Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **525,** 3439–3454 (2023) https://doi.org/10.1093/mnras/stad2495
Advance Access publication 2023 August 21

Towards 21-cm intensity mapping at $z = 2.28$ with uGMRT using the tapered gridded estimator – III. Foreground removal

Kh. Md. Asif Elahi [1]★ Somnath Bharadwaj,[1]★ Srijita Pal,[1] Abhik Ghosh [2] Sk. Saiyad Ali,[3] Samir Choudhuri,[4] Arnab Chakraborty,[5] Abhirup Datta,[6] Nirupam Roy,[7] Madhurima Choudhury [8] and Prasun Dutta[9]

[1]*Department of Physics and Centre for Theoretical Studies, IIT Kharagpur, Kharagpur 721302, India*
[2]*Department of Physics, Banwarilal Bhalotia College, Asansol, West Bengal 713303, India*
[3]*Department of Physics, Jadavpur University, Kolkata 700032, India*
[4]*Centre for Strings, Gravitation and Cosmology, Department of Physics, Indian Institute of Technology Madras, Chennai 600036, India*
[5]*Department of Physics and McGill Space Institute, McGill University, Montreal, QC H3A 2T8, Canada*
[6]*Discipline of Astronomy, Astrophysics and Space Engineering, Indian Institute of Technology Indore, Indore 453552, India*
[7]*Department of Physics, Indian Institute of Science, Bangalore 560012, India*
[8]*ARCO (Astrophysics Research Center), Department of Natural Sciences, The Open University of Israel, 1 University Road, PO Box 808, Ráanana 4353701, Israel*
[9]*Department of Physics, IIT (BHU), Varanasi 221005, India*

Accepted 2023 August 14. Received 2023 August 14; in original form 2023 February 21

ABSTRACT

Neutral hydrogen (H I) 21-cm intensity mapping (IM) is a promising probe of the large-scale structures in the Universe. However, a few orders of magnitude brighter foregrounds obscure the IM signal. Here, we use the tapered gridded estimator to estimate the multifrequency angular power spectrum $C_\ell(\Delta\nu)$ from a 24.4-MHz bandwidth upgraded Giant Metrewave Radio Telescope Band 3 data at 432.8 MHz. In $C_\ell(\Delta\nu)$ foregrounds remain correlated across the entire $\Delta\nu$ range, whereas the 21-cm signal is localized within $\Delta\nu \leq [\Delta\nu]$ (typically, 0.5–1 MHz). Assuming the range $\Delta\nu > [\Delta\nu]$ to have minimal 21-cm signal, we use $C_\ell(\Delta\nu)$ in this range to model the foregrounds. This foreground model is extrapolated to $\Delta\nu \leq [\Delta\nu]$, and subtracted from the measured $C_\ell(\Delta\nu)$. The residual $[C_\ell(\Delta\nu)]_{\rm res}$ in the range $\Delta\nu \leq [\Delta\nu]$ is used to constrain the 21-cm signal, compensating for the signal loss from foreground subtraction. $[C_\ell(\Delta\nu)]_{\rm res}$ is found to be noise-dominated without any trace of foregrounds. Using $[C_\ell(\Delta\nu)]_{\rm res}$, we constrain the 21-cm brightness temperature fluctuations $\Delta^2(k)$, and obtain the 2σ upper limit $\Delta^2_{\rm UL}(k) \leq (18.07)^2$ mK^2 at $k = 0.247$ Mpc^{-1}. We further obtain the 2σ upper limit $[\Omega_{\rm H\,I} b_{\rm H\,I}]_{\rm UL} \leq 0.022$, where $\Omega_{\rm H\,I}$ and $b_{\rm H\,I}$ are the comoving H I density and bias parameters, respectively. Although the upper limit is nearly 10 times larger than the expected 21-cm signal, it is 3 times tighter over previous works using foreground avoidance on the same data.

Key words: methods: data analysis – methods: statistical – techniques: interferometric – diffuse radiation – large-scale structure of Universe.

1 INTRODUCTION

The bulk of the neutral hydrogen (H I) in the post-epoch of reionization (post-EoR) era resides in dense clumps, which are seen as damped Ly-α systems in quasar absorption spectra (Lanzetta, Wolfe & Turnshek 1995; Wolfe et al. 1995; Noterdaeme et al. 2012; Zafar et al. 2013; Ho, Bird & Garnett 2021). These dense H I clumps are believed to be primarily associated with galaxies. The 21-cm intensity mapping (IM) technique aims to measure the integrated redshifted 21-cm emission from the H I distribution instead of individually resolving the distant and faint galaxies (Bharadwaj & Sethi 2001; Bharadwaj, Nath & Sethi 2001). The IM signal is expected to trace the underlying matter distribution, authorizing it as an excellent probe of the cosmological large-scale structures (Bharadwaj & Srikant 2004; Bharadwaj & Ali 2005; Loeb & Wyithe 2008; Bagla, Khandai & Datta 2010; Ansari et al. 2012). Apart from that, the 21-cm IM signal can constrain the evolution of dark energy from the baryon acoustic oscillation (Chang et al. 2008; Wyithe, Loeb & Geil 2008; Battye et al. 2013; Bull et al. 2015b) measurements, put independent limits to various cosmological parameters (Bharadwaj, Sethi & Saini 2009; Visbal, Loeb & Wyithe 2009), quantify non-Gaussianity (Saiyad Ali, Bharadwaj & Pandey 2006; Hazra & Guha Sarkar 2012), and constrain the epoch of reionization (EoR) models (Long et al. 2022).

Appreciating the ample amount of cosmological possibilities, several ongoing experiments, such as the Giant Metrewave Radio Telescope (GMRT[1]; Swarup et al. 1991), MeerKAT[2] (Kennedy & Bull 2021; Cunnington et al. 2023b), the Canadian Hydrogen

★ E-mail: asifelahi999@gmail.com (KMAE); somnath@phy.iitkgp.ac.in (SB)

[1]https://www.gmrt.ncra.tifr.res.in/
[2]https://www.sarao.ac.za/science/meerkat/

Intensity Mapping Experiment (CHIME[3]; CHIME Collaboration 2022), the Ooty Wide Field Array[4] (Subrahmanya, Manoharan & Chengalur 2017), and the Hydrogen Intensity and Real-time Analysis eXperiments (HIRAX[5]; Newburgh et al. 2016), aim at measuring the post-EoR H I IM signal. The forthcoming Square Kilometre Array (SKA-mid[6]; Bull et al. 2015a) and the proposed Packed Ultra-wideband Mapping Array (Slosar et al. 2019) further promise efficient and sensitive measurements of the H I distribution across a wide redshift range.A few experiments (e.g. Chang et al. 2010; Switzer et al. 2013; Wolz et al. 2021; Amiri et al. 2023; Cunnington et al. 2023a) have successfully detected the IM signal at the low redshifts ($z < 1.3$) by cross-correlating the H I maps with galaxy surveys (such as eBOSS; Dawson et al. 2016).Paul et al. (2023) have recently reported a direct detection of the IM signal at $z \approx 0.32$ and $z \approx 0.44$ using the MeerKAT interferometer. However, a high-redshift, autocorrelation detection of the faint 21-cm IM signal is largely confronted by the 4–5 orders of magnitude brighter Galactic and extragalactic foregrounds (see e.g. Shaver et al. 1999; Ali, Bharadwaj & Chengalur 2008; Ghosh et al. 2012).

The foregrounds, which are believed to originate from continuum sources, are expected to be spectrally smoother than the H I 21-cm signal, which is a line emission (Bharadwaj et al. 2001; Di Matteo et al. 2002). The Multifrequency Angular Power Spectrum (MAPS; Zaldarriaga, Furlanetto & Hernquist 2004; Santos, Cooray & Knox 2005; Datta, Choudhury & Bharadwaj 2007) $C_\ell(\nu_a, \nu_b)$ jointly characterizes the angular (ℓ) and spectral (ν) two-point statistics of the sky signal and promises the ability to tell apart the foregrounds from the desired IM signal (Liu & Tegmark 2012; Trott et al. 2022). The MAPS $C_\ell(\Delta\nu)$, with $\Delta\nu = |\nu_a - \nu_b|$, is adequate to assess the IM signal when the frequency bandwidth under consideration is sufficiently small (Mondal, Bharadwaj & Datta 2018; Mondal et al. 2019). Considering $C_\ell(\Delta\nu)$, the foregrounds are expected to remain significantly correlated across the bandwidth, whereas the 21-cm signal is expected to decorrelate rapidly with increasing $\Delta\nu$ (Bharadwaj & Pandey 2003; Bharadwaj & Ali 2005; Datta et al. 2007). In principle, the foregrounds and the signal can be separated using these distinct decorrelation properties of MAPS (Ghosh et al. 2011a, b). However, various factors like baseline migration (Morales et al. 2012), ionospheric fluctuations, and calibration errors (Kumar, Dutta & Roy 2020; Kumar et al. 2022) introduce frequency-dependent structures. In particular, the point sources located at the periphery of the primary beam pattern and at the side lobes show up as oscillations that are extremely difficult to model and remove from the $C_\ell(\Delta\nu)$ (Ghosh et al. 2011a). Further, the period of these oscillations decreases with increasing ℓ (Ghosh et al. 2011a) – a feature that equivalently appears as the 'foreground wedge' in the cylindrical power spectrum (PS) $P(k_\perp, k_\parallel)$ (Datta, Bowman & Carilli 2010; Morales et al. 2012; Trott, Wayth & Tingay 2012; Vedantham, Udaya Shankar & Subrahmanyan 2012; Pober et al. 2016). The reader is referred to Pal et al. (2022, hereafter Paper I) for a discussion on how the oscillation in $C_\ell(\Delta\nu)$ and the wedge are related. Many existing methods attempt to remove the foregrounds from the measured visibilities (e.g. Datta et al. 2010; Paciga et al. 2011; Chapman et al. 2012; Mertens, Ghosh & Koopmans 2018; Trott et al. 2022), assuming the smooth nature of the foregrounds. On the other hand, several works (e.g. Pober et al. 2013, 2014; Dillon et al. 2014, 2015; Liu, Parsons & Trott 2014a, b; Pal et al. 2021; Abdurashidova et al. 2022) have adopted the 'foreground avoidance' strategy where only the region outside the foreground wedge is used to estimate the 21-cm PS.

A novel approach to mitigate the wide-field foreground effects was first proposed by Ghosh et al. (2011b), who tapered the interferometer's sidelobe response to mitigate the oscillations in $C_\ell(\Delta\nu)$. The tapered gridded estimator (TGE) incorporates the tapering by convolving the visibilities with a suitably chosen window function (Choudhuri et al. 2014, 2016a, b). The TGE manifestly has three salient features. First, it tapers the sky response to mitigate the oscillations in $C_\ell(\Delta\nu)$, thereby significantly reducing foreground contamination. Secondly, TGE uses gridded visibilities, reducing computational expenses and improving the signal-to-noise ratio (SNR). Thirdly, the TGE internally estimates the noise bias for an unbiased PS estimation. A two-dimensional version of TGE has been extensively used to study the angular PS of the diffused Galactic foregrounds (Choudhuri et al. 2017, 2020; Chakraborty et al. 2019a; Mazumder et al. 2020) and also Galactic magnetohydrodynamic turbulence (Saha et al. 2019, 2021). Recently, a Tracking TGE (Chatterjee et al. 2022) has been developed to analyse drift scan interferometric observations.

The observed visibility data typically have several frequency channels, which are flagged to avoid radio frequency interference (RFI) contamination. This introduces artefacts in the estimated PS, and several state-of-the-art algorithms (Parsons & Backer 2009; Trott 2016; Ewall-Wice et al. 2021; Kern & Liu 2021; Kennedy et al. 2023) have been proposed to overcome this. The MAPS-based TGE used for this work naturally overcomes this problem by first estimating $C_\ell(\Delta\nu)$ and using this to estimate $P(k_\perp, k_\parallel)$. This advantage arises because we obtain estimates of $C_\ell(\Delta\nu)$ for every $\Delta\nu$ even in the presence of flagged frequency channels. This approach was proposed in Bharadwaj et al. (2018) who validated it using simulations, and it was subsequently demonstrated on observed 150-MHz GMRT data (Pal et al. 2021).

The observations of the upgraded GMRT (uGMRT; Gupta et al. 2017) Band 3 data considered in this paper are described in Chakraborty et al. (2019b), who also outline the initial analysis. This paper is the third in a series of papers that have used MAPS-based TGE to estimate the 21-cm PS from these data. Earlier, Chakraborty et al. (2021, henceforward Ch21) have used the one-dimensional CLEAN algorithm (Parsons & Backer 2009) to estimate the 21-cm PS in delay space. In the first paper of the present series (Paper I), we have estimated the 21-cm PS from a sub-band of these data combining the two available polarizations (RR and LL). In Elahi et al. (2023, hereafter Paper II), we have estimated the 21-cm PS by cross-correlating the two polarizations instead of combining them. We find that this leads to a substantial reduction in the level of foreground contamination as compared to Paper I. Considering this estimated cross-correlation $C_\ell(\Delta\nu)$, in this work we introduce a method to model and remove the foregrounds. We show that this does away with the foreground wedge present in the earlier works, allowing us to use the entire $(k_\perp, k_\parallel)$ plane for estimating the 21-cm PS.

We closely follow the methodology of Ghosh et al. (2011b) who attempted a foreground removal from the $C_\ell(\Delta\nu)$ estimated from a 610-MHz GMRT observation. Considering the estimated $C_\ell(\Delta\nu)$, this is a combination of the 21-cm signal, foregrounds, and noise. For the relevant ℓ range, the 21-cm signal is presumed to be localized within $\Delta\nu \leq 0.5$ MHz, and at large $\Delta\nu$ (>0.5 MHz) the signal's amplitude drops substantially (Bharadwaj & Ali 2005). The key idea is that we may treat the large $\Delta\nu$ range as having very little (practically, zero) 21-cm signal, and use this to model the

[3]https://chime-experiment.ca/en/
[4]https://rac.ncra.tifr.res.in/ort.html
[5]https://hirax.ukzn.ac.za/
[6]https://www.skatelescope.org/

foregrounds. This model is extrapolated to small $\Delta\nu$ and used to subtract out the foregrounds. The residual $C_\ell(\Delta\nu)$ are expected to contain only the 21-cm signal and noise. It should be noted that we only use the small $\Delta\nu$ range ($\leq$0.5 MHz) to put a constraint on the 21-cm signal.

The rest of the paper is arranged as follows: Section 2 briefly summarizes the observations, whereas Section 3 describes the TGE and the foreground removal strategies. The formalism for the cylindrical PS is presented in Section 4, while the spherical PS and $[\Omega_{\rm H\,I} b_{\rm H\,I}]$ are presented in Sections 5 and 6, respectively. Our findings are summarized in Section 7.

The cosmological parameters $\Omega_{\rm m} = 0.309$, $n_{\rm s} = 0.965$, $h = 0.67$, and $\Omega_{\rm b}h^2 = 0.0224$, quoted in Planck Collaboration VI (2020), have been used all through the paper.

2 DATA DESCRIPTION

We have used the uGMRT Band 3 to carry out deep observations of the ELAIS-N1 field covering 1.8 deg^2 in a single pointing. The observations were spread over four nights during 2017 May, with a total observing time of 25 h including all calibration overheads. We have used an integration time of 2 s and a baseband bandwidth of 200 MHz (300–500 MHz), divided into 8196 frequency channels to achieve a spectral resolution ($\Delta\nu_{\rm c}$) of 24.4 kHz. Chakraborty et al. (2019b) detailed the observation and an initial data processing, which is summarized here. We have used the AOFLAGGER (Offringa et al. 2010; Offringa, van de Gronde & Roerdink 2012) for initial RFI flagging, followed by the RFLAG routine of the Common Astronomy Software Applications (CASA; McMullin et al. 2007) to flag residual low-level RFIs from calibrated data. Regarding the calibration, only direction-independent calibration is done on the data using CASA, and no polarization calibration is performed. After flagging and calibration of the visibility data, unresolved and compact sources brighter than 100 μJy are identified and subtracted out using the CASA routine UVSUB. The residual visibility data after UVSUB are used for all the subsequent analyses reported in this work.

