principal components representing the foregrounds, any small- $k_{\perp}$  modes will be degenerate with these and are therefore damped. This agrees with results from Soares et al. (2021) where a damping to  $k_{\perp}$  modes was required to model the foreground contamination.

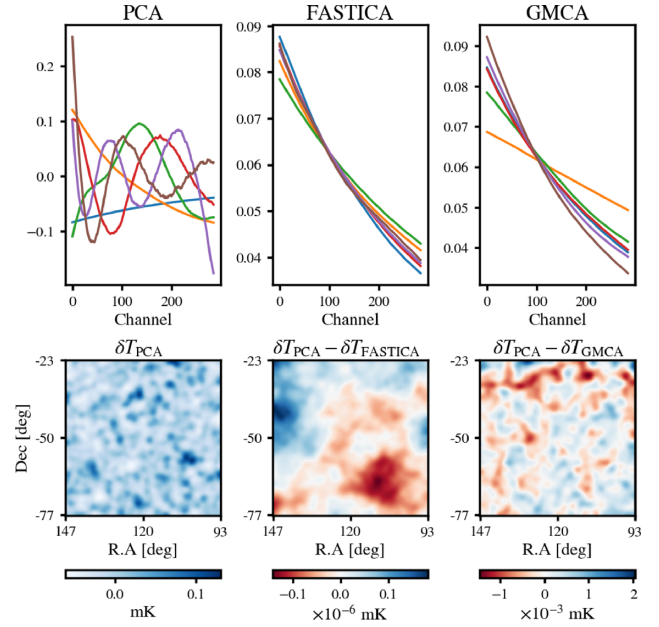
Comparing the middle-left and middle-right panels in Fig. 7, the increased polarization fraction does not appear to have a drastic effect. This is also consistent with the results from the top-panel of Fig. 4 which showed that an increase in polarization fraction, seems to just increase the amplitude of the eigenvalues one would remove anyway in a foreground clean. A polarization fraction above 2 per cent is unlikely to occur in surveys, even in the most pessimistic of forecasts and we therefore stick to using a polarized fraction of 0.5 per cent for the remainder of the paper. However, as we demonstrate in the far-right panels, one way to complicate foreground cleaning is if FR is higher than is being realized in the CRIME model. As explained in Section 2.4, the way we investigate the possibility of a more complex frequency decoherence situation is by taking the CRIME polarization output map from the Galactic plane, and using this as a high FR case. Doing this creates much more oscillating behaviour in the foreground spectra and as Fig. 7 shows, requires a more aggressive clean, with  $N_{\text{FG}} = 15$ .

### 5.1 Comparing blind foreground cleaning methods

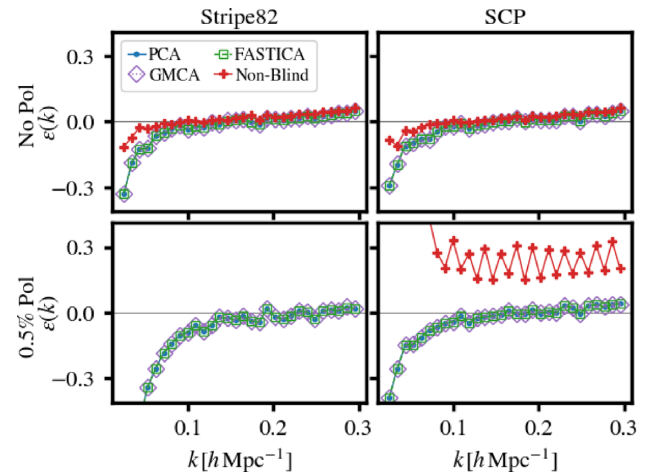
We now compare the cleaning methods we have introduced: PCA, FASTICA, and GMCA. All methods rely on the assumption that we can decompose the signal linearly as in equation (14) and estimate the mixing matrix  $\hat{\mathbf{A}}$ , identifying the subspace of the data set where we expect foregrounds to live, which are then removed as per equation (16).

In the top-panels of Fig. 8, we show the mixing matrices derived by the different methods applied to the same data cube. Each method provides a different estimation for  $\hat{\mathbf{A}}$ , yet, the final cleaned maps are remarkably similar in all cases. We found that there were no discernible differences in the maps and in the bottom panel we demonstrate this. We plot the PCA-cleaned H I intensity map for one channel (bottom-left panel of Fig. 8) and then show the difference between this PCA map and the corresponding FASTICA and GMCA counterparts (middle and right bottom panels). Their differences are orders of magnitude below the amplitudes of the cleaned map and this is true for all channels of all the sky regions explored, with and without the inclusion of polarization leakage. Given the similarity in the cleaned maps, it is unsurprising that at the 2-point statistics level, in all the scenarios considered, the three blind methods output essentially identical power spectra (as shown by the examples in Fig. 9).

The difference in the top-panel between PCA and the other methods is an interesting demonstration of the subtle distinction in techniques. PCA is maximizing the variance into as few modes as possible. The highest ranked mode represents the one that best fits the variance of the data. The second highest rank mode, will be the next best fit but is required to be orthogonal to the first, hence why PCA identifies two dominant smooth modes in Fig. 8, which likely contain the synchrotron and free-free emission. The remaining modes are then more oscillatory and likely identify polarized residuals in a descending order of contribution to the total variance. Applying FASTICA and GMCA algorithms, we are instead identifying a

