Draft version February 24, 2026
Typeset using LaTeX **twocolumn** style in AASTeX7.0.1

Foreground Mitigation and Power Spectrum Analysis for Tianlai Full-Sky 21 cm Survey Observation

Yikai Deng,[1,2] Shifan Zuo,[1,2] Jixia Li,[1,2] Yougang Wang[1,2,3] And Xuelei Chen[1,2,3]
The Tianlai collaboration

[1]*State Key Laboratory of Radio Astronomy and Technology, National Astronomical Observatories, CAS, A20 Datun Road, Chaoyang District, Beijing, 100101, P. R. China*
[2]*School of Astronomy and Space Science, University of Chinese Academy of Sciences, Beijing 100049, China*
[3]*Key Laboratory of Cosmology and Astrophysics (Liaoning) & College of Sciences, Northeastern University, Shenyang 110819, China*

ABSTRACT

We present a comprehensive analysis of the 21 cm intensity mapping (IM) data from the Tianlai Cylinder Pathfinder Array (TCPA), focusing on multi-scale foreground mitigation and three-dimensional power spectrum estimation. Utilizing 20 days of drift-scan observations (714.4–781.7 MHz, corresponding to HI emission at redshift $z \approx 0.82$–0.99), we reconstruct high-fidelity sky maps by incorporating a high-precision, drone-measured primary beam model. This in-situ calibration significantly enhances reconstruction accuracy over previous analytical approximations. To address astrophysical foregrounds, which dominate the cosmological signal by approximately five orders of magnitude, we implement a robust multi-scale subtraction strategy—mPCA-UWTS—that combines an isotropic Undecimated Wavelet Transform on the Sphere (UWTS) with independent Principal Component Analysis (PCA) within each wavelet domain. We subsequently estimate the 3D power spectrum via Spherical Fourier–Bessel (SFB) decomposition, providing a mathematically rigorous treatment of wide-angle and line-of-sight curvature effects inherent in wide-field surveys. Our analysis demonstrates that the SFB framework effectively isolates systematic contaminants and recovers the clustering signal without the biases introduced by conventional flat-sky approximations. This work represents the first application of the SFB formalism to observational 21 cm IM data, establishing it as a computationally efficient and scalable diagnostic tool for the next generation of wide-field 21 cm surveys, including the Square Kilometre Array (SKA) and the full Tianlai array.

Keywords: Cosmology (343) — H I line emission (690) — Dark energy (351) — Radio interferometers (1345)

1. INTRODUCTION

The 21 cm hyperfine transition of neutral hydrogen (H I) has emerged as a preeminent cosmological probe, offering a unique window into the large-scale structure (LSS) of the Universe across a vast redshift range (H. Padmanabhan 2024). Unlike traditional galaxy surveys that detect individual luminous objects, 21 cm intensity mapping (IM) measures the aggregate emission from multiple sources within a single voxel. This approach efficiently maps the three-dimensional density field over enormous volumes (A. Liu & J. R. Shaw 2020), making it particularly well-suited for investigating dark energy, the Epoch of Reionization (EoR), and the growth of cosmic structure.

The primary challenge in 21 cm IM is the presence of astrophysical foregrounds, which are typically four to five orders of magnitude more intense than the cosmological signal (T. J. Mozdzen et al. 2017; M. W. Eastwood et al. 2018; A. Liu & J. R. Shaw 2020). These foregrounds are dominated by diffuse Galactic synchrotron radiation, with additional contributions from free-free emission and extragalactic point sources (S. Cunnington et al. 2021). Fortunately, foregrounds exhibit spectral smoothness, which contrasts with the rapid frequency fluctuations of the 21 cm signal. This spectral contrast underpins various mitigation strategies, including parametric methods like polynomial fitting (X. Wang et al. 2006), Gaussian Pro-

Email: sfzuo@bao.ac.cn, xuelei@cosmology.bao.ac.cn

cess Regression (GPR; F. G. Mertens et al. 2018), and non-parametric blind signal separation techniques such as Principal Component Analysis (PCA; E. R. Switzer et al. 2013), Independent Component Analysis (FastICA; E. Chapman et al. 2012), and Generalized Morphological Component Analysis (GMCA) (I. P. Carucci et al. 2020). However, instrumental systematics—most notably frequency-dependent beam chromaticity—can "leak" foreground power into the cosmological window, significantly complicating the separation process (K. A. Glasscock et al. 2024; E. Chapman & V. Jelić 2019).