We have split a 24.4-MHz bandwidth subset from the above data at the central frequency $\nu_{\rm c} = 432.84$ MHz, which has a visibility rms of 0.43 Jy (Paper I). We also note that the initial rounds of flagging and calibration lead to ∼55 per cent of these data being flagged (Paper I). We have analysed these data in Paper I and Paper II, and we extend our study utilizing the same data here. Similar to Paper II, for our analysis we have used only the baselines with $\boldsymbol{U} \leq 1\,k\lambda$, where the '*uv* coverage' is adequately dense and more or less uniform. Note that in Paper I we have considered a broader baseline range ($\boldsymbol{U} \leq 3\,k\lambda$).

We have used the notation $\mathcal{V}_i^x(\nu_{\rm a})$ to refer to the visibility data corresponding to baseline $\boldsymbol{U}_i$, in polarization state x (RR and LL circular polarization correlation products for the present data) and at an observing frequency of $\nu_{\rm a}$.

3 FOREGROUND REMOVAL FROM MAPS

3.1 The cross TGE

The MAPS $C_\ell(\nu_{\rm a}, \nu_{\rm b})$ is formally defined as

$$C_\ell(\nu_{\rm a}, \nu_{\rm b}) = \left\langle a_{\ell{\rm m}}(\nu_{\rm a})\, a^*_{\ell{\rm m}}(\nu_{\rm b}) \right\rangle, \tag{1}$$

where $a_{\ell{\rm m}}(\nu)$ are the coefficients when we expand $\delta T_{\rm b}(\hat{\boldsymbol{n}}, \nu)$ the brightness temperature fluctuations in terms of spherical harmonics $Y^{\rm m}_\ell(\hat{\boldsymbol{n}})$ on the sky. The ensemble average $\langle ... \rangle$ refers to different statistically independent realizations of the random field $\delta T_{\rm b}(\hat{\boldsymbol{n}}, \nu)$.

The 'cross' TGE for MAPS that cross-correlates the calibrated visibility data $\mathcal{V}_i^x(\nu_{\rm a})$ measured in two orthogonal polarizations (x = RR, LL) is described in Paper II, and we briefly summarize this here. We first calculate $\mathcal{V}^x_{\rm cg}(\nu_{\rm a})$ the convolved–gridded visibility for every grid point $\boldsymbol{U}_g$ on a rectangular grid in the *uv*-plane using

$$\mathcal{V}^x_{\rm cg}(\nu_{\rm a}) = \sum_i \tilde{w}(\boldsymbol{U}_g - \boldsymbol{U}_i)\, \mathcal{V}^x_i(\nu_{\rm a})\, F^x_i(\nu_{\rm a})\,, \tag{2}$$

where $F^x_i(\nu_{\rm a})$, which takes values 0 or 1, accounts for the flagged channels, and $\tilde{w}(\boldsymbol{U})$ is the Fourier transform of the window function $\mathcal{W}(\theta) = {\rm e}^{-\theta^2/[f\theta_0]^2}$, which is introduced to taper the sky response away from the beam centre. Here, θ_0 is 0.6 times the full width at half-maximum of the primary beam (PB), and following Paper I we have used $f = 0.6$ for the present analysis.

The cross TGE is defined as

$$\hat{E}_g(\nu_{\rm a},\, \nu_{\rm b}) = M_g^{-1}(\nu_{\rm a}, \nu_{\rm b}) \mathcal{R}e \Big[\mathcal{V}^{\rm RR}_{\rm cg}(\nu_{\rm a}) \mathcal{V}^{*\rm LL}_{\rm cg}(\nu_{\rm b}) + \mathcal{V}^{\rm LL}_{\rm cg}(\nu_{\rm a}) \mathcal{V}^{*\rm RR}_{\rm cg}(\nu_{\rm b}) \Big]\,, \tag{3}$$

where $M_g(\nu_{\rm a}, \nu_{\rm b})$ is a normalization factor, which we have estimated using simulations corresponding to $C_\ell(\nu_{\rm a}, \nu_{\rm b}) = 1$ referred to as unit MAPS or uMAPS. The simulated visibilities $[\mathcal{V}_i^x(\nu_{\rm a})]_{\rm uMAPS}$ are used to estimate

$$M_g(\nu_{\rm a}, \nu_{\rm b}) = \mathcal{R}e \Big[\mathcal{V}^{\rm RR}_{\rm cg}(\nu_{\rm a}) \mathcal{V}^{*\rm LL}_{\rm cg}(\nu_{\rm b}) + \mathcal{V}^{\rm LL}_{\rm cg}(\nu_{\rm a}) \mathcal{V}^{*\rm RR}_{\rm cg}(\nu_{\rm b}) \Big]_{\rm uMAPS}\,. \tag{4}$$

We have used multiple realizations (50 in this work) of the uMAPS to decrease the statistical uncertainties in the estimated M_g.

The estimator (equation 3) is unbiased, i.e. $\langle \hat{E}_g \rangle = C_{\boldsymbol{\ell}_g}$ at each grid $\boldsymbol{U}_g$, where $\boldsymbol{\ell}_g = 2\,\pi\,\boldsymbol{U}_g$. We subsequently use the notation $C_{\boldsymbol{\ell}}$ and $\boldsymbol{\ell}$ to denote $C_{\boldsymbol{\ell}_g}$ and $\boldsymbol{\ell}_g$, respectively. The cosmological 21-cm signal is statistically isotropic on the sky plane, i.e. $C_{\boldsymbol{\ell}} \equiv C_\ell$, where $\ell = \mid \boldsymbol{\ell} \mid$ corresponds to an angular multipole. It is then possible to do an annular binning of the estimated $\hat{E}_g$ in the *uv*-plane. We have taken this approach in Paper I and Paper II to increase the SNR, and also to reduce the data volume for a PS estimation. However, the isotropy does not hold for the foregrounds, and the foreground contribution differs from grid point to grid point. We find it advantageous to individually model and remove the foregrounds separately at each grid point. Hence, we avoid binning $C_{\boldsymbol{\ell}}$ at this stage, and we perform the binning only after foreground removal.

3.2 The predicted signal

Here, we follow Paper II and consider $C_\ell(\Delta\nu)$ instead of $C_\ell(\nu_{\rm a}, \nu_{\rm b})$. This assumes the statistics of the 21-cm signal to be ergodic along the line of sight (LoS), which is quite justified given the relatively small redshift interval of $\Delta z = 0.19$ centred around $z = 2.28$. We have used

$$\left[C_{\ell_{\rm a}}(\Delta\nu_n)\right]_T = [\Omega_{\rm H\,I} b_{\rm H\,I}]^2\, \frac{\bar{T}^2}{\pi r^2} \int_0^\infty {\rm d}k_\parallel \cos(k_\parallel r' \Delta\nu_n) \times {\rm sinc}^2\left(k_\parallel r' \Delta\nu_{\rm c}/2\right)\, P_m\left(k_{\perp{\rm a}}, k_\parallel\right) \tag{5}$$

to calculate the predicted $\left[C_{\ell_{\rm a}}(\Delta\nu_n)\right]_T$ corresponding to the 21-cm signal. Here, the equation (5) has been used in a slightly modified form in Paper II who took it from Bharadwaj & Ali (2005). In equation (5), $P_m(\boldsymbol{k}_\perp, k_\parallel)$ is the dark matter PS in redshift space at the wave vector $\boldsymbol{k}$ whose LoS and perpendicular components are $k_\parallel$ and $\boldsymbol{k}_\perp = \boldsymbol{\ell}/r$, respectively, and the sinc function takes care of the finite width $\Delta\nu_{\rm c}$ of the frequency channels. Considering $z = 2.28$, we find

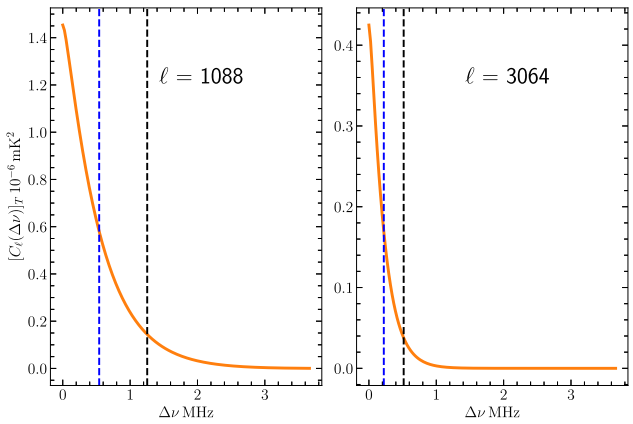

Figure 1. The theoretically predicted 21-cm signal corresponding to $[\Omega_{\rm H\,I} b_{\rm H\,I}] = 10^{-3}$ is shown. The blue and the black dashed vertical lines show $[\Delta\nu]_{0.4}$ and $[\Delta\nu]_{0.1}$, respectively, for the corresponding ℓ values.

that the comoving distance r and its frequency derivative $r^{'}$ have the values 5703 Mpc and 9.85 Mpc MHz^{-1}, respectively (Paper II). The mean brightness temperature (Bharadwaj et al. 2001; Bharadwaj & Ali 2005)

$$\bar{T}(z) = 133\ \mathrm{mK}\,(1+z)^2 \left(\frac{h}{0.7}\right)\left(\frac{H_0}{H(z)}\right) \tag{6}$$

is found to have a value of 406 mK at this redshift. $\Omega_{\rm H\,I}$ and $b_{\rm H\,I}$ here are the comoving H I mass density in units of the present critical density and the bias parameter, respectively. We have used $[\Omega_{\rm H\,I} b_{\rm H\,I}] = 10^{-3}$, which is supported by various observations (e.g. Rhee et al. 2018 and references therein) and simulations (e.g. Sarkar, Bharadwaj & Anathpindika 2016). For the matter PS $P_m(\boldsymbol{k})$, we have used the fitting formula presented in Eisenstein & Hu (1998) ignoring the effect of redshift space distortion (Bharadwaj & Ali 2005).

Fig. 1 shows $[C_\ell(\Delta\nu)]_T$ for two representative ℓ values. Considering both the panels, we see that $[C_\ell(\Delta\nu)]_T$ peaks at $\Delta\nu = 0$, declines with increasing $\Delta\nu$, and $[C_\ell(\Delta\nu)]_T \sim 0$ when $\Delta\nu \gtrsim 2$ MHz. We define $[\Delta\nu]_{0.4}$ and $[\Delta\nu]_{0.1}$, respectively, as the values of $\Delta\nu$ where the amplitude of $[C_\ell(\Delta\nu)]_T$ falls to 40 and 10 per cent of its peak value $[C_\ell(0)]_T$. In each panel, the blue and the black dashed vertical lines indicate $[\Delta\nu]_{0.4}$ and $[\Delta\nu]_{0.1}$, respectively, for the corresponding ℓ value. We see that the 21-cm signal is predominantly localized within a small range of frequency separations, and there is a very little 21-cm signal at $\Delta\nu > [\Delta\nu]_{0.1}$. The quantities $[\Delta\nu]_{0.4}$ and $[\Delta\nu]_{0.1}$ provide estimates of this range of frequency separations within which the 21-cm signal is localized.

The left-hand panel of Fig. 1 shows $[C_\ell(\Delta\nu)]_T$ for $\ell = 1088$. This has a peak value of $\sim 1.4 \times 10^{-6}$ mK^2 at $\Delta\nu = 0$, and $[\Delta\nu]_{0.4} \approx 0.54$ MHz and $[\Delta\nu]_{0.1} \approx 1.25$ MHz, respectively. The right-hand panel shows $[C_\ell(\Delta\nu)]_T$ for a comparatively larger $\ell = 3064$ value. For this ℓ, the peak value is $\sim 0.4 \times 10^{-6}$ mK^2, and we have $[\Delta\nu]_{0.4} \approx 0.22$ MHz and $[\Delta\nu]_{0.1} \approx 0.52$ MHz, respectively. We see that with increasing ℓ, the amplitude of $[C_\ell(\Delta\nu)]_T$ goes down, and it also decorrelates faster, i.e. $[\Delta\nu]_{0.4}$ and $[\Delta\nu]_{0.1}$ get smaller.

3.3 Foreground modelling and removal: polynomial fitting

The foregrounds, which are believed to originate from continuum sources, are expected to be spectrally smooth (Bharadwaj et al. 2001; Di Matteo et al. 2002). However, baseline migration (Morales et al. 2012; Vedantham et al. 2012) and the frequency dependence of the PB (Ghosh et al. 2011a) introduce additional chromaticity in the measured $C_\ell(\Delta\nu)$. Regardless, the foregrounds are expected to be present over the entire $\Delta\nu$ range, whereas the 21-cm signal is largely localized within a small $\Delta\nu$ range, and there is very little 21-cm signal for $\Delta\nu > [\Delta\nu]_{0.1}$. Here, we assume that in the range $\Delta\nu > [\Delta\nu]_{0.1}$ the measured $C_\ell(\Delta\nu)$ can be modelled as

$$C_\ell(\Delta\nu) = [C_\ell(\Delta\nu)]_{\rm FG} + [\mathrm{Noise}], \tag{7}$$

where $[C_\ell(\Delta\nu)]_{\rm FG}$ and [Noise] are the foreground and noise contributions, respectively. In our first approach, we have used polynomial fitting (PF) to model the foregrounds and subtract these out. The estimated $C_\ell(\Delta\nu)$ is, by construction, a symmetric function of $\Delta\nu$, i.e. $C_\ell(\Delta\nu) = C_\ell(-\Delta\nu)$. We have chosen an even polynomial in $\Delta\nu$ to model the foregrounds,

$$[C_\ell(\Delta\nu)]_{\rm FG} = \sum_{m=0}^{n} a_{2m}\ (\Delta\nu)^{2m}, \tag{8}$$

where the coefficients a_{2m} are the free parameters of our foreground model. We have fitted equation (8) to the measured $C_\ell(\Delta\nu)$ in the range $\Delta\nu > [\Delta\nu]_{0.1}$, and used this to obtain the best-fitting values of the parameters a_{2m}. The idea is to extrapolate the best-fitting $[C_\ell(\Delta\nu)]_{\rm FG}$ to $\Delta\nu \leq [\Delta\nu]_{0.1}$, where we use it to subtract out the foregrounds

$$[C_\ell(\Delta\nu)]_{\rm res} = C_\ell(\Delta\nu) - [C_\ell(\Delta\nu)]_{\rm FG}\,. \tag{9}$$

Subsequent to foreground subtraction, the entire analysis is restricted to the range $\Delta\nu \leq [\Delta\nu]_{0.1}$, which was not included in PF. We have used $[C_\ell(\Delta\nu)]_{\rm res}$ to measure (or constrain) the H I 21-cm signal.

We have used maximum likelihood to find the best-fitting parameters a_{2m} and their error covariance C_a. The error covariance C_a also allows us to quantify the uncertainty in the best-fitting $[C_\ell(\Delta\nu)]_{\rm FG}$. We assume that the polynomial coefficients a_{2m} follow a multivariate normal distribution (MND) with mean values as obtained from the maximum likelihood solutions and covariance C_a, i.e. $a_{2m} \sim \mathcal{N}(a_{2m}^{\rm MLE}, C_a)$. We generate 1000 realizations of the polynomial coefficients from the MND and obtain the fits for each realization. The mean of the fits is by construction the best-fitting foreground model $[C_\ell(\Delta\nu)]_{\rm FG}$, whereas the standard deviation across the realizations yields the error in the fits. These fitting errors are added to the error budget for each data point of the residuals $[C_\ell(\Delta\nu)]_{\rm res}$.