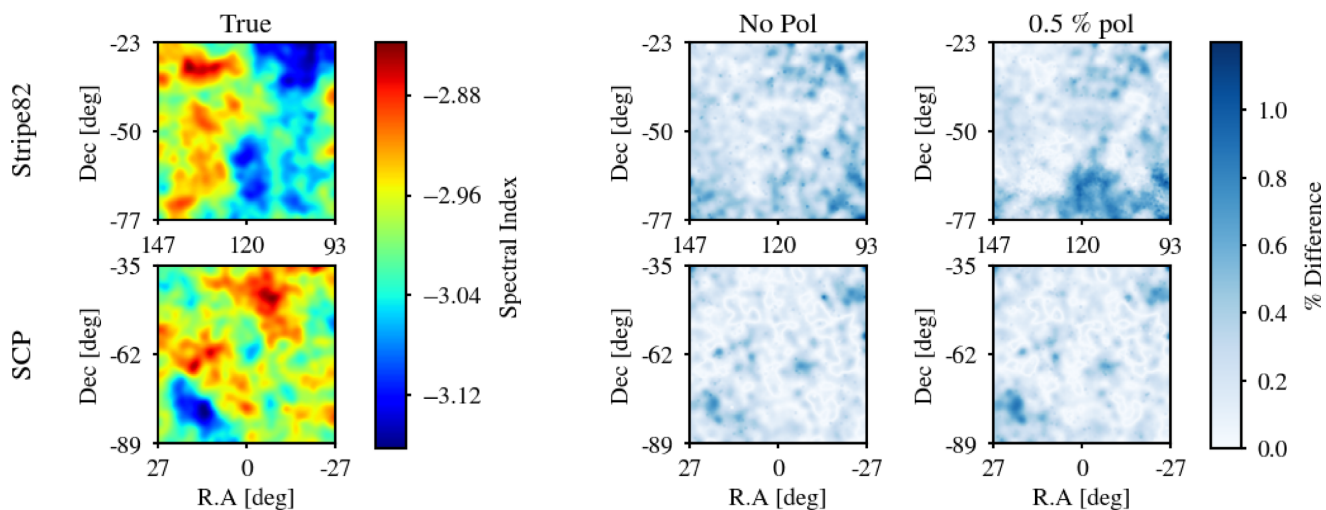


**Figure 8.** Top-panels: Normalized column vectors of the mixing matrix estimated by PCA, FASTICA, and GMCA for the Stripe82 region with 0.5 per cent polarization leakage. Note that we only show the first 6 modes for clarity despite using  $N_{\text{FG}} = 9$ . Bottom-left: resulting H I intensity map from PCA clean for a random frequency channel. Also shown are the difference maps between this PCA cleaned map and those cleaned using FASTICA (bottom-centre) and GMCA (bottom-right). The three methods estimate different mixing matrices, yet they lead to analogous foreground-dominated data cubes and produce extremely similar H I residuals as shown by the bottom maps (noting the small colour-bar scale).



**Figure 9.** Impact from different foreground cleaning methods on the 3D power spectra for both regions. We show the weighted difference  $\epsilon$  in foreground free and cleaned power spectra (equation 30) where positive (negative) values represent undercleaning (over)cleaning. Results are *without* polarization leakage (top row) and with 0.5 per cent polarization leakage (bottom row).

pre-determined  $N_{\text{FG}}$  number of modes within which to maximize statistical independence or sparsity. They achieve this by identifying functions that can share out the contributions to the variance amongst these  $N_{\text{FG}}$  modes, with no requirement of orthogonality, providing they are maximizing independence and sparsity. This allows the modes in FASTICA and GMCA to approximately follow the slope



**Figure 10.** True synchrotron spectral indices (left-hand panel) from our simulations and the absolute percentage difference between the truth and those estimated by parametric fitting for no polarization leakage (centre panel) and 0.5 per cent polarization leakage (right-hand panel). The Stripe82 region is on the top row with SCP on the bottom row.

defined by the dominant spectral indices from synchrotron and free-free emission. These functions will still contain the polarized information demonstrated in the PCA functions, but be contained as sub-dominant oscillations within these modes.

Despite the differences in the identified mixing matrices, the similarity in final results from all three blind methods can be understood by considering the difference in the assumptions they make when linearly decomposing the signal. PCA merely identify the eigenvectors corresponding to the largest eigenvalues in the frequency–frequency covariance of the data. FASTICA adds on this by promoting non-Gaussianity in the estimated sources, as a proxy for their statistical independence. GMCA promotes sparsity in the spatial domain, after having wavelet-transformed the patches, relying on highlighting the specific morphologies of the components to facilitate the source separation process. Since all three approaches result in essentially identically cleaned maps and power spectra, this leads to the conclusion that the FASTICA and GMCA assumptions do not hold for these data sets: we are dealing with fairly Gaussian (non-sparse) components in the spatial dimensions, thus FASTICA and GMCA are not optimized to improve upon their pre-processing PCA step. However, this statement cannot be generalized and is specific to our tested simulated data i.e. the size and resolution of the patches we work with. For instance, Carucci et al. (2020a) show how FASTICA and GMCA behave differently in presence of non-continuous, RFI-flagged data on the full sky. Also work is still needed to understand a realistic beam effect on the intensity maps, which could add complexity to the foreground removal process.

### 5.2 Non-blind foreground cleaning

The appeal of a parametric method is its ability to yield estimates for both the cosmological signal and each individual foreground. In our parametric approach, we have assumed knowledge of the free-free spectral index, enlisted a least-squares fit assuming pure synchrotron emission to determine the synchrotron spectral index, and finally solved a least-squares optimization to the data based on the synchrotron and free-free spectral forms. As our approach only considers synchrotron and free-free emission explicitly, all other data contributions get absorbed into either our cosmological, synchrotron or free-free estimates. Specifically, the instrumental noise is included

in our H I estimate, point sources are contained in our synchrotron estimate due to their similar spectral forms and polarization leakage is seen to degrade the quality of the H I, free-free, and synchrotron estimates.

With the free-free spectral form held at the true value, the key to our approach is accurate determination of the synchrotron spectral indices. Fig. 10 shows the absolute percentage difference maps between the true synchrotron spectral indices and those estimated by our non-blind fit. The right-hand column shows a slight increase in percentage error when polarization leakage is present. However, it appears it is only a mild prohibitive factor for accurate recovery of the synchrotron spectral index in the high Galactic latitude regions we have studied and for the polarization leakage model we have used.

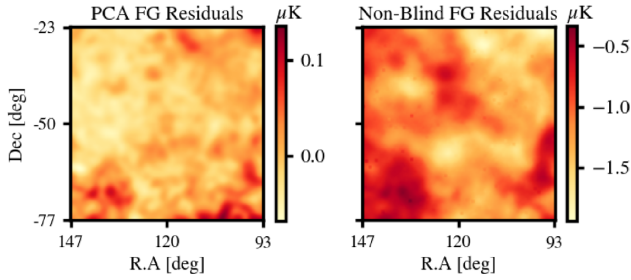
We find parametric fitting is competitive with the blind methods in the absence of polarization leakage. However, as polarization leakage is not explicitly accounted for within our parametric fit, nor does it display any degenerate behaviour with the foregrounds we do account for, we find that parametric fitting is not capable of being used ‘as is’ to get to the H I signal level. This is explicitly shown in Fig. 9; in the case of the polarized Stripe82, the non-blind residuals are too large to occupy the same axis ranges as the blind residuals. In conclusion: our parametric fit cannot compete with the blind methods in the face of polarization leakage.

### 5.3 Hybrid foreground cleaning

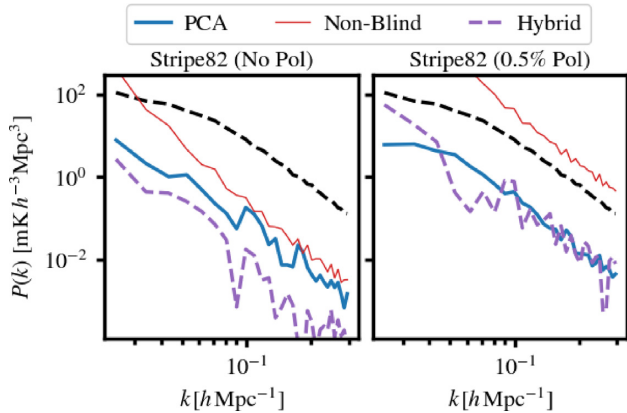
With several available methods for performing 21-cm foreground cleaning, an obvious question to ask is whether any of them can be combined into a hybrid method to produce better results. Hybridized techniques have been explored in Planck Collaboration IV (2020), where SMICA was used in a semiblind way and in Bobin et al. (2019), where GMCA was used in a semiblind way.