Observational 21 cm cosmology follows two complementary paths: global signal experiments and power spectrum interferometry. Global signal experiments, such as EDGES (J. D. Bowman et al. 2018) and SARAS (S. Singh et al. 2022), utilize single-element radiometers to target sky-averaged absorption features from the "Cosmic Dawn" and EoR. Conversely, interferometric arrays like the Low-Frequency Array (LOFAR; B. Greig et al. 2021), the Hydrogen Epoch of Reionization Array (HERA; HERA Collaboration et al. 2023), and the Tianlai Cylinder Pathfinder Array (X. Chen 2012) target the spatial fluctuations of 21 cm emission. By mapping the 3D power spectrum, these instruments aim to resolve LSS and the Baryon Acoustic Oscillation (BAO) scale, providing critical constraints on the expansion history and growth of structure.

Modern 21 cm interferometers often employ exceptionally wide fields of view (FoV) to achieve the survey volumes required for cosmological precision. Conventional analysis typically relies on uv-plane Fourier transforms, which are strictly valid only for narrow tracked fields where the flat-sky approximation holds. For wide-field drift-scan observations, the m-mode formalism (J. R. Shaw et al. 2014) uses spherical harmonic decomposition to naturally account for Earth's rotation and the all-sky nature of the data, providing a robust framework for sky map reconstruction from zenith-pointed arrays.

However, the wide-field nature of these surveys invalidates the traditional flat-sky approximation used in standard cosmological analyses. In the flat-sky limit, a single, fixed line-of-sight (LoS) is assumed for the entire survey volume—a simplification valid only for small patches of the sky. For wide-field and all-sky experiments, the LoS direction varies significantly across the survey area, leading to "wide-angle effects" that couple different physical scales and distort the recovered signal. Consequently, the conventional Cartesian Fourier power spectrum, $P(\mathbf{k})$, is susceptible to significant biases on large scales (B. Khek et al. 2024).

To overcome these geometric limitations, Spherical Fourier–Bessel (SFB) decomposition has emerged as a rigorous framework. Originally proposed in the 1990s (J. Binney & T. Quinn 1991; O. Lahav 1993) and applied to galaxy surveys (e.g., K. B. Fisher et al. 1995; A. F. Heavens & A. N. Taylor 1995), SFB provides an optimal basis for 3D fields on the sphere by combining spherical harmonics for angular distribution with spherical Bessel functions for the radial (redshift) dimension. While early adoption was limited by computational cost, recent optimizations (L. Samushia 2019) and efficient numerical codes (B. Leistedt et al. 2012; H. S. Grasshorn Gebhardt & O. Doré 2021) have made SFB analysis tractable. Studies have demonstrated the superiority of SFB over spherical harmonic tomography (A. Nicola et al. 2014; F. Lanusse et al. 2015) and its robustness in measuring large-scale cosmological effects (F. Semenzato et al. 2025). Nevertheless, most SFB research has focused on galaxy surveys or simulations, with very few applications to observational 21 cm IM data (A. Liu et al. 2016).

The Tianlai project is a pathfinder experiment designed to validate technologies for 21 cm IM during the post-reionization epoch (X. Chen 2012; Y. Xu et al. 2015; J. Li et al. 2020; F. Wu et al. 2021). Located at the radio-quiet Hongliuxia site in Xinjiang, China, the experiment comprises two co-located interferometric arrays: a cylinder pathfinder and a dish pathfinder. The Tianlai Cylinder Pathfinder Array (TCPA) consists of three fixed parabolic cylindrical reflectors, each 15 m (E-W) $\times$ 40 m (N-S), operating in drift-scan mode. Currently targeting the 685–810 MHz (lijixia: 700–800) band ($z \approx 0.77$–1.03), the TCPA utilizes 96 dual-linear polarization feeds in a $31+32+33$ configuration. Feeds are uniformly spaced along the central 12.4 m of the focal line to suppress grating lobes while maximizing sensitivity (J. Zhang et al. 2016b). The primary goal is to map the LSS and constrain cosmological parameters, particularly the dark energy equation of state.