The parameter n, which decides the order of the fitting polynomial, is an extra free parameter in our foreground model. We have considered different values $0 \leq n \leq 10$, and chosen the one that best matches the measured $C_\ell(\Delta\nu)$ in the range $\Delta\nu \leq [\Delta\nu]_{0.1}$ and hence gives the smallest residual $C_\ell(\Delta\nu)$. We have quantified this using the mean squared error (MSE)

$$\mathrm{MSE} = \frac{\sum_i^k [C_\ell(\Delta\nu_i)]_{\rm res}^2}{(k-n)}, \tag{10}$$

where k is the number of $\Delta\nu$ data points in the range $\Delta\nu \leq [\Delta\nu]_{0.1}$. We have used the value of n that minimizes MSE.

The fitting procedure is illustrated in Fig. 2, where the left-hand panel shows the measured $C_\ell(\Delta\nu)$ for a representative $\boldsymbol{\ell}$ with $\ell = 2740$. Note that we have estimated $C_\ell(\Delta\nu)$ for the entire $\Delta\nu$ range of 24.4 MHz, which corresponds to the bandwidth of the data. The results for the entire $\Delta\nu$ range are presented in fig. 2 of Paper II. We find there that various $\Delta\nu$ dependent structures appear in $C_\ell(\Delta\nu)$ at larger $\Delta\nu$ separations (>4 MHz). These structures are possibly due to the low sampling at larger $\Delta\nu$, and also the frequency dependence of the PB pattern. The difficulty arises because PF starts to model

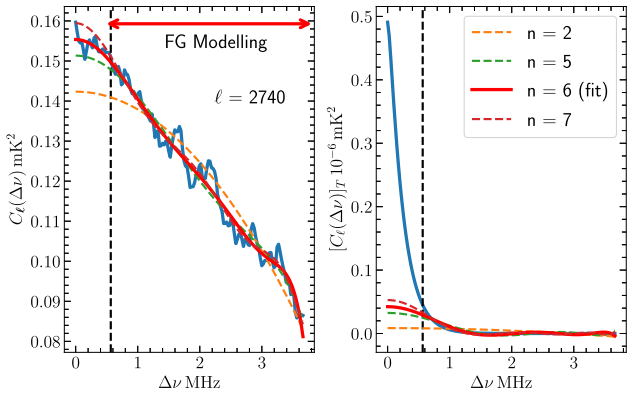

Figure 2. The left- and the right-hand panels, respectively, show (blue solid line) the measured $C_\ell(\Delta\nu)$ and $[C_\ell(\Delta\nu)]_T$ for a representative $\ell = 2740$. The black dashed vertical lines show $[\Delta\nu]_{0.1}$. In the left-hand panel, the dashed curves show the best-fitting $[C_\ell(\Delta\nu)]_{\rm FG}$ corresponding to different values of n indicated in the legend. The red solid curve shows the $[C_\ell(\Delta\nu)]_{\rm FG}$ that minimizes the residual $[C_\ell(\Delta\nu)]_{\rm res}$ in the range $\Delta\nu < [\Delta\nu]_{0.1}$. The dashed curves and the red solid curve in the right-hand panel show $[C_\ell(\Delta\nu)]_{\rm TP}$, the best-fitting polynomials for $[C_\ell(\Delta\nu)]_T$, for the same values of n.

these structures instead of tracing the smooth $\Delta\nu$ dependence of the foregrounds. In this work, we have restricted the $\Delta\nu$ range to 3.66 MHz, which corresponds to 150 $\Delta\nu$ data points. We see that $C_\ell(\Delta\nu)$ falls smoothly across the $\Delta\nu$ range considered here, albeit with relatively small wriggles superimposed on this. The right-hand panel shows $[C_\ell(\Delta\nu)]_T$ predicted for the same value of ℓ, for which the vertical dashed line shows $[\Delta\nu]_{0.1} = 0.56$ MHz in both the panels. Our aim is to use the range $\Delta\nu > [\Delta\nu]_{0.1}$ to model the overall smooth nature of the measured $C_\ell(\Delta\nu)$. The left-hand panel shows the best-fitting polynomials for several values of n. Considering $n = 2$, we see that a quadratic polynomial does not provide a very good fit. The fit improves as n is increased; however, the polynomial starts to model features from the wriggles and other rapid fluctuations if n is increased beyond a certain point. Considering $n = 5$, 6, and 7, we see that these are nearly indistinguishable in the range $\Delta\nu > [\Delta\nu]_{0.1}$; however, we see that $n = 6$ provides the best fit to the measured $C_\ell(\Delta\nu)$ at $\Delta\nu \leq [\Delta\nu]_{0.1}$. The value of MSE is also minimized for $n = 6$, and we have adopted this to set the order of the polynomial for this particular value of ℓ. We have followed the same procedure for all the ℓ values in our analysis, the value of n was restricted to the range $0 \leq n \leq 10$.

PF is also expected to cause some loss of the 21-cm signal. To quantify this loss, we consider the expected signal $[C_\ell(\Delta\nu)]_T$, which is shown in the right-hand panel. Here also, we have used the range $\Delta\nu > [\Delta\nu]_{0.1}$ to obtain $[C_\ell(\Delta\nu)]_{\rm TP}$, which is the best-fitting polynomial for $[C_\ell(\Delta\nu)]_T$. The best-fitting polynomials are shown for several values of n, we see that the signal loss increases as the order of the polynomial increases. Here, we adopt $n = 6$, which is the value used for the measured $C_\ell(\Delta\nu)$. We see that $[C_\ell(\Delta\nu)]_{\rm TP}$ has a peak value of $\sim 0.04 \times 10^{-6}$ mK^2, which is less than 10 per cent of the peak value of $[C_\ell(\Delta\nu)]_T$, i.e. the signal loss is less than 10 per cent. In the subsequent analysis, we have accounted for the signal loss using the loss-corrected 21-cm signal

$$[C_\ell(\Delta\nu)]_{\rm TC} = [C_\ell(\Delta\nu)]_T - [C_\ell(\Delta\nu)]_{\rm TP}\,. \tag{11}$$

It may be noted that the actual signal loss is expected to be less than the conservative estimates used here. This is because we will generally have a combination of foregrounds and 21-cm signal, and the foregrounds are expected to have a slower $\Delta\nu$ variation in comparison to the 21-cm signal.

3.4 Foreground modelling and removal: Gaussian process regression

The second approach conforms to the same idea (presented in Section 3.3) that the foregrounds are expected to span the entire $\Delta\nu$ range in contrast to the 21-cm signal, which largely remains localized within a small $\Delta\nu$ range. However, instead of using a polynomial (equation 8), we model the foregrounds with a Gaussian process (GP).

A GP is a collection of an infinite number of random variables $f(x)$ such that any finite number of these variables follow a joint MND. A mean $m(x)$ and a covariance function $K(x, x')$ completely define a GP,

$$f(x) \sim \mathcal{GP}(m(x), K(x, x'))\,. \tag{12}$$

In GP regression (GPR), the observed data $\boldsymbol{d}$ are assumed to be the outcome of a GP whose mean and covariance functions are unknown. The mean and the covariance functions have forms involving hyperparameters whose optimal values are inferred from the data. Once the optimal values of the hyperparameters are known, one can make predictions for unobserved data points $\boldsymbol{d}^*$ (Williams & Rasmussen 1996; Rasmussen & Williams 2006). The PYTHON library GEORGE (Ambikasaran et al. 2015) has been used for the GPR analysis presented here.

In our case, the measured $C_\ell(\Delta\nu)$ values in the range $\Delta\nu > [\Delta\nu]_{0.1}$ {or $[\Delta\nu]_{0.4}$} are the observed data points $\boldsymbol{d}$. We model these using equation (7), where $[C_\ell(\Delta\nu)]_{\rm FG}$ and [Noise] are both assumed to be independent, zero mean, Gaussian random variables. We further assume the measured $C_\ell(\Delta\nu)$ to be the outcome of a GP with covariance

$$\boldsymbol{K}(\boldsymbol{d}, \boldsymbol{d}) \equiv K(\Delta\nu_m, \Delta\nu_n) = k_{\rm FG}(\Delta\nu_m, \Delta\nu_n) + \sigma_{\rm N}^2\delta_{nm}, \tag{13}$$

where $\sigma_{\rm N}^2$, the noise variance, is estimated from noise-only simulations and $\boldsymbol{k}_{\rm FG}(\boldsymbol{d}, \boldsymbol{d}) \equiv k_{\rm FG}(\Delta\nu_m, \Delta\nu_n)$ is a positive semi-definite kernel that quantifies the covariance of $[C_\ell(\Delta\nu_m)]_{\rm FG}$ and $[C_\ell(\Delta\nu_n)]_{\rm FG}$.

The choice of the functional form of the kernel $k_{\rm FG}(\Delta\nu_m, \Delta\nu_n)$ is crucial in GPR predictions as it encodes our assumptions (or prior knowledge) about $[C_\ell(\Delta\nu)]_{\rm FG}$. There are several commonly used kernels such as squared exponential, rational quadratic, and Matérn that are stationary, i.e. the kernel depends only on the difference $R = |\Delta\nu_m - \Delta\nu_n|$. However, in our case $\Delta\nu$ is already a frequency difference. We further find that these kernels are more sensitive to the rapid fluctuations relative to the smooth, slowly varying features in the observed data $\boldsymbol{d}$.

For the present analysis, we have used the non-stationary polynomial kernel given by (Rasmussen & Williams 2006)

$$k_{\rm FG}(\Delta\nu_m, \Delta\nu_n) = (\Delta\nu_m\, \Delta\nu_n + b)^P, \tag{14}$$

where the constant b is a hyperparameter and P, which denotes the order of the polynomial kernel, is not a hyperparameter and is held fixed. We have considered different values of P and find that larger values provide a better fit, however, at the expense of a larger signal loss. With this in view, we have restricted ourselves to $P = 2$ and 3 for the entire analysis. For a given P, we have used maximized the log-likelihood to estimate the optimal value of b from the observed data $\boldsymbol{d}$.

In the final step, we use GPR, with the inferred optimal hyperparameter b, to make predictions for $\boldsymbol{d}^*$ the unobserved data points. In our case, $\boldsymbol{d}^*$ corresponds to the foreground model predictions

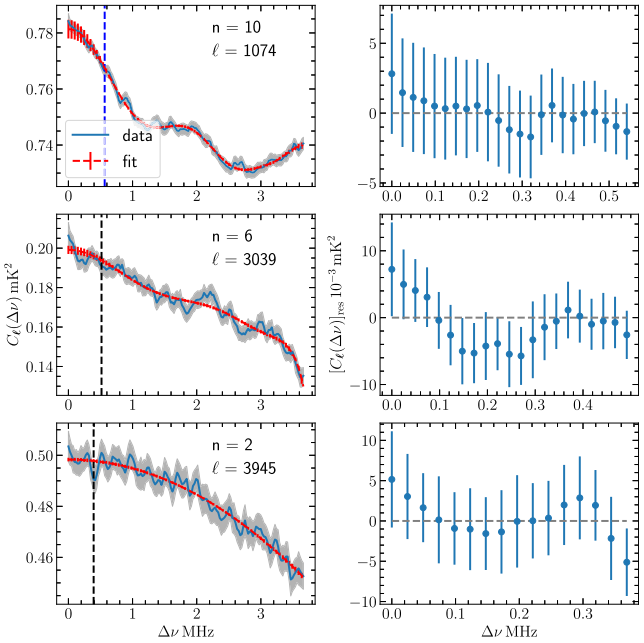

Figure 3. The blue solid lines in the left-hand panel show the measured $C_\ell(\Delta\nu)$ with 2σ uncertainties (grey shaded region) expected from noise. The red dashed lines and the associated error bars show the best-fitting polynomial foreground models and their 2σ uncertainties. The vertical dashed lines show $[\Delta\nu]_{0.4}$ (blue) in the top panel and $[\Delta\nu]_{0.1}$ (black) in the two bottom panels, respectively. The values of ℓ and n (used for foreground modelling) are mentioned in the respective panels. The residuals $[C_\ell(\Delta\nu)]_{\rm res}$ in the range $\Delta\nu \le [\Delta\nu]$ are shown in the right-hand panels where the 2σ error bars combine the noise and fitting errors.

$[C_\ell(\Delta\nu)]_{\rm FG}$ for the range $\Delta\nu \le [\Delta\nu]_{0.1}$ {or $[\Delta\nu]_{0.4}$}. The mean and covariance of $\boldsymbol{d}^*$ (given the observed data $\boldsymbol{d}$) are, respectively, predicted to be (see e.g. Bishop 2006)

$$\mu(\boldsymbol{d}^* \mid \boldsymbol{d}) = \boldsymbol{k}_{\rm FG}(\boldsymbol{d}^*, \boldsymbol{d})\, \boldsymbol{K}^{-1}(\boldsymbol{d}, \boldsymbol{d})\, \boldsymbol{d}, \tag{15}$$

$$\Sigma(\boldsymbol{d}^* \mid \boldsymbol{d}) = \boldsymbol{k}_{\rm FG}(\boldsymbol{d}^*, \boldsymbol{d}^*) - \boldsymbol{k}_{\rm FG}(\boldsymbol{d}^*, \boldsymbol{d})\, \boldsymbol{K}^{-1}(\boldsymbol{d}, \boldsymbol{d})\, \boldsymbol{k}_{\rm FG}(\boldsymbol{d}, \boldsymbol{d}^*). \tag{16}$$

Here, $\mu(\boldsymbol{d}^*|\boldsymbol{d})$ is the predicted foreground model $[C_\ell(\Delta\nu)]_{\rm FG}$ and $\Sigma(\boldsymbol{d}^*|\boldsymbol{d})$ is the covariance of the fitting error. We have added the diagonal elements of $\Sigma(\boldsymbol{d}^*|\boldsymbol{d})$ to the noise variance to estimate the total error variance for $[C_\ell(\Delta\nu)]_{\rm res}$. To account for the signal loss, we have applied GPR to the expected 21-cm signal $[C_\ell(\Delta\nu)]_T$ and corrected for the 21-cm signal loss in exactly the same way as in Section 3.3.

3.5 Fit and residual $C_\ell(\Delta\nu)$

We have divided the measured $C_\ell(\Delta\nu)$ into three sets, $\ell < 2000$ (Set I), $2000 < \ell < 4000$ (Set II), and $\ell > 4000$ (Set III), and analysed them separately. For Set I, which covers the smaller ℓ values, we see that $[C_\ell(\Delta\nu)]_T$ decorrelates slowly with increasing $\Delta\nu$ (left-hand panel of Fig. 1), and $[\Delta\nu]_{0.1} \sim 1.25$ MHz. We have found that for this case the measured $C_\ell(\Delta\nu)$ contains structures within $[\Delta\nu]_{0.1}$ that cannot be modelled by extrapolating the polynomial from $\Delta\nu > [\Delta\nu]_{0.1}$. In this case, we find that it is advantageous to reduce the $\Delta\nu$ range for signal estimation and increase the $\Delta\nu$ range used for PF. Here, we have used $[\Delta\nu]_{0.4}$ for the ℓ values of Set I, and $[\Delta\nu]_{0.1}$ for the ℓ values in the other two sets.

The left-hand panels of Fig. 3 show the measured $C_\ell(\Delta\nu)$ (blue solid lines) along with the best-fitting polynomial foreground models $[C_\ell(\Delta\nu)]_{\rm FG}$ (red dashed lines), considering three representative ℓ values one from each set, respectively. The grey shaded regions on the measured $C_\ell(\Delta\nu)$ show errors due to noise, which were estimated using simulations as presented in Paper II (also in Section 4 of this paper), whereas the red error bars on $[C_\ell(\Delta\nu)]_{\rm FG}$ show fitting errors. All the panels here show 2σ error bars.