The approach we adopt here differs slightly since we investigate combinations of foreground cleaning methods by cross-correlating two maps cleaned using different approaches. The main potential benefits from this approach come from the possibility of the cross-correlation reducing the residual foregrounds.

Due to the inherently similar resulting maps in our three blind methods, it is not surprising that we found no benefit in combining these techniques. However, for cases where the non-blind approach



**Figure 11.** Pure foreground residuals remaining after PCA (left-hand panel) and parametric (right-hand panel) cleans on Stripe82 region without polarization leakage. We have averaged along the lines of sight.



**Figure 12.** Power spectra for foreground residuals remaining after cleaning relative to the foreground-free original H I signal (black-dashed line). Hybrid result refers to the cross-power spectrum between a PCA clean map and one cleaned using our non-blind approach.

was providing a good fit to the foregrounds and thus producing a reliable foreground clean, we found that a cross-correlation between a map cleaned with this approach and one using PCA, does produce a reduction in residual foreground correlation. We plot the pure foreground residuals (equation 24) from PCA and the non-blind method for the Stripe82 region with no polarization leakage in Fig. 11. It is clear by eye that there are some significant differences in these maps and thus cross-correlating these methods should result in a reduction from residual foreground contributions. We demonstrate this in Fig. 12 (left-hand panel) showing the contribution to the power from the pure-foreground residuals left after a clean using PCA (thick blue line) and parametric fit (thin red line). The aim is for the foreground residuals to be as low as possible, ideally far below the original H I signal (black dashed line), which we include for reference. As shown, by using a hybrid approach (dashed purple line) that cross-correlates the two differently cleaned maps, foreground residuals are reduced, although we find no such reduction in the polarized case (right-hand panel). We found no discernible improvement in the final power spectrum measurement of all the components, i.e. this current approach would make no change to the measured  $\varepsilon$ . However, this reduction in residuals is important, especially for future surveys where maximizing precision and reducing errorbars is paramount (see discussion in Section 4.5.2). This technique could also yield benefits in more realistic situations containing more systematics. If the different methods respond differently to these systematics, then reductions in residual contamination could be significant. For now, we highlight these results as an area for potential further investigation.

#### 5.4 Balancing foreground residuals and H I damping

As intensity mapping surveys continue to produce data, the focus will be on how to optimize these surveys for constraining cosmological and astrophysical parameters. Until now we have used a consistent defined choice of  $N_{\text{FG}}$  for each region and polarization case. Here, we begin to examine the consequences of varying this parameter and explore whether an optimal choice can be made. We seek a balance between overcleaning foregrounds by using a high- $N_{\text{FG}}$  that causes damping to H I power on large scales, and undercleaning using a low- $N_{\text{FG}}$  leaving higher residual foregrounds that potentially bias results or boost errors.

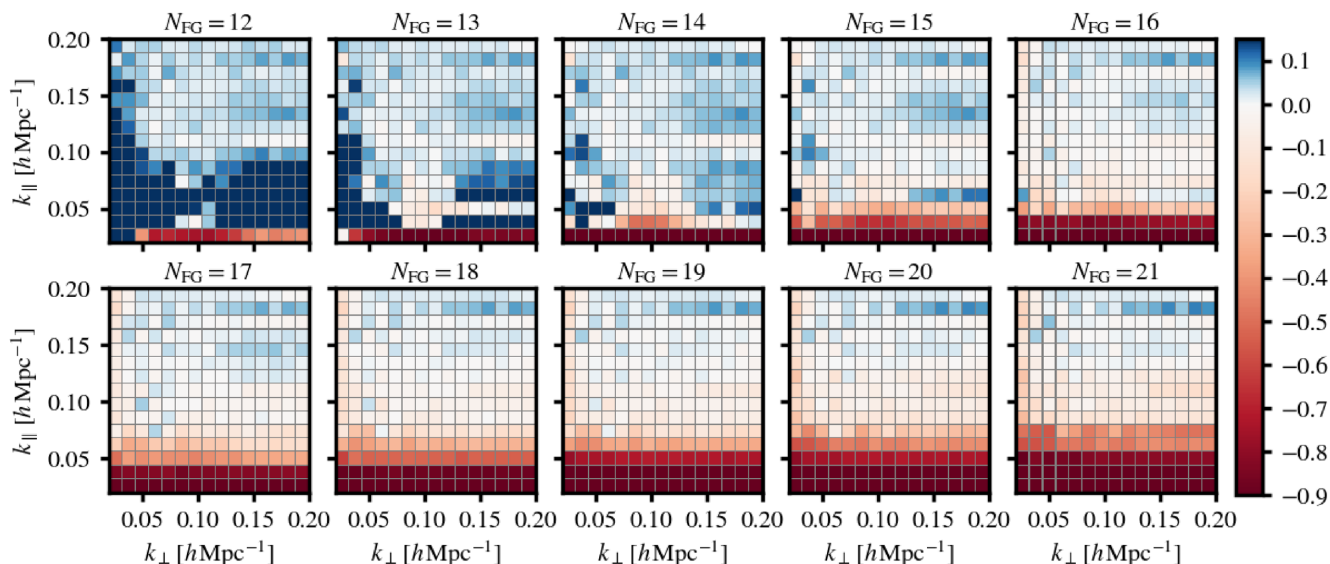
In the case of cross-correlations (which we focus on in Section 5.5) this is more straightforward to analyse since residual foregrounds, provided they are not too large, will just boost the errors in the power spectrum measurement. Thus identifying an optimal number of  $N_{\text{FG}}$  modes to remove is primarily based on minimizing errors. In autocorrelation, the process for optimizing the choice of  $N_{\text{FG}}$  is more difficult (see discussion in Section 4.5.2).

Fig. 13 shows how the 2D power spectrum evolves with an increasing  $N_{\text{FG}}$  by analysing  $\varepsilon_{2\text{D}}$ , the weighted difference between foreground-cleaned and foreground-free power spectra (equation 31). The results are only for the Stripe82 region in the case of high FR polarization leakage, cleaned using a PCA method. This demonstrates how increasing the aggressiveness of the clean mitigates foreground residuals (blue regions) but at the expense of severely damping small- $k_{\parallel}$  modes (shown by red regions).