In this paper, we analyze a 20-day observational dataset from the TCPA. We introduce an analysis pipeline that integrates a high-fidelity primary beam model derived from in-situ drone measurements with a multi-scale foreground mitigation strategy. We use an isotropic Undecimated Wavelet Transform on the Sphere (UWTS) combined with scale-dependent PCA for high-fidelity foreground removal. We then apply SFB decomposition to extract the 3D power spectrum, accounting for wide-angle and LoS curvature effects. This work demonstrates the viability of the SFB framework for extracting cosmological signals from real-world, wide-field interferometric datasets. The paper is organized as follows: Section 2 details data reduction, beam modeling, and foreground mitigation. Section 3 presents the SFB

power spectrum results, followed by a discussion in Section 4. We conclude with a summary in Section 5.

2. DATA REDUCTION AND ANALYSIS

2.1. *Observations and Data Reduction*

We analyze 20 days of drift-scan observations from the TCPA, spanning two epochs in 2018: January 21–February 3 (14 days) and March 22–27 (6 days). The processing pipeline begins with raw visibilities. Due to storage and throughput constraints during commissioning, the analyzed bandwidth is restricted to 712.9–783.1 MHz, divided into 576 channels (0.122 MHz width). Raw data underwent automated RFI flagging and calibration using Cygnus A, supplemented by an on-site artificial calibration noise source (CNS) for phase stabilization. Nightly mean subtraction and coupling corrections were applied to the calibrated visibilities. During map-making, bright celestial sources—Cassiopeia A, Cygnus A, and Centaurus A—were modeled and subtracted. Additionally, approximately 30 minutes of data around the Sun's daily transit were excised to mitigate sidelobe contamination. The m-mode map-making framework was presented in J. R. Shaw et al. (2014); J. Zhang et al. (2016a), the detailed pipeline and calibration strategy and are documented in S. Zuo et al. (2021), S. Zuo et al. (2019) respectively, and the error in the calibration were also analyzed by simulation in K. Yu et al. (2023). We reconstruct independent sky maps for the XX and YY polarizations to account for their distinct primary beam responses.

Previous studies utilized the analytical beam model in `tlpipe` (S. Zuo et al. 2021), which assumes a separable profile in the east–west (EW) and north–south (NS) directions (J. R. Shaw et al. 2014). However, simulations (S. Sun et al. 2022) and measurements with a drone (J. Li et al. 2025) show that the actual beam response deviates significantly from this idealized form. To enhance reconstruction accuracy, we incorporate a high-precision beam profile for the A26 feed, derived from drone-based far-field measurements (J. Li et al. 2025). This model covers the 693.24–801.64 MHz range with 0.98 MHz resolution.

Drone operational constraints limited the measurements to zenith angle ranges of $|\theta_{\rm NS}| < 70^\circ$ and $|\theta_{\rm EW}| < 3^\circ$. To construct a comprehensive all-sky model, we extend these to $\pm 90^\circ$ in both directions across all frequencies. In the measured EW domain, we use cubic spline interpolation for fine-scale structures. For unmeasured regions, a piecewise Gaussian model accounts for beam asymmetry; the two sides of the EW profile are fitted independently, with scaling factors applied to the Gaussian tails for continuity. For the NS direction, the sparsity and irregularity of the data necessitate a single Gaussian fit over the $|\theta_{\rm NS}| < 70^\circ$ range.

Figure 1 shows the beam profile along the East-West and North-South directions as measured by the drone and fitting models for a representative frequency. The full frequency-dependent profile is shown in Figure 2. Not very clear: are these measurement result from the drone, or based on some model? To align these with the 576 frequency channels of our visibility data, we implement a multi-stage interpolation. First, we shift the peaks of the fitted profiles to a common zenith angle to avoid interpolation artifacts from minor frequency-dependent offsets in the raw data. These offsets are then re-applied to the interpolated profiles (Figure 3). The final EW and NS profiles are projected onto the celestial sphere using the HEALPix (K. M. Górski et al. 2005) Mollview projection (Figure 4). Assuming beam separability, our 2D primary beam model is constructed as the product of these orthogonal 1D profiles.