The right-hand panels of Fig. 3 show $[C_\ell(\Delta\nu)]_{\rm res}$, the residuals after foreground subtraction, and the blue error bars are the combined errors from noise and foreground modelling. The $\Delta\nu$ range is restricted to $\Delta\nu \le [\Delta\nu]_{0.4}$ and $[\Delta\nu]_{0.1}$ for the top panel and the two lower panels, respectively. We see that in all the panels shown here the residuals $[C_\ell(\Delta\nu)]_{\rm res}$ are largely consistent with zero (grey dashed lines). While the foreground subtraction is successful at several ℓ values, there also are many ℓ values where the residuals are clearly not consistent with zero. In some cases, our method overpredicts $[C_\ell(\Delta\nu)]_{\rm FG}$ and we are left with large negative $[C_\ell(\Delta\nu)]_{\rm res}$, whereas we also find some other cases where $[C_\ell(\Delta\nu)]_{\rm FG}$ is underpredicted and we have large positive $[C_\ell(\Delta\nu)]_{\rm res}$. It is necessary to identify and flag the ℓ where our foreground subtraction fails and $[C_\ell(\Delta\nu)]_{\rm res}$ is clearly not consistent with zero.

We have assessed the effectiveness of foreground subtraction by fitting the residual $[C_\ell(\Delta\nu)]_{\rm res}$ using our model prediction for the expected signal $[C_\ell(\Delta\nu)]_T$ (equation 5) with $\boldsymbol{A} \equiv [\Omega_{\rm H\,I} b_{\rm H\,I}]^2$ as a free parameter. The quantity $\mathbf{A}$ effectively quantifies the amplitude of $[C_\ell(\Delta\nu)]_{\rm res}$, with $\mathbf{dA}$ denoting the predicted uncertainties. We have used the criteria $|\mathbf{A}|/\mathbf{dA} > 2$ to identify the ℓ values where foreground subtraction fails, which are then flagged. We finally have 8, 33, and 20 unflagged ℓ values for Sets I, II, and III, respectively. The first six figures of Appendix A show the measured $C_\ell(\Delta\nu)$, the corresponding foreground model $[C_\ell(\Delta\nu)]_{\rm FG}$, and the residual $[C_\ell(\Delta\nu)]_{\rm res}$ for all the ℓ values that were accepted for the subsequent analysis. A visual inspection shows these $[C_\ell(\Delta\nu)]_{\rm res}$ values to be largely within the 2σ error bars. We have also considered the flagging criteria ($|\mathbf{A}|/\mathbf{dA} > 1$) and ($|\mathbf{A}|/\mathbf{dA} > 3$ and 5) to investigate what happens if the cut is tightened or relaxed. We find that a few more ℓ values are flagged if the cut is tightened, whereas we pick up some ℓ values having residual foregrounds if the cut is relaxed. Either way, the final conclusions are not much changed, and we have used $|\mathbf{A}|/\mathbf{dA} > 2$, which provides the best results.

Fig. 4 is the same as Fig. 3, except that we have used GPR instead of PF for foreground modelling. The flagging here has also been carried out in the same way as for PF. For GPR, visual inspection reveals a further four ℓ grid points in Set III where $[C_\ell(\Delta\nu)]_{\rm res}$ has strong oscillatory features, which can be attributed to residual point source foregrounds. In addition to the flagging criteria described earlier, we have also flagged these four visually identified ℓ grids. We finally get 2, 21, and 13 unflagged ℓ values in GPR analysis for Sets I, II, and III, respectively. The last two figures of Appendix A show the measured $C_\ell(\Delta\nu)$, the corresponding foreground model $[C_\ell(\Delta\nu)]_{\rm FG}$, and the residual $[C_\ell(\Delta\nu)]_{\rm res}$ for all the ℓ values that were accepted for the subsequent analysis. We visually assess that these $[C_\ell(\Delta\nu)]_{\rm res}$ values are largely within the 2σ error bars.

We note a few features common to both PF and GPR analysis presented here. Considering the foreground fits, both the methods show quite reasonable fits to the measured $C_\ell(\Delta\nu)$, and the residuals $[C_\ell(\Delta\nu)]_{\rm res}$ are largely within 2σ error bounds. Strictly speaking, we could interpret the residuals as being consistent with the noise predictions if the $[C_\ell(\Delta\nu)]_{\rm res}$ values were randomly distributed around zero. However, we find that the residuals are not random but exhibit correlations that extend over several adjacent $\Delta\nu$ values. Here, we note that this can arise from uncertainties in foreground modelling, which are, in general, predicted to be correlated (e.g.

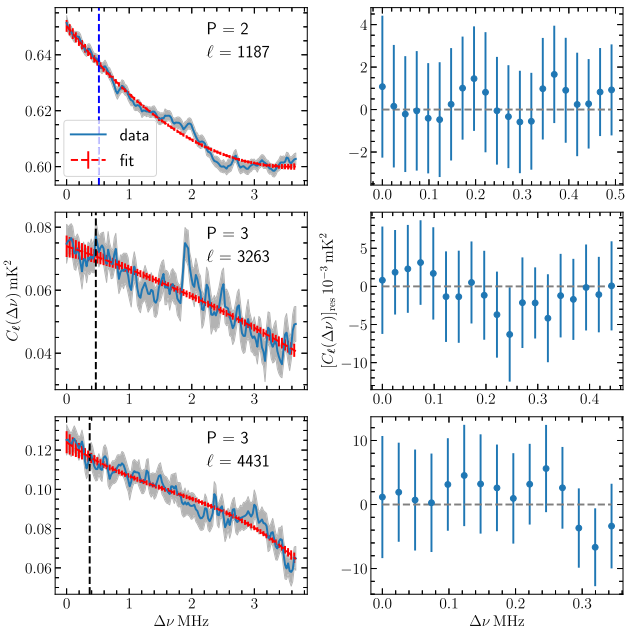

Figure 4. Same as Fig. 3 but for GPR. The quoted values of P denote the orders of the polynomial covariance function.

equation 16). Low-level residual foregrounds can also lead to such correlations. In order to keep the treatment simple, for the present analysis we have ignored these correlations and we have treated the errors at each $[C_{\boldsymbol{\ell}}(\Delta\nu)]_{\rm res}$ as being independent.

4 THE CYLINDRICAL PS

The cylindrical PS $P(k_\perp, k_\parallel)$ of the 21-cm signal is the Fourier transform of the MAPS $C_\ell(\Delta\nu)$ along $\Delta\nu$ (Datta et al. 2007):

$$P(k_\perp, k_\parallel) = r^2\, r' \int_{-\infty}^{\infty} {\rm d}(\Delta\nu)\, {\rm e}^{-{\rm i}k_\parallel r'\Delta\nu}\, C_\ell(\Delta\nu)\,. \tag{17}$$

Following Paper II, we model the measured $C_\ell(\Delta\nu)$ as

$$C_\ell(\Delta\nu_n) = \sum_m \mathbf{A}_{nm}\, P(k_\perp, k_{\parallel m}) + [{\rm Noise}]_n, \tag{18}$$

where $\mathbf{A}$ contains the Fourier transform coefficients and $[{\rm Noise}]_n$ is the noise in each estimated $C_\ell(\Delta\nu_n)$. The maximum likelihood estimate of $P(k_\perp, k_{\parallel m})$ is given by

$$P(k_\perp, k_{\parallel m}) = \sum_n \left[\left(\mathbf{A}^\dagger \mathbf{N}^{-1}\mathbf{A}\right)^{-1}\mathbf{A}^\dagger\mathbf{N}^{-1}\right]_{mn} C_\ell(\Delta\nu_n), \tag{19}$$

where $\mathbf{N}$ is the noise covariance matrix.

The noise covariance estimate is detailed in Paper II, which we briefly restate here. We have simulated multiple (50) realizations of visibility data having zero mean and standard deviation $\sigma_{\rm N} = 0.43$ Jy, which is the visibility rms present in the actual data. We identically analyse the simulated data using the TGE (equation 3) to estimate $\mathbf{N}$. However, the true noise level in the data is ∼4.77 times larger than what we obtain from the system 'noise-only' simulations (Paper II), and the noise levels are scaled up with this factor for this work.

We have binned the foreground-subtracted residual $[C_{\boldsymbol{\ell}}(\Delta\nu)]_{\rm res}$ at different grid points $\boldsymbol{\ell}_g$ into equally separated annular bins in the uv-plane. We have calculated the bin-averaged values $C_{\ell_a}(\Delta\nu_n) = \sum_g w_g [C_{\boldsymbol{\ell}}(\Delta\nu)]_{\rm res}/\sum_g w_g$ at the bin-averaged multipoles $\ell_a = \sum_g w_g \ell_g/\sum_g w_g$, where the weights w_g of the grid points are taken to be unity implying that all grids are equally weighted. The ℓ_a values here span the range $1000 \lesssim \ell_a \lesssim 5800$ with 5 ℓ bins; this corresponds to the $k_\perp$ range 0.18 Mpc$^{-1} \lesssim {\rm k}_\perp \lesssim 1.0$ Mpc^{-1} for the estimated $P(k_\perp, k_\parallel)$. The $\Delta\nu$ extent available for 21-cm signal estimation is different for each $\boldsymbol{\ell}$. However, for estimating the cylindrical PS we have used a fixed range of 0.488 MHz, which corresponds to $N_E = 20$ frequency separations $\Delta\nu$ for all the bins. As a consequence, the estimated $P(k_\perp, k_\parallel)$ all span the same $k_\parallel$ range of 0 Mpc$^{-1} \leq {\rm k}_\parallel \leq 13.1$ Mpc^{-1} with the resolution $\Delta k_\parallel = 0.69$ Mpc^{-1}. The actual available $\Delta\nu$ range is larger than 0.488 MHz for some of the small $\boldsymbol{\ell}$ values for which this assumption leads to some additional 21-cm signal loss. The actual available range is somewhat smaller than 0.488 MHz for the large $\boldsymbol{\ell}$, in which case we end up introducing excess noise into the analysis. Note that the exact available $\Delta\nu$ ranges are shown for all the $\boldsymbol{\ell}$ values in Appendix A. We further note that the cylindrical PS has not been used for any of the final quantitative outcomes quoted in this paper. Here, the cylindrical PS has primarily been used for a visual representation of the residual data and for exploring the noise statistics. It may also be noted that we have not corrected $P(k_\perp, k_\parallel)$ for the signal loss associated with the foreground removal. We do not expect these simplifying assumptions to have a severe impact on the two issues under consideration in this section.

Fig. 5 shows $|P(k_\perp, k_\parallel)|$ before and after foreground removal using PF in the left-hand and the middle panels, respectively. The cylindrical PS are visually very similar for GPR analysis, and we do not explicitly show them here. The rightmost panel shows $\delta P_{\rm N}(k_\perp, k_\parallel)$, the expected rms statistical fluctuation due to noise. We note that $\delta P_{\rm N}(k_\perp, k_\parallel)$ is obtained by applying the maximum likelihood estimator (MLE; equation 19) on the $C_\ell(\Delta\nu)$ obtained from multiple realizations of the noise-only simulations mentioned earlier. The black dashed lines show the predicted location of the foreground wedge boundary $[k_\parallel]_H = (r/r'\nu_{\rm c})k_\perp$. The left-hand panel shows that before foreground subtraction $|P(k_\perp, k_\parallel)|$ has a dynamic range of $\sim 10^6$ starting from $\sim 10^8$ mK^2 Mpc^3 at $k_\parallel = 0$ to $\sim 10^2$ mK^2 Mpc^3 at the higher $k_\parallel$. The most affected LoS mode is $k_\parallel = 0$, which corresponds to the DC of the signal. Note that we have shifted the $k_\parallel = 0$ point slightly to place it in a log scale. We notice that the major contribution of the power is within $[k_\parallel]_H$. A significant foreground leakage is also noticeable up to $k_\parallel \sim 1$ Mpc^{-1}. The $|P(k_\perp, k_\parallel)|$ values beyond $k_\parallel \sim 1$ Mpc^{-1} appear to be comparatively foreground-free.

Considering the middle panel, we see that the $|P(k_\perp, k_\parallel)|$ values are found to lie within 10^2–10^6 mK^2 Mpc^3. The $|P(k_\perp, k_\parallel)|$ values at $k_\parallel < 1$ Mpc^{-1} have an amplitude $\sim 10^5$ mK^2 Mpc^3. We see that the smooth foreground components that appear at low $k_\parallel$ have been successfully subtracted out from the data. In the last two $k_\perp$ bins, we see some residual foregrounds at large $k_\parallel$ ($\gtrsim$2 Mpc^{-1}). In these larger $k_\perp$ modes, the noise is found to be high (right-hand panel), and we could not separate a rapidly varying foreground component (or possibly unknown systematics) from the noisy data.

After foreground subtraction, we can use the entire $(k_\perp, k_\parallel)$ space for constraining the 21-cm PS. However, it is necessary to ensure that the measured $P(k_\perp, k_\parallel)$ values are either strictly positive or be of either signs with the negative values being consistent with the expected noise level. Due to the non-uniform baseline sampling in the uv-plane and also in $\Delta\nu$, we do not expect the noise level to be uniform across the $(k_\perp, k_\parallel)$ plane. We account for this by considering the quantity X, which is defined as (Pal et al. 2021)

$$X = \frac{P(k_\perp, k_\parallel)}{\delta P_{\rm N}(k_\perp, k_\parallel)}. \tag{20}$$

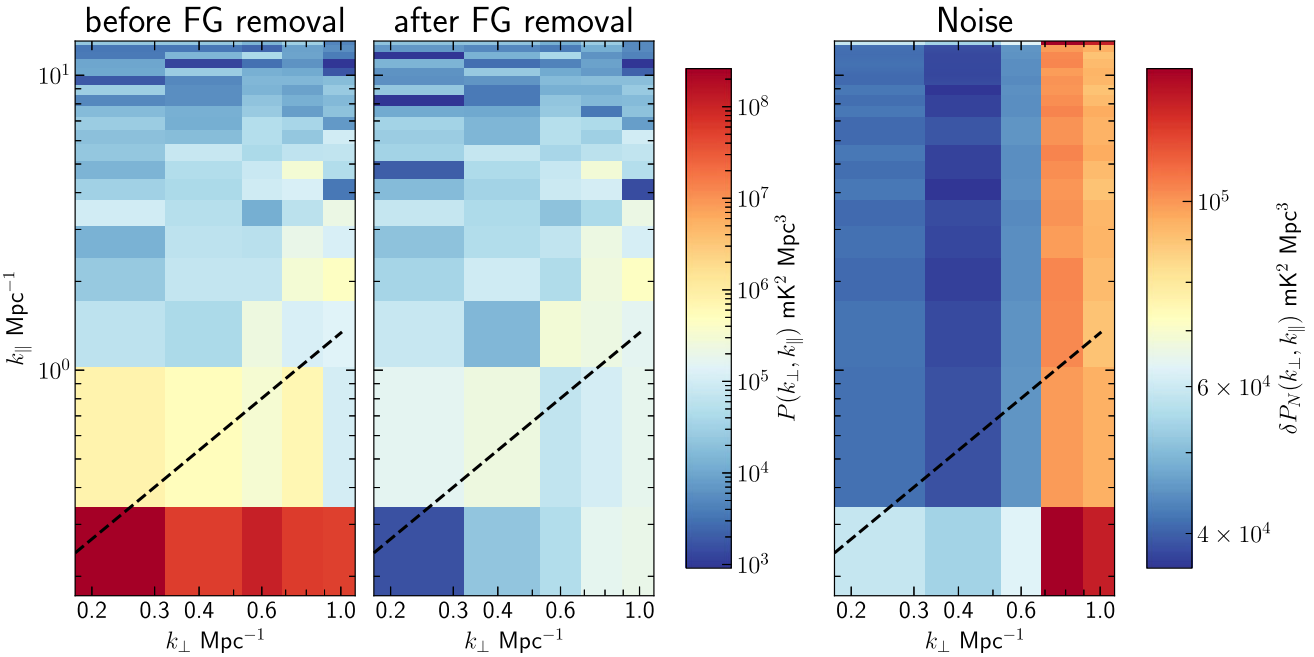

Figure 5. The first two panels show the cylindrical PS $|P(k_\perp, k_\parallel)|$ before and after foreground removal (using PF). The third panel shows the 1σ statistical fluctuation $\delta P_N(k_\perp, k_\parallel)$ in the estimated $P(k_\perp, k_\parallel)$. The black dashed line shows the predicted boundary of the foreground wedge $[k_\parallel]_H$.