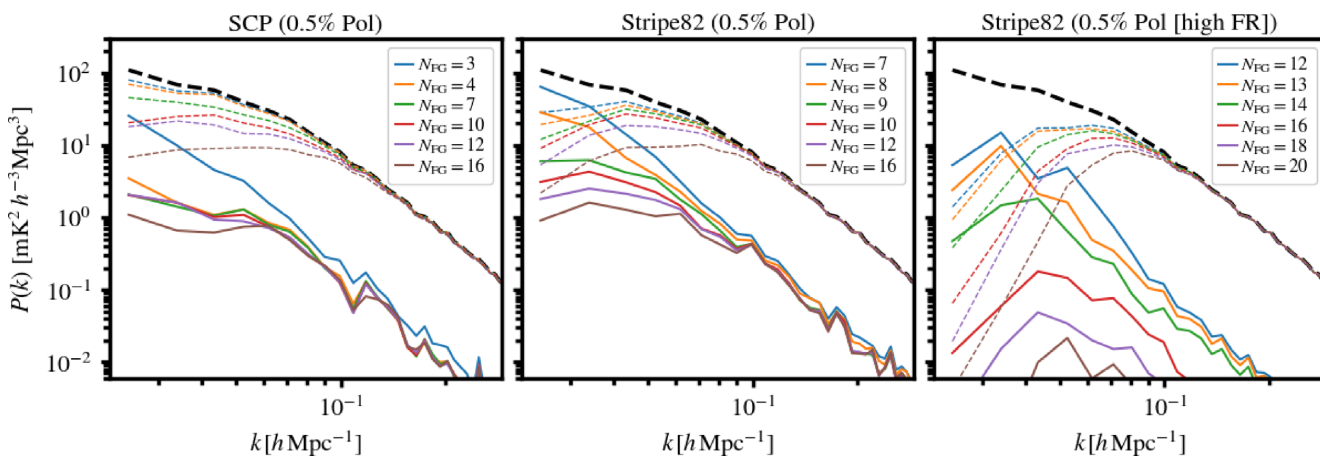
In Fig. 14, the levels of foreground residuals for both regions with 0.5 per cent polarization leakage (and also the Stripe82 high-FR case) are shown after PCA cleans with varying  $N_{\text{FG}}$ . These are calculated using the methods outlined in Section 4.5.2, which reconstruct the exact maps of the foreground-only signal remaining in the cleaned data (shown by solid lines in Fig. 14). For comparison, we also plot the H I-only power spectra (dashed lines) calculated in a similar way by projecting the H I-only simulated data along the  $N_{\text{FG}}$  subtracted eigenvectors to precisely reconstruct the residual H I in the maps after the PCA clean.

Fig. 14 shows that the residuals decrease with increasing  $N_{\text{FG}}$  as expected, but the H I signal is also damped. The aim is to reach a level where the H I signal dominates over the residual foregrounds, so they no longer bias results. It is encouraging to see that in general H I dominates by an order of magnitude across small scales ( $k \gtrsim 0.1 \text{ Mpc } h^{-1}$ ). In the Stripe82 high-FR case, a high  $N_{\text{FG}}$  is required to bring the foreground residuals below the H I-only power and even using  $N_{\text{FG}} = 14$ , foreground residuals remain at a similar level to the H I for modes with  $k \lesssim 0.04 \text{ Mpc } h^{-1}$ . We can see how results differ between the SCP and Stripe82 regions, for example in the SCP, a very mild clean ( $N_{\text{FG}} = 4$ ) is sufficient to achieve a residual foreground level an order of magnitude lower than the H I across all scales, even in this 0.5 per cent polarization leakage case.

It is interesting to note how much the residual-H I (dashed lines) differ in the high-FR case relative to the standard polarization cases (e.g. comparing the H I residual for  $N_{\text{FG}} = 16$ ) despite it being the same underlying simulated H I data. The reason for this is down to how the eigenvectors are constructed in each case and we show the first six in Fig. 15. For each of the cases, we see little distinction between the first two eigenmodes, which are likely picking out the synchrotron slope and the free-free emission. But in the SCP, the modes start to oscillate after this, which suggests that smaller scale cosmological H I or noise is leaking into these modes. Whereas in the Stripe82 high-FR case, the eigenmodes are relatively smooth (except for the long wavelength oscillations caused by polarization



**Figure 13.** Impact of a varying  $N_{\text{FG}}$  on the efficacy of the foreground clean. We show the weighted difference  $\varepsilon_{2\text{D}}$  in foreground free and cleaned 2D power spectra, defined by equation (31), where positive (negative) values represent undercleaning (over)cleaning. This is shown for a range of  $N_{\text{FG}}$  values in a PCA clean on the Stripe82 region for the high-FR polarization leakage case. We see too low  $N_{\text{FG}}$  leaves large foreground residuals and increasing  $N_{\text{FG}}$  removes these residuals at the cost of damping small  $k_{\parallel}$  modes.



**Figure 14.** Power spectra for the foreground residuals (solid lines) from a PCA clean with varying  $N_{\text{FG}}$  modes removed from each region. Left-hand and centre panels are for both regions with 0.5 per cent polarization leakage and the right-hand panel is the Stripe82 region for the high FR case. For comparison, we also plot the foreground-free original HI-only signal (black-dashed line) along with the residual HI-only power spectra (coloured-dashed lines) after each PCA clean. Ideally one would require a scenario where the residual HI dominates over the residual foregrounds. Calculations outlined in equations (24) and (23).

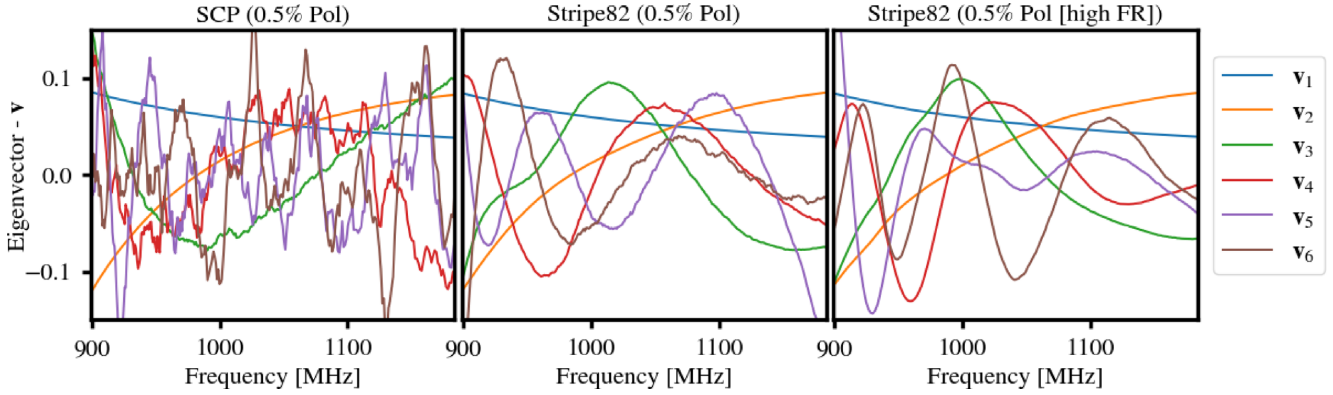
leakage – which can be compared to the polarized spectra in Fig. 3) indicating that more large-scale information will be removed. This is what we see in Fig. 14 too, if we compare e.g. the  $N_{\text{FG}} = 16$  case for the two regions. In the Stripe82 high-FR case, the small- $k$  is severely damped relative to the SCP, but comparing the scales around  $k = 0.08\text{--}0.1 h \text{Mpc}^{-1}$  we see that the Stripe82 high-FR  $N_{\text{FG}} = 16$  case actually has slightly larger power in the residual HI, owing to the fact that the eigenmodes being removed are smoother and contain less small-scale HI power.