The distribution of X is expected to be symmetric with mean $\mu = 0$ and standard deviation $\sigma_{\rm Est} = 1$ if the estimated $P(k_\perp, k_\parallel)$ are completely due to uncorrelated Gaussian random noise in the measured visibilities. Fig. 5 of Paper II shows the histogram of X prior to foreground subtraction considering only the $(k_\perp, k_\parallel)$ modes within the '21-cm window' (TW). This was found to be largely symmetric around $\mu = 0.61$ but with $\sigma_{\rm Est} = 4.77$. The small, positive mean ($\mu > 0$) was interpreted as indicating the presence of some low-level foreground leakage into the TW, whereas the relatively large $\sigma_{\rm Est}$ indicates that the actual $\delta P_N(k_\perp, k_\parallel)$ are larger than those expected from system noise alone. As noted in Paper II, possible sources for this excess noise include artefacts due to imperfect calibration, low-level residual RFIs, inaccuracy in point source removal, and various other factors that are not well known at present (Mertens et al. 2020; Gan et al. 2022). As mentioned earlier, we have scaled up the predictions from the system noise-only simulations by the factor $\sigma_{\rm Est} = 4.77$ to account for this excess noise. It is important to note that this scaling factor plays a crucial role in interpreting the statistical significance of the results presented here and also in Paper II as most of the PS measurements would be well above (or below) the expected noise level if this scaling were not applied.

Fig. 6 of this paper shows the histogram of X after foreground subtraction using PF (left) and GPR (right) considering all the available $(k_\perp, k_\parallel)$ modes. Note that in addition to the scaled system noise, $\delta P_N(k_\perp, k_\parallel)$ now also has a contribution from the uncertainties in foreground modelling. We find $\mu = 0.21$ and $\sigma_{\rm Est} = 1.1$ for PF and $\mu = 0.13$ and $\sigma_{\rm Est} = 1.04$ for GPR, whereas ideally we expect $\mu = 0$ if the foregrounds have been perfectly subtracted and $\sigma_{\rm Est} = 1$ if our predictions for the expected statistical fluctuations are correct. The small, positive value of μ indicates that some low-level foregrounds possibly still remain in the residuals. However, note that in both cases the value of μ is smaller compared to that before foreground subtraction. This indicates an overall reduction in the level of foreground contamination for 21-cm signal estimation as compared to Paper II. The fact that we obtain $\sigma_{\rm Est} \approx 1$ for both PF and GPR roughly validates our error predictions including our treatment of the uncertainties due to foreground modelling.

For both PF and GPR, we see that the probability density function (PDF) of X is largely symmetric around a small positive mean value. We find that the bulk (∼90–95 per cent) of the X values lie in the central region $|X| \leq 2$. We find a few samples (∼1 per cent) at larger values of X (e.g. $X > 4$), which are trace amounts of unsubtracted residual foregrounds and can be considered outliers of the distribution. The fact that we do not see any significant outliers indicates that foregrounds have largely been subtracted from the data. Similar to Paper II, we find that PDFs are not well described by a Gaussian, whereas a Lorentzian distribution (also known as Cauchy distribution) represents the statistics better. The magenta dashed lines in Fig. 6 show the best-fitting Lorentzian PDFs

$$\rho(x) = \frac{1}{\pi\gamma}\left[\frac{\gamma^2}{(x - x_0)^2 + x_0^2}\right], \tag{21}$$

with x_0 and γ, respectively, denoting the peak location and the spread (half-width at half-maximum) of the distribution. We find $x_0 = 0.08$ and $\gamma = 0.30$ for PF, and $x_0 = 0.06$ and $\gamma = 0.26$ for GPR. Fig. 6 also shows the cumulative distribution function (CDF) of X, along with the CDF predicted by the best-fitting Lorentzian distribution. We see that for both the foreground subtraction methods, the CDF of X and the corresponding best-fitting Lorentzian CDF are in close agreement.

The quantity of our interest is the statistical uncertainties in the estimated $P(k_\perp, k_\parallel)$. We have found that X is well represented by a Lorentzian distribution, but it is not guaranteed that the statistical fluctuations in the estimated $P(k_\perp, k_\parallel)$ will also be described by the same. Wilensky, Brown & Hazelton (2023) recently found that this is likely to follow a Laplacian distribution, and converge to a Gaussian in most situations. We also note that the standard deviation is ill-defined for the Lorentzian distribution in our analysis. Considering this uncertainty, we have followed the standard practice of quoting

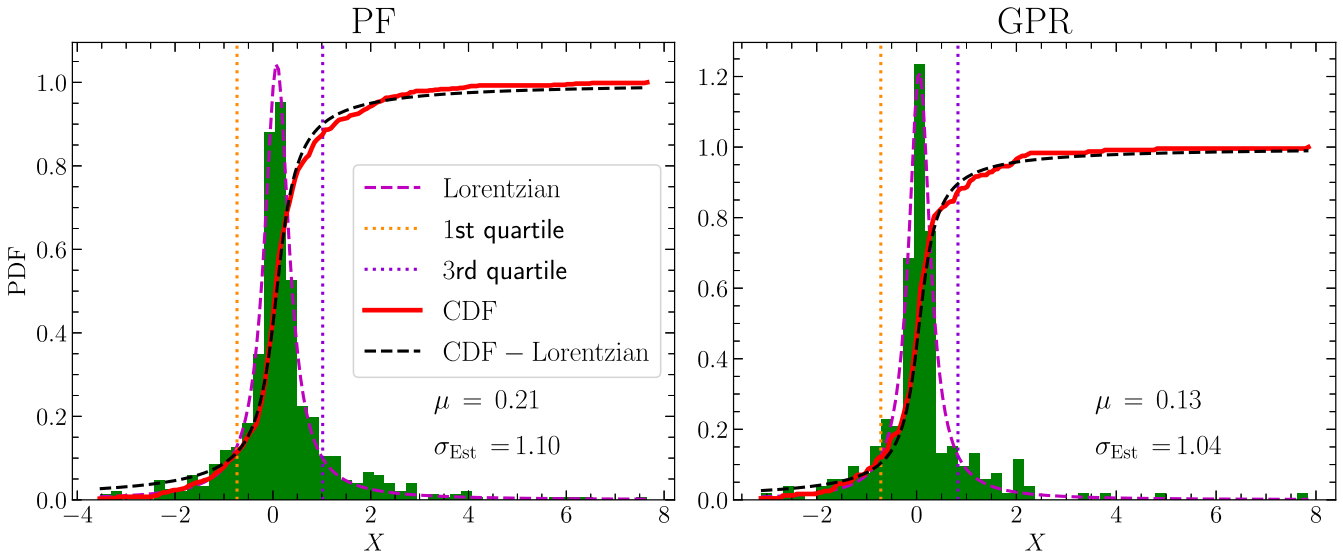

Figure 6. This figure shows the PDF (green vertical bars) and the CDF (red solid line) for $X = P(k_\perp, k_\parallel)/\delta P_N(k_\perp, k_\parallel)$. Lorentzian fits of the PDF and the CDF are shown by the magenta and black dashed lines, respectively. The orange and violet vertical lines show the first and the third quartiles of the best-fitting Lorentzian distribution. The mean (μ) and the standard deviation (σ) of X are annotated.

the 2σ error bars throughout the analysis. However, it is questionable whether these error bars actually represent the 95 per cent confidence intervals.

Following Paper II, for the subsequent analysis we have further scaled the error estimates with $\sigma_{\rm Est}$. Note that this additional scaling leads to a very small increase in the error bars as $\sigma_{\rm Est} \approx 1$ for both PF and GPR.

5 THE SPHERICAL PS

We can think of the foreground-subtracted residual $[C_\ell(\Delta\nu)]_{\rm res}$ in terms of two components:

$$[C_\ell(\Delta\nu)]_{\rm res} = [C_\ell(\Delta\nu)]_T + [C_\ell(\Delta\nu)]_R \ . \quad (22)$$

Here, $[C_\ell(\Delta\nu)]_T$ is the spatially isotropic 21-cm signal and $[C_\ell(\Delta\nu)]_R$ refers to the anisotropic components of $C_\ell(\Delta\nu)$, which is ideally expected to be within the noise level. Unsubtracted foregrounds, if present, will also contribute to $[C_\ell(\Delta\nu)]_R$.

Here, the 21-cm signal is modelled as

$$\left[C_{\ell_a}(\Delta\nu_n)\right]_T = \sum_i B_i(a,n)\,[P(k_i)]_T \ , \quad (23)$$

where $[P(k_i)]_T$ refers to the value of the spherical 21-cm PS in the ith bin and $B_i(a, n) = \sum_m A_{nm}$ is summed over the $(k_{\perp a}, k_{\parallel m})$ modes within this bin. We have applied an MLE (Paper II) to directly determine PS $P(k)$ from the measured $[C_\ell(\Delta\nu)]_{\rm res}$.

We have used all the available $[C_\ell(\Delta\nu)]_{\rm res}$ values to obtain the the model parameters $P(k)$ at six spherical k bins spanning the range $0.247\ {\rm Mpc}^{-1} < {\rm k} < 9.931\ {\rm Mpc}^{-1}$. We find that the goodness-of-fit parameter (reduced-χ^2) has a value of 1.14 with 629 degrees of freedom. Although this is a slightly poor fit for $[C_\ell(\Delta\nu)]_{\rm res}$, we have used the best-fitting $P(k)$ values to calculate the mean squared brightness temperature $\Delta^2(k) \equiv k^3P(k)/2\pi^2$. The red and blue curves in Fig. 7 show $|\Delta^2(k)|$ along with the corresponding 2σ error bars for PF and GPR, respectively. Note that we have marked the negative $\Delta^2(k)$ values with a (green) cross (X). Considering PF, we see that the $\Delta^2(k)$ values (red squares) are positive in all the k bins except the first k bin ($k = 0.247\ {\rm Mpc}^{-1}$). For GPR, $\Delta^2(k)$ is found to be negative

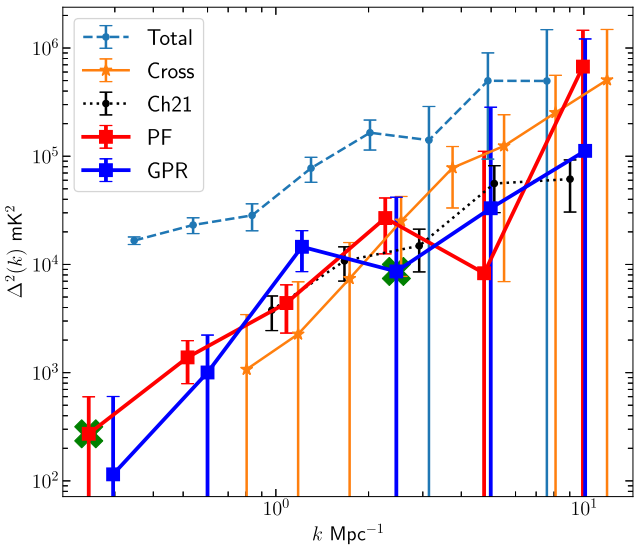

Figure 7. The mean squared brightness temperature fluctuations $\Delta^2(k)$ and the associated 2σ errors are shown. The red and blue squares show the values obtained in this work using PF and GPR, respectively. The green crosses (X) indicate negative values. The light blue (dashed), orange (solid), and black (dotted) curves present the results of Paper I, Paper II, and Ch21, respectively.

in the fourth bin ($k = 2.461\ {\rm Mpc}^{-1}$). While we expect $\Delta^2(k)$ for the 21-cm signal to be positive, the values inferred from observations can be negative due to statistical fluctuations. We see that the negative $\Delta^2(k)$ values are all within $0 \pm 2\sigma$, and we interpret these measurements as being consistent with zero with the negative values arising from statistical fluctuations. The values of $\Delta^2(k)$, σ, and the SNR ($=\Delta^2(k)/\sigma$) at different k bins are presented in Tables 1 and 2 for PF and GPR, respectively. Note that the correction due to signal loss, which we will discuss shortly, is considered for all the values quoted here. The 2σ upper limits $\Delta^2_{\rm UL}(k) = \Delta^2(k) + 2\sigma$ obtained from the two methods are also presented in their corresponding tables. Note that for the negative $\Delta^2(k)$ values, we have conservatively taken

Table 1. The mean squared brightness temperature fluctuations $\Delta^2(k)$, their errors σ, SNR, 2σ upper limits $\Delta^2_{\rm UL}(k) = \Delta^2(k) + 2\sigma$, and $[\Omega_{\rm H\,I} b_{\rm H\,I}]_{\rm UL}$ values are tabulated for PF.

k ($\rm Mpc^{-1}$)	$\Delta^2(k)$ ($\rm mK^2$)	1σ ($\rm mK^2$)	SNR	$\Delta^2_{\rm UL}(k)$ ($\rm mK^2$)	$[\Omega_{\rm H\,I} b_{\rm H\,I}]_{\rm UL}$
0.247	$-(16.50)^2$	$(12.78)^2$	-1.67	$(18.07)^2$	0.036
0.517	$(37.20)^2$	$(17.21)^2$	4.67	$(44.46)^2$	0.064
1.082	$(66.35)^2$	$(32.26)^2$	4.23	$(80.52)^2$	0.089
2.266	$(163.84)^2$	$(84.59)^2$	3.75	$(202.86)^2$	0.180
4.744	$(91.06)^2$	$(226.59)^2$	0.16	$(333.14)^2$	0.245
9.931	$(820.54)^2$	$(626.28)^2$	1.72	$(1207.37)^2$	0.759

Table 2. Same as Table 1 but for GPR.

k ($\rm Mpc^{-1}$)	$\Delta^2(k)$ ($\rm mK^2$)	1σ ($\rm mK^2$)	SNR	$\Delta^2_{\rm UL}(k)$ ($\rm mK^2$)	$[\Omega_{\rm H\,I} b_{\rm H\,I}]_{\rm UL}$
0.296	$(10.72)^2$	$(15.61)^2$	0.47	$(24.54)^2$	0.045
0.600	$(31.68)^2$	$(24.71)^2$	1.64	$(47.17)^2$	0.064
1.215	$(120.66)^2$	$(54.65)^2$	4.87	$(143.29)^2$	0.153
2.461	$-(92.83)^2$	$(128.75)^2$	-0.52	$(182.08)^2$	0.158
4.985	$(181.84)^2$	$(353.91)^2$	0.26	$(532.51)^2$	0.387
10.098	$(334.80)^2$	$(741.99)^2$	0.20	$(1101.45)^2$	0.681

Table 3. The upper limits from 21-cm IM experiments using this uGMRT Band 3 data. The $[\Omega_{\rm H\,I} b_{\rm H\,I}]_{\rm UL}$ values quoted inside the parentheses (...) are obtained when a single, k-independent value of $[\Omega_{\rm H\,I} b_{\rm H\,I}]$ is directly constrained from $C_\ell(\Delta\nu)$ (Section 6).

Works	z	k ($\rm Mpc^{-1}$)	$[\Delta^2(k)]_{\rm UL}$ ($\rm mK^2$)	$[\Omega_{\rm H\,I} b_{\rm H\,I}]_{\rm UL}$
Ch21	1.96	0.99	$(58.57)^2$	0.09
	2.19	0.97	$(61.49)^2$	0.11
	2.62	0.95	$(60.89)^2$	0.12
	3.58	0.99	$(105.85)^2$	0.24
Paper I	2.28	0.35	$(133.97)^2$	0.23
Paper II	2.28	0.80	$(58.67)^2$	0.072 (0.061)
This work				
PF	2.28	0.25	$(18.07)^2$	0.036 (**0.022**)
GPR		0.30	$(24.54)^2$	0.045 (**0.031**)

$\Delta^2(k) = 0$ and quoted the 2σ values as the upper limits. The tightest constraint is found to be $\Delta^2_{\rm UL}(k) \leq (18.07)^2\,\rm mK^2$ at $k = 0.247\,\rm Mpc^{-1}$ from PF.

The results obtained from previous works with the same observational data are also shown in Fig. 7. Paper I (blue dashed) used the total TGE, whereas Paper II (solid orange) has used the cross TGE, which was also adopted for this work. We also present the results from Ch21 (black dotted line) who have estimated the 21-cm PS in delay space using 8-MHz bandwidth data taken from the same observations at $\nu_{\rm c} = 445$ MHz (z = 2.19). All of these works have used foreground avoidance and the PS estimates are restricted to $k > 0.35$, 0.80, and 1 $\rm Mpc^{-1}$, respectively, due to the presence of the foreground wedge.

In the present analysis, we are able to foray deeper into the small k range extending up to $k = 0.247$ $\rm Mpc^{-1}$. While our results are consistent with the earlier estimates of Paper II and Ch21 at larger k where there is an overlap, we obtain a tighter upper limit of $\Delta^2_{\rm UL}(k) \leq (18.07)^2\,\rm mK^2$ at the smallest k bin of $k = 0.247$ $\rm Mpc^{-1}$. Note that a comparison of all the upper limits obtained from the same observation is presented in Table 3.

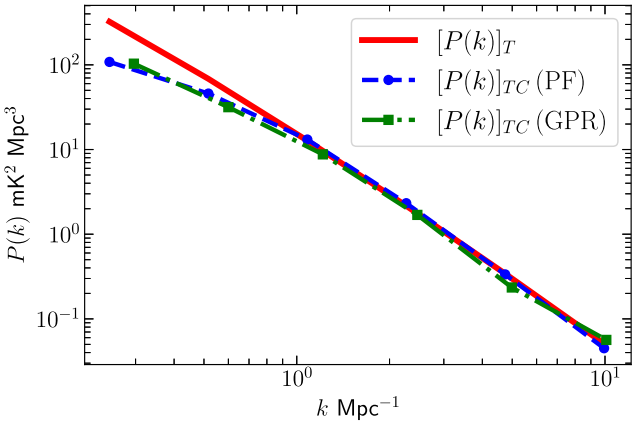

Figure 8. Correction in the spherical PS $P(k)$ to take into account the signal loss due to the foreground removal algorithms. The red line shows the predicted 21-cm PS, whereas the blue and green lines show the loss-corrected 21-cm PS from PF and GPR, respectively.

The results presented above have all been corrected for possible 21-cm signal loss due to foreground subtraction. To estimate the correction factor, we have calculated the spherically binned PS using both $[C_\ell(\Delta\nu)]_T$ (equation 5) and $[C_\ell(\Delta\nu)]_{\rm TC}$ (equation 11). The analysis has been done identical to the analysis of the measured $C_{\boldsymbol{\ell}}(\Delta\nu)$. Fig. 8 shows $[P(k)]_T$ and $[P(k)]_{\rm TC}$ estimated from $[C_\ell(\Delta\nu)]_T$ and $[C_\ell(\Delta\nu)]_{\rm TC}$, respectively. We have multiplied the estimated $\Delta^2_{\rm UL}(k)$ at each k bin with the corresponding correction factor $[P(k)]_T/[P(k)]_{\rm TC}$ to account for the signal loss. We see that the correction factor is very close to unity for all the k bins expect for the smallest two bins. Considering PF (GPR), the correction factor has values $\sim$3.0 ($\sim$2.2) and $\sim$1.5 ($\sim$1.5) for the first two k bins, respectively. The rather large signal loss in the lowest k bin can be attributed to the fact that we have used $[\Delta\nu]_{0.4}$ for Set I, which has the smallest $k_\perp$ modes.

The upper limits on $\Delta^2(k)$ allow us to constrain the cosmological H I abundance parameter $[\Omega_{\rm H\,I} b_{\rm H\,I}]$. The assumption that the H I distribution traces the underlying matter distribution allows us to express $P_T(\boldsymbol{k})$ in terms of $P^s_m(\boldsymbol{k})$, the matter PS in redshift space through the relation (equation 17 of Paper II)

$$P_T(\boldsymbol{k}) = [\Omega_{\rm H\,I} b_{\rm H\,I}]^2\,\bar{T}^2\,P^s_m(\boldsymbol{k}), \tag{24}$$

with the mean brightness temperature $\bar{T}$ given in equation (6). Using the estimated $\Delta^2_{\rm UL}(k)$ values, we place 2σ upper limits $[\Omega_{\rm H\,I} b_{\rm H\,I}]_{\rm UL}$, which are also presented in Tables 1 and 2. The best constraint is obtained using PF, where we find $[\Omega_{\rm H\,I} b_{\rm H\,I}]_{\rm UL} \leq 0.036$ from the first bin $k = 0.247$ $\rm Mpc^{-1}$.

6 CONSTRAINING $[\Omega_{\rm H\,I} b_{\rm H\,I}]$

Here, we use the foreground-subtracted residuals $[C_{\boldsymbol{\ell}}(\Delta\nu)]_{\rm res}$ to directly measure $[\Omega_{\rm H\,I} b_{\rm H\,I}]$ without estimating the spherical PS $P(k)$. A similar approach was taken in Paper II (section 6) in the context of foreground avoidance. The idea here is to use the full available information to find the maximum likelihood solution for the only parameter $[\Omega_{\rm H\,I} b_{\rm H\,I}]^2$. We have modelled $[C_{\boldsymbol{\ell}}(\Delta\nu)]_{\rm res}$ using equation (22), and quantify the 21-cm signal $[C_\ell(\Delta\nu)]_T$ using equation (5). The formalism allows us to parametrize the entire 21-cm signal with the parameter $[\Omega_{\rm H\,I} b_{\rm H\,I}]^2$, which we estimate by the MLE. We have used $[C_\ell(\Delta\nu)]_{\rm TC}$ instead of $[C_\ell(\Delta\nu)]_T$ as the model 21-cm signal to account for the signal loss.

Table 4. The upper limits on $[\Omega_{\rm H\,I} b_{\rm H\,I}]$ from PF.

Set	ℓ range	$[\Omega_{\rm H\,I} b_{\rm H\,I}]^2$ $\times 10^{-4}$	SNR	$[\Omega_{\rm H\,I} b_{\rm H\,I}]_{\rm UL}$ $\times 10^{-2}$
I	<2000	-3.13 ± 2.53	−1.24	2.249
II	2000–4000	6.33 ± 4.49	1.41	3.913
III	>4000	44.67 ± 12.87	3.47	8.391
Combined	<6300	0.44 ± 2.17	0.20	**2.187**

Table 5. The upper limits on $[\Omega_{\rm H\,I} b_{\rm H\,I}]$ from GPR.

Set	ℓ range	$[\Omega_{\rm H\,I} b_{\rm H\,I}]^2$ $\times 10^{-4}$	SNR	$[\Omega_{\rm H\,I} b_{\rm H\,I}]_{\rm UL}$ $\times 10^{-2}$
I	<2000	3.55 ± 5.46	0.65	3.805
II	2000–4000	-3.59 ± 6.63	−0.54	3.641
III	>4000	12.67 ± 16.50	0.77	6.758
Combined	<6300	1.40 ± 4.09	0.34	**3.093**

As mentioned in Section 3.5, we have divided the measured $C_\ell(\Delta\nu)$ into three sets. We have applied the MLE on the $[C_\ell(\Delta\nu)]_{\rm res}$ values from the individual sets to constrain $[\Omega_{\rm H\,I} b_{\rm H\,I}]^2$ for which the results for PF are presented in Table 4. Here, Set I yields a tight constraint, where we have $[\Omega_{\rm H\,I} b_{\rm H\,I}]^2 = -3.13 \times 10^{-4} \pm 2.53 \times 10^{-4}$ with the corresponding upper limit $[\Omega_{\rm H\,I} b_{\rm H\,I}]_{\rm UL} \leq 2.25 \times 10^{-2}$ at 2σ level. Note that while calculating the upper limit we have conservatively set $[\Omega_{\rm H\,I} b_{\rm H\,I}]^2$ to zero when it is negative. We note that Set II also yields a considerably tight upper limit $[\Omega_{\rm H\,I} b_{\rm H\,I}]_{\rm UL} \leq 3.91 \times 10^{-2}$ on the H I abundance. Set III, however, gives a relatively poor constraint $[\Omega_{\rm H\,I} b_{\rm H\,I}]_{\rm UL} \leq 8.39 \times 10^{-2}$. This weak upper limit from Set III can be attributed to larger noise and residual foregrounds in these $C_\ell(\Delta\nu)$ values. We have also combined $[C_\ell(\Delta\nu)]_{\rm res}$ for all the three sets to maximize the SNR and constrain $[\Omega_{\rm H\,I} b_{\rm H\,I}]^2$. We find $[\Omega_{\rm H\,I} b_{\rm H\,I}]^2 = 4.4 \times 10^{-5} \pm 2.17 \times 10^{-4}$, which yields the tightest constraint $[\Omega_{\rm H\,I} b_{\rm H\,I}]_{\rm UL} \leq 2.19 \times 10^{-2}$ on the upper limit. The best-fitting values obtained from the combined set are also presented in Table 4.

The results from GPR are presented in Table 5. Here, the constraints from Set II are better than those from Sets I and III, while the tightest limit from GPR $[\Omega_{\rm H\,I} b_{\rm H\,I}]_{\rm UL} \leq 3.09 \times 10^{-2}$ is obtained by combining all three sets. Comparing Tables 4 and 5, we see that GPR performs better than PF for Sets II and III, whereas PF yields tighter constraints for Set I and when all the sets are combined. Further, we also find that the SNR values are always smaller for GPR for which all the $[\Omega_{\rm H\,I} b_{\rm H\,I}]^2$ are within $0 \pm \sigma$.

7 SUMMARY AND CONCLUSIONS

This paper is the third in a series of papers that focus on 21-cm IM considering uGMRT data centred at 432.8 MHz, which corresponds to $z = 2.28$. The data analysed here are described in Section 2, and also our earlier work Paper I. In this work, we have used the MAPS $C_\ell(\Delta\nu)$ estimated in Paper II using the 'cross' TGE, which uses the cross-correlation of the RR and LL polarizations. This has the advantage of naturally avoiding the noise bias and other systematics that are uncorrelated between the two polarizations. While both of our earlier works used foreground avoidance, this work implements foreground removal, which does away with the foreground wedge allowing us to use the entire $(k_\perp, k_\parallel)$ plane for estimating the 21-cm PS.

The foregrounds generally exhibit a smooth spectral behaviour, and the contribution to the measured $C_\ell(\Delta\nu)$ is expected to remain correlated even at large $\Delta\nu$. In contrast, the predicted 21-cm signal $[C_\ell(\Delta\nu)]_T$ decorrelates rapidly with increasing $\Delta\nu$, and is close to zero beyond a characteristic frequency scale $\Delta\nu > [\Delta\nu]$. We find $[\Delta\nu] \sim 0.5$–1.0 MHz for most of the ℓ range considered here (Fig. 1).

Here, we have considered two different approaches for foreground modelling and subtraction, the first being PF and the second being GPR. For both, we have used the range $\Delta\nu > [\Delta\nu]$ to estimate the foreground contribution to $C_\ell(\Delta\nu)$. In PF, this is modelled as an even polynomial in $\Delta\nu$. The values of $[\Delta\nu]$ and the polynomial order are different for each $\boldsymbol{\ell}$, and the details are presented in Section 3.3. For GPR, the details of foreground modelling are presented in Section 3.4. For both PF and GPR, we have extrapolated the foreground model predictions to the range $\Delta\nu \leq [\Delta\nu]$, and subtracted out the foreground predictions from $C_\ell(\Delta\nu)$. We have used the residual $[C_\ell(\Delta\nu)]_{\rm res}$, after foreground subtraction, in the range $\Delta\nu \leq [\Delta\nu]$ to constrain the 21-cm signal. The foreground subtraction introduces a loss in the 21-cm signal, which we have corrected in the quantitative results presented in this work. The cylindrical PS $P(k_\perp, k_\parallel)$, estimated from $[C_\ell(\Delta\nu)]_{\rm res}$, appears to be largely free of foreground contribution, and there is no indication of the foreground wedge (Fig. 5). The noise statistics, quantified by X (equation 20), are well described by a Lorentzian distribution that is largely symmetric around $X = 0$. The distribution of X does not exhibit any large outliers, both positive or negative, indicating that the estimated PS is free of systematics.

The residual $[C_\ell(\Delta\nu)]_{\rm res}$ has been used to estimate the spherical PS $P(k)$ using MLE, which utilizes the statistical isotropy of the expected 21-cm signal (Section 5). The estimated $\Delta^2(k)$ exhibits values of both signs; however, the negative values are within $0 \pm 2\sigma$ and we interpret these as arising from statistical fluctuations. We have found that the $|\Delta^2(k)|$ values and their uncertainties increase monotonically with increasing k. The results from PF and GPR are presented in Tables 1 and 2, respectively. Our best result $\Delta^2_{\rm UL}(k) \leq (18.07)^2$ mK^2 comes from PF at the smallest k bin $k = 0.247$ Mpc^{-1}. This upper limit corresponds to $[\Omega_{\rm H\,I} b_{\rm H\,I}]_{\rm UL} \leq 0.036$. Considering GPR, the corresponding values are $\Delta^2_{\rm UL}(k) \leq (24.54)^2$ mK^2 and $[\Omega_{\rm H\,I} b_{\rm H\,I}]_{\rm UL} \leq 0.045$ at $k = 0.296$ Mpc^{-1}, which is the smallest k bin for GPR. Although the best upper limit comes from PF, we find that overall the $\Delta^2_{\rm UL}(k)$ values from PF and GPR are quite comparable (Fig. 7).

In a different approach, we have combined the estimated $[C_\ell(\Delta\nu)]_{\rm res}$ to directly constrain the single parameter $[\Omega_{\rm H\,I} b_{\rm H\,I}]$ (Section 6), for which the results for PF and GPR are presented in Tables 4 and 5, respectively. We see that PF provides the best upper limits. Considering a particular subset of the $\boldsymbol{\ell}$ grids (Set I), we can place a tight upper limit $[\Omega_{\rm H\,I} b_{\rm H\,I}]_{\rm UL} \leq 2.25 \times 10^{-2}$. It may, however, be noted that this particular limit incorporates a large correction factor to account for the 21-cm signal loss due to foreground subtraction. The $\Delta\nu$ range used for estimating and extrapolating the foreground models is also different from that used for the other sets. The tightest upper limit $[\Omega_{\rm H\,I} b_{\rm H\,I}]_{\rm UL} \leq 2.19 \times 10^{-2}$, however, is obtained when we combine the entire $\boldsymbol{\ell}$ grid (Sets I, II, and III). This limit is a factor of 3 improvement over the earlier works that have used foreground avoidance. A comparison of the upper limits obtained from this uGMRT Band 3 observation is presented in Table 3.