Fig. 14 demonstrates the potential problem from additive biases caused by residual foregrounds, which will correlate in the autopower spectrum. We also see in Figs 7 and 13 how these relatively small foreground residual levels can cause non-negligible additive biases in the total measured power spectrum. Of course, in our simulated scenario we are able to decompose the contribution from residual foregrounds and HI to the total power spectrum and make a well-

informed choice on the optimal choice of  $N_{\text{FG}}$ . However, if dealing with real data, residual foregrounds and HI would not be easily separable at the required precision. This highlights a central challenge to using HI intensity mapping in autocorrelation for data that include foregrounds that cannot be efficiently removed (see discussion in Section 4.5.2).

### 5.5 Cross-correlation with optical galaxy surveys

Previous HI intensity mapping surveys with the GBT (Masui et al. 2013; Wolz et al. 2017, 2021) and Parkes telescopes (Anderson et al. 2018; Li, Staveley-Smith & Rhee 2020b) have relied on cross-correlations with optical surveys for successful detections of cosmological HI. This is both due to noise and systematic effects in these pathfinder intensity mapping experiments, but also due to foreground residuals. As we have already stated,



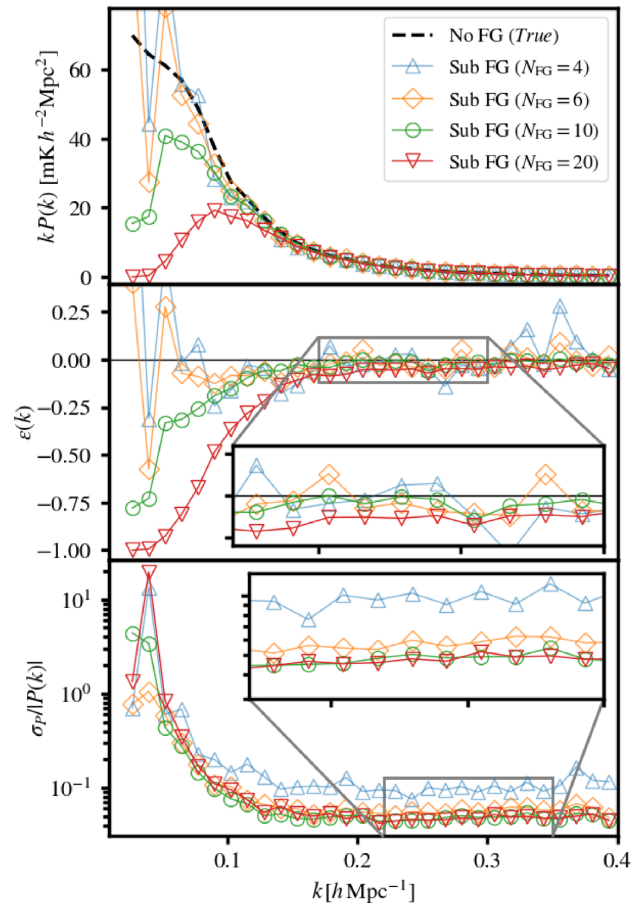
**Figure 15.** First six eigenvectors from the frequency–frequency covariance matrix for both regions with 0.5 per cent polarization leakage. Right-hand panel is for the Stripe82 region for the high-FR case.

intensity mapping simulations are typically idealized and foregrounds can be removed relatively straightforwardly. However, the real data in these early experiments have shown that to achieve low enough foreground residuals for successful detections, more aggressive foreground cleans are required compared to simulations. This is likely due to calibration issues and effects from polarization leakage or chromatic beams, which can cause some frequency decoherence in the otherwise continuous foreground signals.

Cross-correlations with optical galaxy surveys are important because they allow for systematics to be mitigated (they drop out in cross-correlation), and a detection can be more easily achieved. We demonstrate this process with our simulated data sets and in doing so we can investigate the optimal level of foreground cleaning required.

To do this we exclusively use our most dominant foreground region (the Galactic Plane), including polarization leakage. This is an attempt to mimic real-data experiments, which as discussed often need high levels of foreground cleaning ( $\sim 10\text{--}20 N_{\text{FG}}$  modes removed from the data). As we have demonstrated in Fig. 6 (third panel) and Fig. 13, the H I autocorrelation is highly affected by the presence of such dominant foregrounds. Either dominant foreground residuals remain in the data from choosing  $N_{\text{FG}}$  which is too low, or the largest scales are completely destroyed from choosing a higher  $N_{\text{FG}}$ .

Fig. 16 shows the improvements that can be made with cross-correlations. Here, we cross-correlate the simulated optical data (outlined in Section 3.2) with the H I intensity map data contaminated with a polarized Galactic Plane foreground and PCA cleaned. The top-panel immediately shows that a foreground clean with much fewer modes removed is sufficient for a cross-correlation measurement that has reasonable agreement with the no-foreground case. This is in spite of the large levels of residual foregrounds that will inevitably be remaining from such a mild clean. This is also shown in the middle panel where the weighted difference  $\epsilon$  (equation 30) is plotted for each variant of  $N_{\text{FG}}$ . This shows that  $N_{\text{FG}} = 4$  and  $N_{\text{FG}} = 6$  deliver a reasonably consistent agreement across all scales with the foreground-free power spectrum. This also shows how more aggressive cleans still damp the power spectrum even in this case of cross-correlation. This is especially noticeable at large scales (small- $k$ ) but even at mid-range scales in the zoomed-in section at  $0.17 < k < 0.3 h \text{ Mpc}^{-1}$ , we can see power being damped as  $N_{\text{FG}}$  is increased.