While we have removed the foregrounds to a large extent, we still have not detected the 21-cm signal due to noise. On a positive note, we have put tight constraints on the upper limit of the 21-cm IM signal. Several different studies, both observational (e.g. Rhee et al. 2018; Chowdhury et al. 2020; Ho et al. 2021; Amiri et al. 2023; Cunnington et al. 2023a) and theoretical (e.g. Padmanabhan,

Choudhury & Refregier 2015; Sarkar et al. 2016), indicate that $[\Omega_{\rm H\,I} b_{\rm H\,I}] \sim 10^{-3}$ for the redshift we have considered here. Our present upper limits on $[\Omega_{\rm H\,I} b_{\rm H\,I}]_{\rm UL} \leq 0.022$ is ∼10 times larger than currently estimated values. However, this upper limit is nearly three times tighter over previous IM measurements at this redshift.

ACKNOWLEDGEMENTS

We thank the anonymous reviewer for the detailed comments, which helped us to improve the work. We thank the staff of GMRT for making this observation possible. GMRT is run by National Centre for Radio Astrophysics (NCRA) of the Tata Institute of Fundamental Research (TIFR). AG would like to thank IUCAA, Pune for providing support through the associateship programme. SB would like to acknowledge funding provided under the MATRICS grant SERB/F/9805/2019-2020 and AG would like to acknowledge funding provided under the SERB-SURE grant SUR/2022/000595 of the Science & Engineering Research Board, a statutory body of Department of Science & Technology (DST), Government of India. Part of this work has used the Supercomputing facility 'PARAM Shakti' of IIT Kharagpur established under National Supercomputing Mission (NSM), Government of India, and supported by Centre for Development of Advanced Computing (CDAC), Pune.

DATA AVAILABILITY

The observed data are publicly available through the GMRT data archive[7] under the proposal code 32_120. The simulated data used here are available upon reasonable request to the corresponding author.

REFERENCES

Abdurashidova Z. et al., 2022, ApJ, 925, 221
Ali S. S., Bharadwaj S., Chengalur J. N., 2008, MNRAS, 385, 2166
Ambikasaran S., Foreman-Mackey D., Greengard L., Hogg D. W., O'Neil M., 2015, IEEE Trans. Pattern Anal. Mach. Intell., 38, 252
Amiri M. et al., 2023, ApJ, 947, 16
Ansari R. et al., 2012, A&A, 540, A129
Bagla J. S., Khandai N., Datta K. K., 2010, MNRAS, 407, 567
Battye R. A., Browne I. W. A., Dickinson C., Heron G., Maffei B., Pourtsidou A., 2013, MNRAS, 434, 1239
Bharadwaj S., Ali S. S., 2005, MNRAS, 356, 1519
Bharadwaj S., Pandey S. K., 2003, JA&A, 24, 23
Bharadwaj S., Sethi S. K., 2001, JA&A, 22, 293
Bharadwaj S., Srikant P. S., 2004, JA&A, 25, 67
Bharadwaj S., Nath B. B., Sethi S. K., 2001, JA&A, 22, 21
Bharadwaj S., Sethi S. K., Saini T. D., 2009, Phys. Rev. D, 79, 083538
Bharadwaj S., Pal S., Choudhuri S., Dutta P., 2018, MNRAS, 483, 5694
Bishop C. M., 2006, Pattern Recognition and Machine Learning. Springer, New York
Bull P., Camera S., Raccanelli A., Blake C., Ferreira P., Santos M., Schwarz D. J., 2015a, in Proceedings of Advancing Astrophysics with the Square Kilometre Array (AASKA14) Vol. 215, Measuring Baryon Acoustic Oscillations with Future SKA Surveys. p. 24. Available at: https://arxiv.org/abs/1501.04088
Bull P., Ferreira P. G., Patel P., Santos M. G., 2015b, ApJ, 803, 21
Chakraborty A. et al., 2019a, MNRAS, 487, 4102
Chakraborty A. et al., 2019b, MNRAS, 490, 243
Chakraborty A. et al., 2021, ApJ, 907, L7
Chang T.-C., Pen U.-L., Peterson J. B., McDonald P., 2008, Phys. Rev. Lett., 100, 091303
Chang T.-C., Pen U.-L., Bandura K., Peterson J. B., 2010, Nature, 466, 463
Chapman E. et al., 2012, MNRAS, 423, 2518
Chatterjee S., Bharadwaj S., Choudhuri S., Sethi S., Patwa A. K., 2022, MNRAS, 519, 2410
CHIME Collaboration, 2022, ApJS, 261, 29
Choudhuri S., Bharadwaj S., Ghosh A., Ali S. S., 2014, MNRAS, 445, 4351
Choudhuri S., Bharadwaj S., Roy N., Ghosh A., Ali S. S., 2016a, MNRAS, 459, 151
Choudhuri S., Bharadwaj S., Chatterjee S., Ali S. S., Roy N., Ghosh A., 2016b, MNRAS, 463, 4093
Choudhuri S., Bharadwaj S., Ali S. S., Roy N., Intema H. T., Ghosh A., 2017, MNRAS, 470, L11
Choudhuri S., Ghosh A., Roy N., Bharadwaj S., Intema H. T., Ali S. S., 2020, MNRAS, 494, 1936
Chowdhury A., Kanekar N., Chengalur J. N., Sethi S., Dwarakanath K. S., 2020, Nature, 586, 369
Cunnington S. et al., 2023a, MNRAS, 518, 6262
Cunnington S. et al., 2023b, MNRAS, 523, 2453
Datta K. K., Choudhury T. R., Bharadwaj S., 2007, MNRAS, 378, 119
Datta A., Bowman J. D., Carilli C. L., 2010, ApJ, 724, 526
Dawson K. S. et al., 2016, AJ, 151, 44
Di Matteo T., Perna R., Abel T., Rees M. J., 2002, ApJ, 564, 576
Dillon J. S. et al., 2014, Phys. Rev. D, 89, 023002
Dillon J. S. et al., 2015, Phys. Rev. D, 91, 123011
Eisenstein D. J., Hu W., 1998, ApJ, 496, 605
Elahi K. M. A. et al., 2023, MNRAS, 520, 2094
Ewall-Wice A. et al., 2021, MNRAS, 500, 5195
Gan H. et al., 2022, A&A, 663, A9
Ghosh A., Bharadwaj S., Ali S. S., Chengalur J. N., 2011a, MNRAS, 411, 2426
Ghosh A., Bharadwaj S., Ali S. S., Chengalur J. N., 2011b, MNRAS, 418, 2584
Ghosh A., Prasad J., Bharadwaj S., Ali S. S., Chengalur J. N., 2012, MNRAS, 426, 3295
Gupta Y. et al., 2017, Curr. Sci., 113, 707
Hazra D. K., Guha Sarkar T., 2012, Phys. Rev. Lett., 109, 121301
Ho M.-F., Bird S., Garnett R., 2021, MNRAS, 507, 704
Kennedy F., Bull P., 2021, MNRAS, 506, 2638
Kennedy F., Bull P., Wilensky M. J., Burba J., Choudhuri S., 2023, ApJS, 266, 23
Kern N. S., Liu A., 2021, MNRAS, 501, 1463
Kumar J., Dutta P., Roy N., 2020, MNRAS, 495, 3683
Kumar J., Dutta P., Choudhuri S., Roy N., 2022, MNRAS, 512, 186
Lanzetta K. M., Wolfe A. M., Turnshek D. A., 1995, ApJ, 440, 435
Liu A., Tegmark M., 2012, MNRAS, 419, 3491
Liu A., Parsons A. R., Trott C. M., 2014a, Phys. Rev. D, 90, 023018
Liu A., Parsons A. R., Trott C. M., 2014b, Phys. Rev. D, 90, 023019
Loeb A., Wyithe J. S. B., 2008, Phys. Rev. Lett., 100, 161301
Long H., Morales-Gutiérrez C., Montero-Camacho P., Hirata C. M., 2022, preprint (arXiv:2210.02385)
Mazumder A., Chakraborty A., Datta A., Choudhuri S., Roy N., Wadadekar Y., Ishwara-Chandra C. H., 2020, MNRAS, 495, 4071
McMullin J. P., Waters B., Schiebel D., Young W., Golap K., 2007, in Shaw R. A., Hill F., Bell D. J., eds, ASP Conf. Ser. Vol. 376, Astronomical Data Analysis Software and Systems XVI. Astron. Soc. Pac., San Francisco, p. 127
Mertens F. G., Ghosh A., Koopmans L. V. E., 2018, MNRAS, 478, 3640
Mertens F. G. et al., 2020, MNRAS, 493, 1662
Mondal R., Bharadwaj S., Datta K. K., 2018, MNRAS, 474, 1390
Mondal R., Bharadwaj S., Iliev I. T., Datta K. K., Majumdar S., Shaw A. K., Sarkar A. K., 2019, MNRAS, 483, L109
Morales M. F., Hazelton B., Sullivan I., Beardsley A., 2012, ApJ, 752, 137
Newburgh L. B. et al., 2016, in Hall H. J., Gilmozzi R., Marshall H. K., eds, Proc. SPIE Conf. Ser. Vol. 9906, Ground-based and Airborne Telescopes VI. SPIE, Bellingham, p. 99065X
Noterdaeme P. et al., 2012, A&A, 547, L1
Offringa A. R., de Bruyn A. G., Biehl M., Zaroubi S., Bernardi G., Pandey V. N., 2010, MNRAS, 405, 155

[7] https://naps.ncra.tifr.res.in/goa/

Offringa A. R., van de Gronde J. J., Roerdink J. B. T. M., 2012, A&A, 539, A95
Paciga G. et al., 2011, MNRAS, 413, 1174
Padmanabhan H., Choudhury T. R., Refregier A., 2015, MNRAS, 447, 3745
Pal S., Bharadwaj S., Ghosh A., Choudhuri S., 2021, MNRAS, 501, 3378
Pal S. et al., 2022, MNRAS, 516, 2851
Parsons A. R., Backer D. C., 2009, AJ, 138, 219
Paul S., Santos M. G., Chen Z., Wolz L., 2023, preprint (arXiv:2301.11943)
Planck Collaboration VI, 2020, A&A, 641, A6
Pober J. C. et al., 2013, ApJ, 768, L36
Pober J. C. et al., 2014, ApJ, 782, 66
Pober J. C. et al., 2016, ApJ, 819, 8
Rasmussen C., Williams C., 2006, Gaussian Processes for Machine Learning, Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA, USA
Rhee J., Lah P., Briggs F. H., Chengalur J. N., Colless M., Willner S. P., Ashby M. L. N., Le Fèvre O., 2018, MNRAS, 473, 1879
Saha P., Bharadwaj S., Roy N., Choudhuri S., Chattopadhyay D., 2019, MNRAS, 489, 5866
Saha P., Bharadwaj S., Chakravorty S., Roy N., Choudhuri S., Günther H. M., Smith R. K., 2021, MNRAS, 502, 5313
Saiyad Ali S., Bharadwaj S., Pandey S. K., 2006, MNRAS, 366, 213
Santos M. G., Cooray A., Knox L., 2005, ApJ, 625, 575
Sarkar D., Bharadwaj S., Anathpindika S., 2016, MNRAS, 460, 4310
Shaver P. A., Windhorst R. A., Madau P., de Bruyn A. G., 1999, A&A, 345, 380
Slosar A. et al., 2019, Bull. Am. Astron. Soc., 51, 53
Subrahmanya C. R., Manoharan P. K., Chengalur J. N., 2017, JA&A, 38, 10
Swarup G., Ananthakrishnan S., Kapahi V. K., Rao A. P., Subrahmanya C. R., Kulkarni V. K., 1991, Curr. Sci., 60, 95
Switzer E. R. et al., 2013, MNRAS, 434, L46
Trott C. M., 2016, MNRAS, 461, 126
Trott C. M., Wayth R. B., Tingay S. J., 2012, ApJ, 757, 101
Trott C. M., Mondal R., Mellema G., Murray S. G., Greig B., Line J. L. B., Barry N., Morales M. F., 2022, A&A, 666, A106
Vedantham H., Udaya Shankar N., Subrahmanyan R., 2012, ApJ, 745, 176
Visbal E., Loeb A., Wyithe S., 2009, J. Cosmol. Astropart. Phys., 2009, 030
Wilensky M. J., Brown J., Hazelton B. J., 2023, MNRAS, 521, 5191
Williams C., Rasmussen C., 1996, Advances in Neural Information Processing Systems 8. MIT Press, Cambridge, MA, USA, p. 514
Wolfe A. M., Lanzetta K. M., Foltz C. B., Chaffee F. H., 1995, ApJ, 454, 698
Wolz L. et al., 2021, MNRAS, 510, 3495
Wyithe J. S. B., Loeb A., Geil P. M., 2008, MNRAS, 383, 1195
Zafar T., Péroux C., Popping A., Milliard B., Deharveng J. M., Frank S., 2013, A&A, 556, A141
Zaldarriaga M., Furlanetto S. R., Hernquist L., 2004, ApJ, 608, 622

APPENDIX A: RESULTS FOR SETS I, II, AND III

This appendix contains figures showing the measured $C_\ell(\Delta\nu)$ along with the respective foreground models $[C_\ell(\Delta\nu)]_{\rm FG}$ for all the unflagged ℓ grid points that were used to constrain the 21-cm signal. The corresponding residuals $[C_\ell(\Delta\nu)]_{\rm res}$ after foreground subtraction are shown in separate figures. As mentioned in Section 3.5, we have divided the measured $C_\ell(\Delta\nu)$ into three sets, $\ell < 2000$ (Set I), $2000 < \ell < 4000$ (Set II), and $\ell > 4000$ (Set III), and analysed these separately. Considering PF, Figs (A1, A2), (A3, A4), and (A5, A6) show the results for Sets I, II, and III, respectively. The results for GPR from all three sets are shown together in Figs A7 and A8.

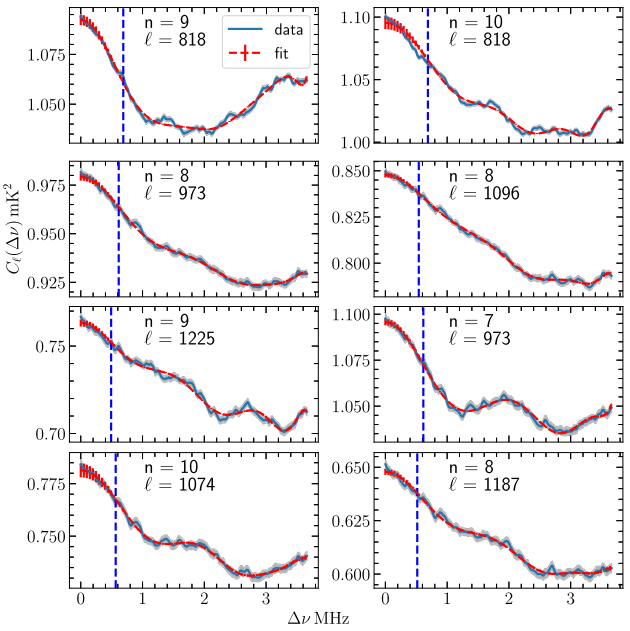

Figure A1. For Set I, the blue solid lines show the measured $C_\ell(\Delta\nu)$ with 2σ uncertainties (grey shaded region) expected from scaled system noise. The red dashed lines and the associated error bars show the best-fitting foreground models for PF and their 2σ uncertainties. The values of ℓ and n (used for foreground modelling) are mentioned in the respective panels. The vertical blue dashed lines show $[\Delta\nu]_{0.4}$.