**Figure 16.** Results from the cross-correlation of optical galaxy data with H I intensity maps cleaned using PCA with a range of  $N_{\text{FG}}$ . (Middle panel) shows  $\epsilon(k)$ , the weighted difference defined by equation (30), where positive (negative) values represent undercleaning (over)cleaning. We show the most foreground affected case, the Stripe82 region with high-FR polarization leakage. Comparison with previous results for this region, where an aggressive clean is needed and severely damps power in H I autocorrelation, demonstrates how much cross-correlation improves results. Bottom-panel shows that a less aggressive clean (lower  $N_{\text{FG}}$ ) generally results in higher fractional errors  $\sigma_P/P(k)$ .

It is perhaps tempting to conclude that an optimal choice of  $N_{\text{FG}}$  in cross-correlations can be entirely based on what delivers the best agreement with the foreground-free data, i.e. the  $\varepsilon$  which is closest to zero. However, as is already slightly discernible from the middle-panel, a low choice of  $N_{\text{FG}}$  does provide a higher variance in the results, as shown from the  $N_{\text{FG}} = 4$  case whose value for  $\varepsilon$  fluctuates more than all other cases. This can be understood, by considering that in this situation a large amount of foreground residuals will be in the cleaned intensity map which cause more random and spurious correlations between the optical data and the foreground residuals. In other words, the higher levels of foreground residuals resulting from a lower  $N_{\text{FG}}$  clean, will inevitably boost uncertainties in cross-correlations.

We investigated this further in the bottom panel of Fig. 16 which show the estimated fractional errors  $\sigma_p/|P(k)|$  for each data point. We calculated these using the same approaches used in real-data studies (e.g. Masui et al. 2013; Anderson et al. 2018; Wolz et al. 2021) by treating our MULTIDARK simulated data sets as real data. Alongside this we produced 100 lognormal mocks for both the H I intensity maps and optical galaxy maps. From this we can calculate a foreground transfer function (see Section 4.5.1). Applying this transfer function to each of the mocks and then measuring the variance in their results provides an estimate for the power spectrum errors  $\sigma_p$ .

The bottom panel of Fig. 16 shows that errors are generally largest for the  $N_{\text{FG}} = 4$  case, highlighting the important point that foreground residuals will boost errors. At low- $k$ , we also see the fractional errors are extremely high for the  $N_{\text{FG}} = 20$  case. This is because the power at these scales is damped so severely that the transfer function is essentially trying to recover a power spectrum from a position of  $P(k) \sim 0$ , which inevitably causes noisy results and boosts the variance. This demonstrates the balance required for an optimal choice of  $N_{\text{FG}}$ . An  $N_{\text{FG}}$  which is too low causes too much foreground residual leading to large errors. An  $N_{\text{FG}}$  which is too high, damps the power spectrum drastically causing too much scatter in the power recovered through the transfer function. There is also a requirement for an  $N_{\text{FG}}$  balance at higher  $k$ -ranges. Looking at the zoomed-in section at  $0.22 < k < 0.35 h \text{ Mpc}^{-1}$  we clearly see errors are highest for  $N_{\text{FG}} = 4$ . They begin to decrease with an increasing  $N_{\text{FG}}$  but we eventually find this saturates and there is little or no improvement in error even with a big jump from  $N_{\text{FG}} = 10$  to  $N_{\text{FG}} = 20$ . This means going arbitrarily high in  $N_{\text{FG}}$  will slowly stop improving errors but continue to bias results (increase  $|\varepsilon|$ ).

The results from Fig. 16 are therefore strong evidence that a fine balance must be reached for an optimal choice of  $N_{\text{FG}}$  in cross-correlations. This choice will also depend on the cosmological parameters being probed. For example, an investigation into primordial non-Gaussianity involves attempts to constrain the parameter  $f_{\text{NL}}$ , which requires large scales (small- $k$ ) measurements. Going to  $N_{\text{FG}} \sim 10$  in an attempt to minimize errors may not be plausible in this situation if the bias induced on these large scales is too strong. Conversely, probing something like the H I abundance ( $\Omega_{\text{H I}}$ ), which generally just affects the amplitude scaling of the power spectrum and can thus be probed at most scales, could potentially allow for a more aggressive clean that controls errors but still does not heavily bias the higher- $k$  scales where this parameter can still be constrained. Therefore, it is unlikely that a universally optimal foreground treatment can ever be selected. This also supports conclusions from previous work that attempted to model the effects of foreground cleaning e.g. Cunnington et al. (2020b) and Soares et al. (2021). These works employed subtly different foreground modelling, which is likely due to the different range of scales they targeted given their science goals.

## 6 DISCUSSION

Evidence from pathfinder 21-cm intensity mapping data (Masui et al. 2013; Wolz et al. 2017, 2021; Anderson et al. 2018) suggests that we will, at least initially, be requiring fairly aggressive foreground cleans in H I intensity mapping surveys. This is potentially due to instrumental responses to foregrounds causing effects such as polarization leakage. Understanding the impact this has on probes of large-scale cosmic structure using H I intensity maps is therefore paramount. In this paper, we have provided a study into these issues by presenting a set of test data with a differing range of foreground contamination, both with and without effects from polarization leakage. We stress that the polarization leakage model we use is far from being a precise emulation of these complex effects, nevertheless it captures the main issue with this systematic: its non-smooth and spatial-dependent behaviour in frequency, and a more advanced model is currently lacking.

Given our limited understanding of polarization effects on intensity mapping, we have aimed to test extreme cases where a high- $N_{\text{FG}}$  clean is required, which in this sense, is similar to early pathfinder experiments. This provides a means to begin investigating some of these issues in a simulated setting where we have full control and can separate contributions to the final observed signal. We found that a varying polarization leakage fraction does not create problems for a foreground clean (see Fig. 7) and an increasing fraction mainly just increases the amplitude of the eigenvalues one would remove anyway in a foreground clean (Fig. 4). The main issue comes from the frequency decoherence to the foreground spectra which could plausibly be higher than predicted by the CRIME model we use. To investigate this further, we created a high FR test using the polarization leakage output from the Galactic plane region of CRIME. Whilst this mixing of regions does not represent a physically realistic scenario, it provides the necessary test case to investigate these problems with the models currently available.

The contamination to the data from a foreground clean can manifest in two distinct ways: damping of cosmological H I modes, and foreground residuals.

### 6.1 Damping cosmological H I

Inevitably, some cosmological H I modes will be degenerate with the 21-cm foregrounds and their information will be contained in the same modes that are being removed. This has the effect of damping the H I power spectrum mostly on large scales (small- $k$ ). For our mild cases without polarization leakage (and even the SCP with polarization leakage), this effect is minimal (see Figs 6 and 7). However, where higher  $N_{\text{FG}}$  cleans are needed, as was required for our high FR cases, this damping becomes severe and generally isolated to small- $k_{\parallel}$  modes (see Fig. 13). In all cases though, a foreground transfer function (Section 4.5.1) should be able to compensate for this damping or alternatively, a foreground model with free nuisance parameters that can be marginalized over (Soares et al. 2021), although neither approaches avoid the inevitable reduced sensitivity caused by the foreground clean.

### 6.2 Foreground residuals

Another challenging effect from foreground contamination comes from residuals left in the data after a clean. In mild cases where foregrounds are removed relatively easily, foreground residuals are small but we have shown evidence that there can still be an additive bias (see high- $k$  values in Fig. 9 with positive  $\varepsilon$  values). However,

previous work (e.g. Cunnington et al. 2020a; Soares et al. 2021) has shown that unbiased cosmological parameter estimates can be obtained even when contribution from residuals is not included, thus showing that for mild cases with low enough residuals, their effect is minimal. However, our results indicate that foreground residuals can be exacerbated by polarization leakage. For example, Fig. 14 demonstrated how even with large  $N_{\text{FG}}$ , the foreground residuals can still have a similar amplitude to the remaining H I signal in some cases. This would cause additive biases and boost errors and needs to be modelled, as we outlined in equation (28) and equation (29). Quantifying the contribution from foreground residuals is not trivial with real data and this will be a key challenge for autocorrelation measurements. However, for cross-correlations with an overlapping optical galaxy survey, the situation is somewhat simplified. Here, the foreground residuals do not correlate with the foreground-free galaxy data, allowing less extreme cleans at the expense of a boost to errors (as demonstrated by Fig. 16).

Our investigation suggests that drawing general conclusions or recommendations for an optimal foreground treatment is not possible. An optimal method depends on the region being targeted, the instrumental calibration (e.g. susceptibility to polarization leakage), and also on the scales being targeted depending on the survey's key science goals. But if one can achieve near-perfect instrument calibration, the results do appear fairly general. The first column of Fig. 7 shows all regions can be cleaned by blindly removing three principal components and this delivers similar damping to H I power and a similar level of residual bias. However, residuals may differ between regions for smaller- $k$  but without affecting the accuracy (measured by  $\epsilon$ ) and instead just affect the precision (boosting errors).

We introduced and tested three commonly employed blind foreground cleaning techniques; FASTICA, GMCA, and PCA, the latter being mathematically equivalent to SVD (Section 4.1.1), and a polynomial fit. We found all three blind methods deliver essentially equivalent results in all cases. We discussed this in detail in Section 5.1 – in summary, this is due to FASTICA and GMCA performing an initial PCA which they then try to improve upon by imposing spatial statistical independence and sparsity, respectively. However, perhaps due to the sky size we use and the resolution, the foregrounds not included in the initial PCA reconstruction are not sufficiently sparse or non-Gaussian to be identified by GMCA or FASTICA and no discernible improvement is made.

We also trialed a non-blind approach to foreground removal (Section 4.4). Our tests revealed that this method can potentially be competitive with a blind approach in the absence of polarization leakage (see Fig. 9). However, the method in its current form is not robust to polarization leakage effects and performs poorly when this is included. Thus, this approach would be reliant on a near-perfect calibration of the intensity mapping instrument. For these reasons, this approach would only likely be viable for future surveys where calibration strategies are highly optimized and astrophysical parameters can be tightly constrained, allowing more precise fits to the foregrounds. A potential benefit from this is the possibility of further combinations with a full-blind approach in a hybridization. Section 5.3 examined the cross-correlation between a non-blind cleaned map and a PCA cleaned one. This revealed that while little improvement can be gained in the accuracy of the final power spectrum, the foreground residuals in the two are subtly different and result in a lower contribution in the cross-measurement (Fig. 12).

We hope our findings can be useful for analysing H I intensity mapping data from the MeerKAT intensity mapping survey (Santos et al. 2017; Li et al. 2020a; Wang et al. 2020), and for preparing cross-correlation strategies for MeerKAT and the SKA (Carucci,

Villaescusa-Navarro & Viel 2017; Pourtsidou, Bacon & Crittenden 2017; SKA Cosmology SWG et al. 2020).

## ACKNOWLEDGEMENTS

We are grateful to Paula Soares for helpful comments on the draft. SC was supported by STFC grant ST/S000437/1. JB, IPC, and MI are supported by the European Union through the grant LENA (ERC StG no. 678282) within the H2020 Framework Program. AP is a UK Research and Innovation Future Leaders Fellow, grant MR/S016066/1, and also acknowledges support by STFC grant ST/S000437/1. This research utilized Queen Mary's Apocrita HPC facility, supported by QMUL Research-IT <http://doi.org/10.5281/zenodo.438045>. We acknowledge the use of open source software (Jones et al. 2001; Hunter 2007; McKinney 2010; Van Der Walt, Colbert & Varoquaux 2011; Harris et al. 2020). Some of the results in this paper have been derived using the HEALPY and HEALPIX package. We thank New Mexico State University (USA) and Instituto de Astrofísica de Andalucía CSIC (Spain) for hosting the Skies and Universes site for cosmological simulation products.

## DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author.

## REFERENCES

- Alonso D., Ferreira P. G., Santos M. G., 2014, *MNRAS*, 444, 3183  
 Alonso D., Bull P., Ferreira P. G., Santos M. G., 2015, *MNRAS*, 447, 400  
 Anderson C. et al., 2018, *MNRAS*, 476, 3382  
 Ansari R. et al., 2012, *A&A*, 540, A129  
 Asorey J. et al., 2020, *MNRAS*, 495, 1788  
 Battye R. A., Davies R. D., Weller J., 2004, *MNRAS*, 355, 1339  
 Battye R. A., Browne I. W. A., Dickinson C., Heron G., Maffei B., Pourtsidou A., 2013, *MNRAS*, 434, 1239  
 BeyondPlanck Collaboration I, 2020, preprint ([arXiv:2011.05609](https://arxiv.org/abs/2011.05609))  
 Bharadwaj S., Nath B., Nath B. B., Sethi S. K., 2001, *J. Astrophys. Astron.*, 22, 21  
 Bigot-Sazy M. A. et al., 2015, *MNRAS*, 454, 3240  
 Blake C., Carter P., Koda J., 2018, *MNRAS*, 479, 5168  
 Bobin J., Starck J.-L., Fadili J., Moudden Y., 2007, *IEEE Transact. Image Process.*, 16, 2662  
 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, A105  
 Bobin J., Acero F., Picquetot A., 2019, 2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)  
 Carucci I. P., Villaescusa-Navarro F., Viel M., 2017, *J. Cosmol. Astropart. Phys.*, 2017, 001  
 Carucci I. P., Irfan M. O., Bobin J., 2020a, *MNRAS*, 499, 304  
 Carucci I. P., Irfan M. O., Bobin J., 2020b, *MNRAS*, 499, 304  
 Castorina E., White M., 2018, *MNRAS*, 476, 4403  
 Chang T.-C., Pen U.-L., Peterson J. B., McDonald P., 2008, *Phys. Rev. Lett.*, 100, 091303  
 Chapman E. et al., 2012, *MNRAS*, 423, 2518  
 Cora S. A., 2006, *MNRAS*, 368, 1540  
 Croton D. J. et al., 2016, *ApJS*, 222, 22  
 Cunnington S., Wolz L., Pourtsidou A., Bacon D., 2019, *MNRAS*, 488, 5452  
 Cunnington S., Pourtsidou A., Soares P. S., Blake C., Bacon D., 2020a, *MNRAS*, 496, 415  
 Cunnington S., Camera S., Pourtsidou A., 2020b, *MNRAS*, 499, 4054  
 Delabrouille J., Cardoso J.-F., Patanchon G., 2003, *MNRAS*, 346, 1089  
 Dickinson C., Davies R. D., Davis R. J., 2003, *MNRAS*, 341, 369