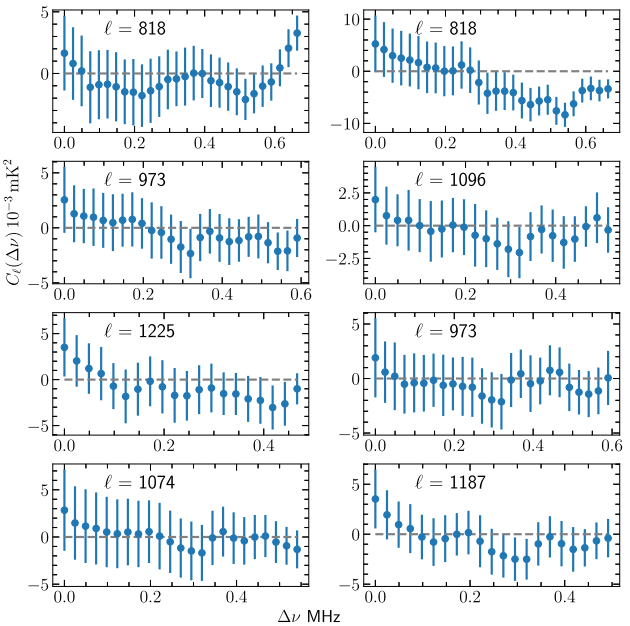

Figure A2. For Set I, the residual $[C_\ell(\Delta\nu)]_{\rm res}$ in the range $\Delta\nu \le [\Delta\nu]_{0.4}$ with the 2σ error bars (combining the noise and fitting errors), for the ℓ quoted in the panel.

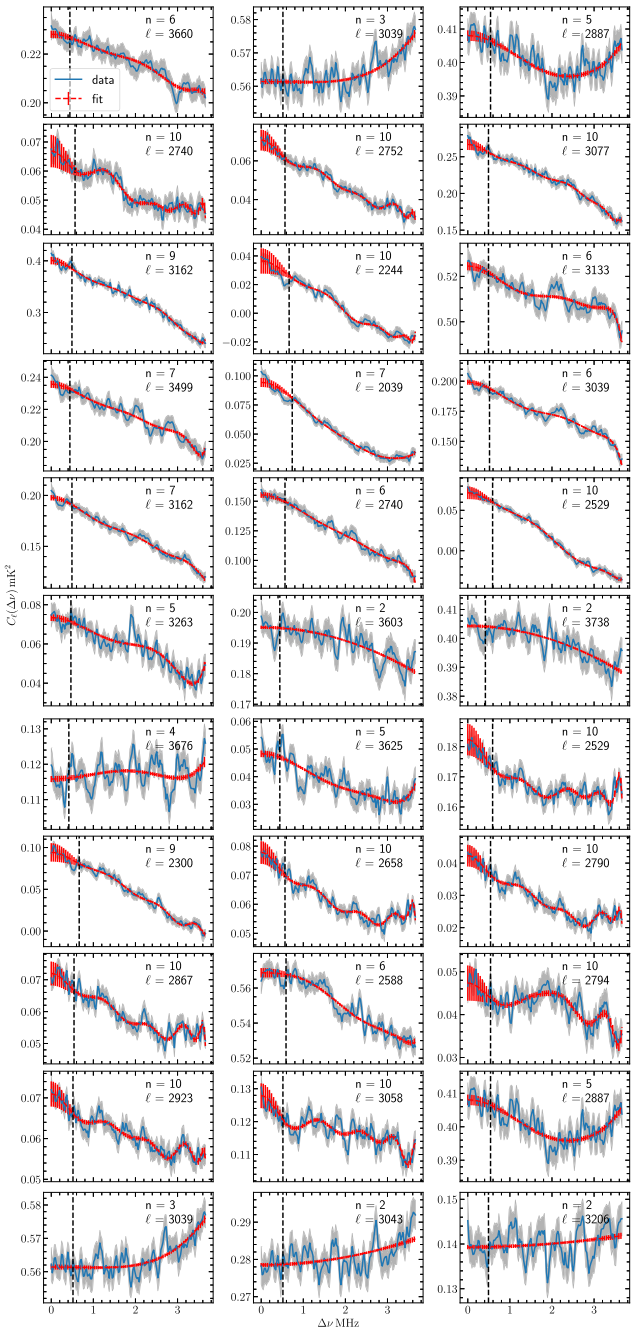

Figure A3. Same as Fig. A1 but for Set II. Here, the vertical black dashed line shows $[\Delta\nu]_{0.1}$.

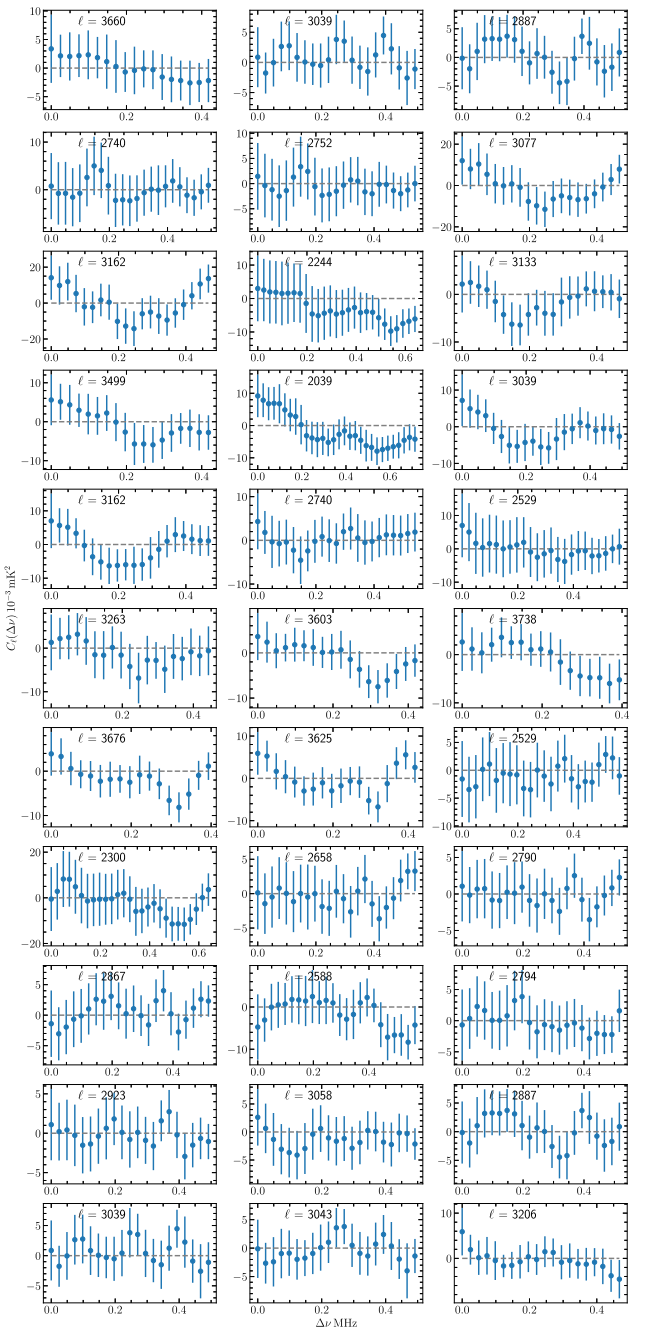

Figure A4. Same as Fig. A2 but for Set II considering the range $\Delta\nu \leq [\Delta\nu]_{0.1}$.

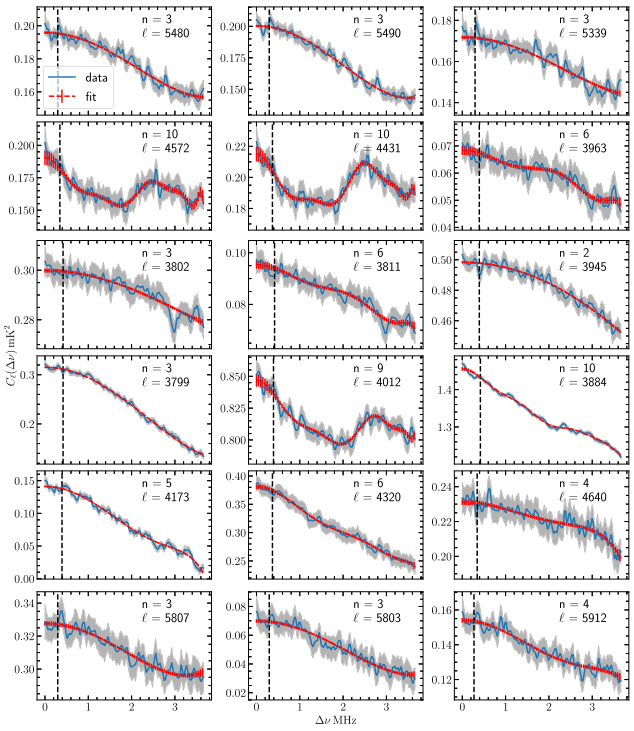

Figure A5. Same as Fig. A1 but for Set III. Here, the vertical black dashed line shows $[\Delta\nu]_{0.1}$.

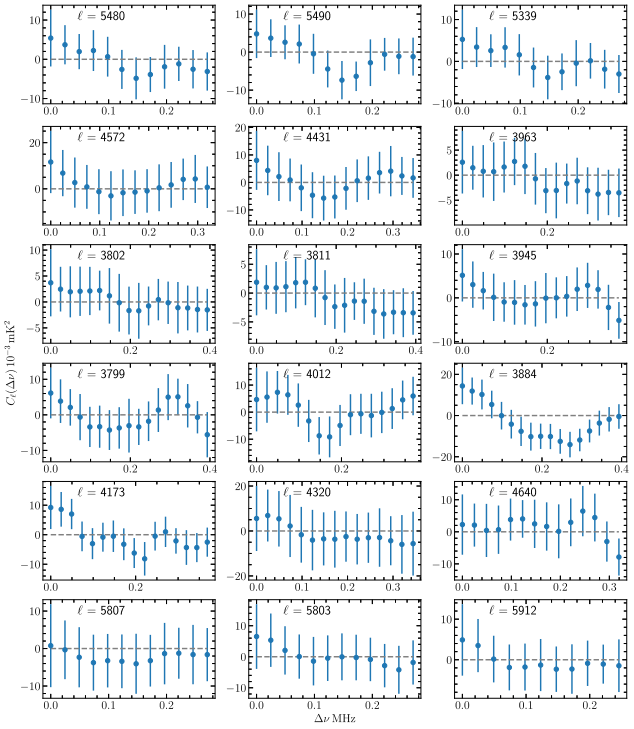

Figure A6. Same as Fig. A2 but for Set III considering the range $\Delta\nu \leq [\Delta\nu]_{0.1}$.

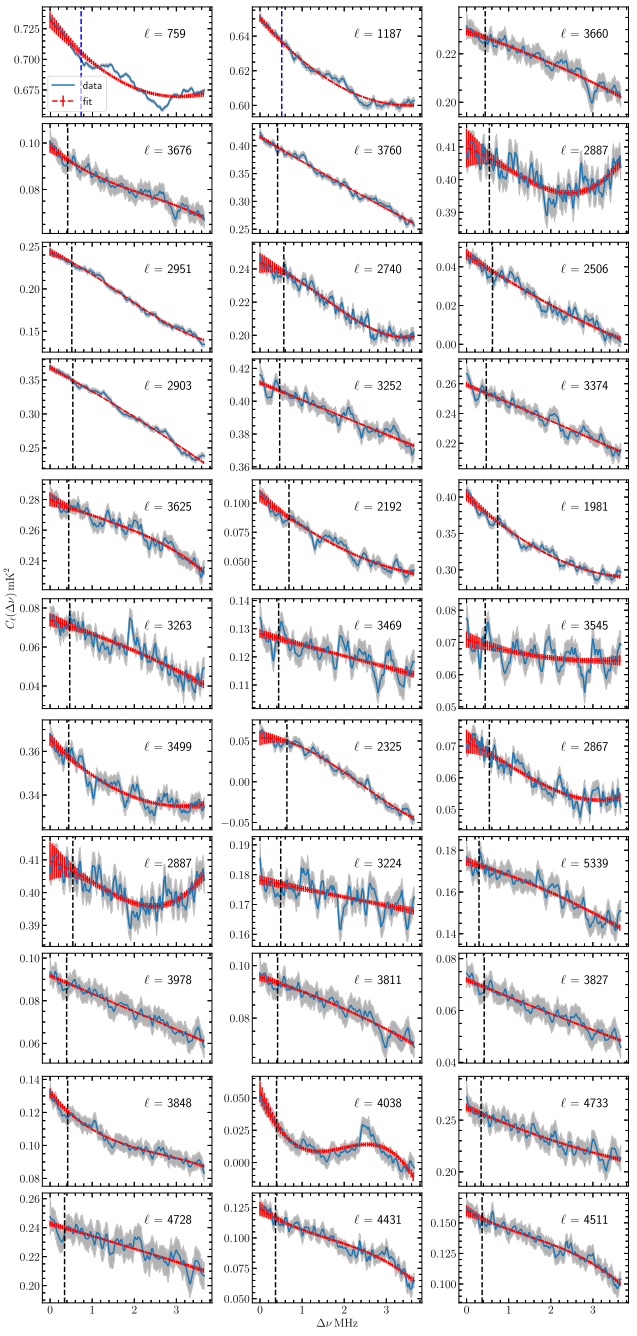

Figure A7. Same as Fig. A1 but for GPR and for all the sets. Here, the vertical blue dashed lines show $[\Delta\nu]_{0.4}$ in the first two panels that correspond to $\ell = 759$ and 1187, and the vertical black dashed lines show $[\Delta\nu]_{0.1}$ in all the other panels.

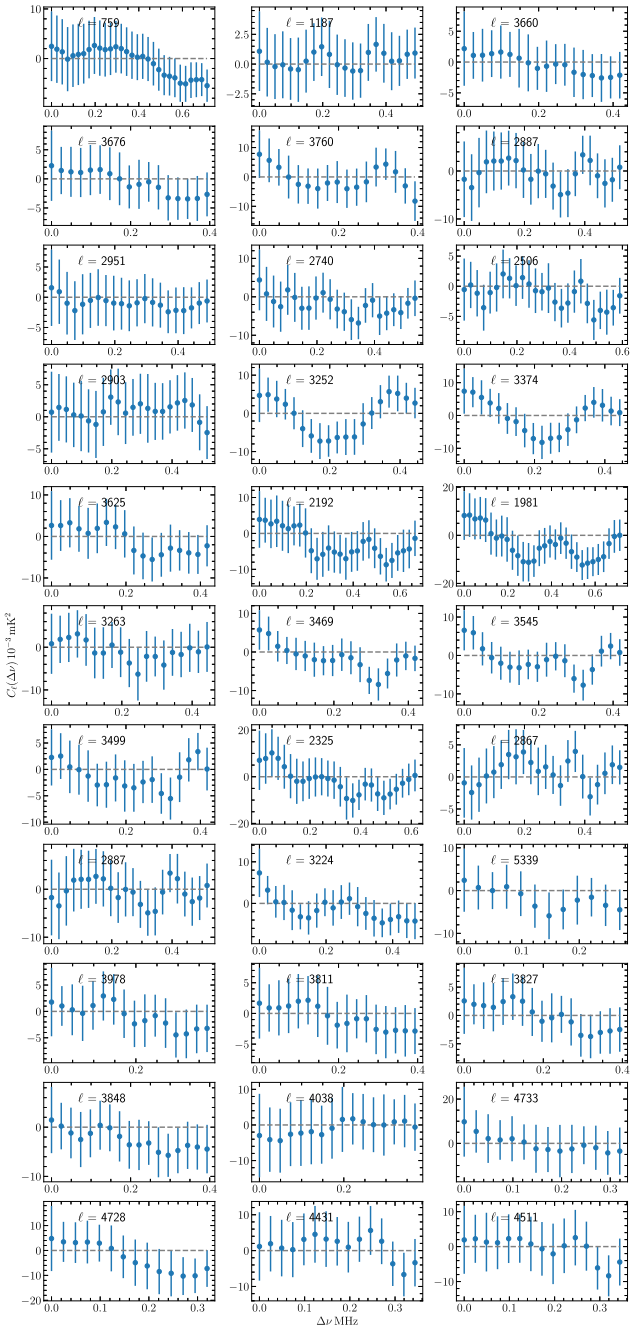

Figure A8. Same as Fig. A2 but for GPR (Fig. A7).

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