- Eriksen H. K., Jewell J. B., Dickinson C., Banday A. J., Górski K. M., Lawrence C. R., 2008, *ApJ*, 676, 10
- Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, *PASP*, 125, 306
- Harper S., Dickinson C., 2018, *MNRAS*, 479, 2024
- Harper S., Dickinson C., Battye R., Roychowdhury S., Browne I., Ma Y.-Z., Olivari L., Chen T., 2018, *MNRAS*, 478, 2416
- Harris C. R. et al., 2020, *Nature*, 585, 357–362
- Hunter J. D., 2007, *Comput. Sci. Eng.*, 9, 90
- Hyvärinen A., 1999, *IEEE Transact. Neural Netw.*, 10 3, 626
- Hyvärinen A., Oja E., 2000, *Neural Netw.*, 13, 411
- Jelic V. et al., 2008, *MNRAS*, 389, 1319
- Jelic V., Zaroubi S., Labropoulos P., Bernardi G., de Bruyn A. G., Koopmans L. V. E., 2010, *MNRAS*, 409, 1647
- Jones E., Oliphant T., Peterson P. et al., 2001, SciPy: Open Source Scientific Tools for Python. <http://www.scipy.org/>
- Klypin A., Yepes G., Gottlober S., Prada F., Hess S., 2016, *MNRAS*, 457, 4340
- Knebe A. et al., 2018, *MNRAS*, 474, 5206
- Lacy M. et al., 2020, *Publ. Astron. Soc. Pac.*, 132, 035001
- Leach S. et al., 2008, *A&A*, 491, 597
- Li Y., Santos M. G., Grainge K., Harper S., Wang J., 2021up, *MNRAS*, 501, 4344
- Li L., Staveley-Smith L., Rhee J., 2020b, *Res. Astron. Astrophys.*, 21, 030
- Liao Y.-W., Chang T.-C., Kuo C.-Y., Masui K. W., Oppermann N., Pen U.-L., Peterson J. B., 2016, *Astrophys. J.*, 833, 289
- Liu A., Shaw J. R., 2020, *Publ. Astron. Soc. Pac.*, 132, 062001
- Liu A., Tegmark M., 2011, *Phys. Rev. D*, 83, 103006
- Mandelbaum R. et al., 2011, *MNRAS*, 410, 844
- Masui K. W. et al., 2013, *ApJ*, 763, L20
- Matshawule S. D., Spinelli M., Santos M. G., Ngobese S., 2020, preprint ([arXiv:2011.10815](https://arxiv.org/abs/2011.10815))
- McKinney W., 2010, in van der Walt S., Millman J., eds, Proceedings of the 9th Python in Science Conference. p. 51
- Mertens F., Ghosh A., Koopmans L., 2018, *MNRAS*, 478, 3640
- Mertens F. et al., 2020, *MNRAS*, 493, 1662
- Miville-Deschênes M.-A., Ysard N., Lavabre A., Ponthieu N., Macías-Pérez J. F., Aumont J., Bernard J. P., 2008, *A&A*, 490, 1093
- Moore D. F., Aguirre J. E., Parsons A. R., Jacobs D. C., Pober J. C., 2013, *Astrophys. J.*, 769, 154
- Olivari L. C., 2018, PhD thesis, Univ. Manchester
- Olivari L. C., Remazeilles M., Dickinson C., 2016, *MNRAS*, 456, 2749
- Olivari L. C., Dickinson C., Battye R. A., Ma Y. Z., Costa A. A., Remazeilles M., Harper S., 2018, *MNRAS*, 473, 4242
- Oppermann N. et al., 2012, *A&A*, 542, A93
- Patil A. H. et al., 2017, *ApJ*, 838, 65
- Pedregosa F. et al., 2011, *J. Mach. Learn. Res.*, 12, 2825
- Picquetot A., Acero F., Bobin J., Maggi P., Ballet J., Pratt G. W., 2019, *A&A*, 627, A139
- Planck Collaboration XIII, 2016, *A&A*, 594, A13
- Planck Collaboration IV, 2020, *A&A*, 641, A4
- Pourtsidou A., Bacon D., Crittenden R., 2017, *MNRAS*, 470, 4251
- Remazeilles M., Dickinson C., Banday A. J., Bigot-Sazy M.-A., Ghosh T., 2015, *MNRAS*, 451, 4311
- Santos M. et al., 2015, Proc. Sci., Cosmology from a SKA HI intensity mapping survey. SISSA, Trieste, PoS#19
- Santos M. G. et al., 2017, Proceedings, MeerKAT Science: On the Pathway to the SKA (MeerKAT2016). Stellenbosch, South Africa
- Santos M. G., Cooray A., Knox L., 2005, *ApJ*, 625, 575
- Shaw J., Sigurdson K., Sitwell M., Stebbins A., Pen U.-L., 2015, *Phys. Rev. D*, 91, 083514
- SKA Cosmology SWG et al., 2020, *Publ. Astron. Soc. Austral.*, 37, e007
- Soares P. S., Cunnington S., Pourtsidou A., Blake C., 2021, *MNRAS*, 502, 2549
- Spinelli M., Zoldan A., De Lucia G., Xie L., Viel M., 2020, *MNRAS*, 493, 5434
- Switzer E. 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, preprint ([arXiv:1102.1523](https://arxiv.org/abs/1102.1523))
- Villaescusa-Navarro F. et al., 2018, *ApJ*, 866, 135
- Villaescusa-Navarro F., Alonso D., Viel M., 2017, *MNRAS*, 466, 2736
- Wang J. et al., 2020, preprint ([arXiv:2011.13789](https://arxiv.org/abs/2011.13789))
- Wolz L. et al., 2017, *MNRAS*, 464, 4938
- Wolz L. et al., 2021, preprint ([arXiv:2102.04946](https://arxiv.org/abs/2102.04946))
- Wolz L., Abdalla F., Blake C., Shaw J., Chapman E., Rawlings S., 2014, *MNRAS*, 441, 3271
- Zuo S., Chen X., Ansari R., Lu Y., 2018, *Astron. J.*, 157, 4

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