Figure 5 (top panels) shows the reconstructed sky maps for the XX (left) and YY (right) polarizations at 748.11 MHz. The maps are dominated by Galactic diffuse emission and bright extragalactic point sources. We identify a prominent "ghost" artifact—an aliased image of the Galactic plane at the upper center-right of the figure, resulting from incomplete angular sampling. Additionally, two parabolic arcs from solar sidelobes are visible near the center of the figure, caused by the Sun's shift between the two 2018 epochs. Despite excising 30 minutes of peak transit data, residual solar power remains detectable. Failure to remove the peak solar signal would cause severe contamination across the map via the instrument's Point Spread Function (PSF).

The Maps reconstructed with this drone-measured beam exhibit improved fidelity over those using the default analytic model. Given the array's geographic latitude (44°09′N), sensitivity drops precipitously south of $\delta \approx -15^\circ$. Consequently, maps become noise-dominated at colatitudes $\theta > 105^\circ$. We restrict subsequent cosmological analyses to the colatitude range $[0^\circ, 105^\circ]$. Additionally, frequency channels at both ends of the band are discarded to mitigate edge effects and band-pass instabilities. The final science-ready data cube comprises $N_\nu = 552$ channels spanning 714.42–781.68 MHz with $N_{\rm side} = 512$ resolution. Note that in this figure the a_{00} spherical harmonic component (global mean temperature $\bar{T}$) is filtered out to suppress RFI and common-mode pickup, resulting in zero-mean maps.

We also reconstruct maps using only cross-cylinder baselines (Figure 5, bottom panels). These longer baselines (≥ 15.0 m) are less susceptible to crosstalk. While the Galactic plane remains visible, its large-scale struc-

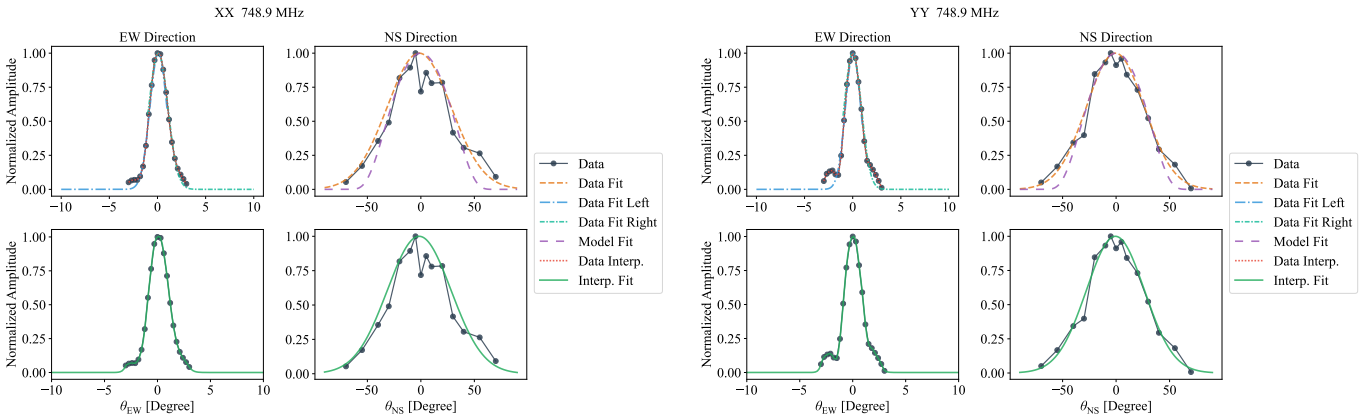

Figure 1. One-dimensional primary beam response for the A26 feed as a function of zenith angle at a representative frequency(748.9 MHz) for the XX (left panels) and YY (right panels) polarizations. Drone-based measurements (markers) are compared with three modeling approaches: a Gaussian fit, the default analytical model from `tlpipe`, and cubic spline interpolation. The upper panels illustrate the fitting performance within the measured domain, while the lower panels show the extended all-sky beam profiles. There is no difference in the the range of the upper and lower panels? Also, the meaning of the different curves in the legend is not clear

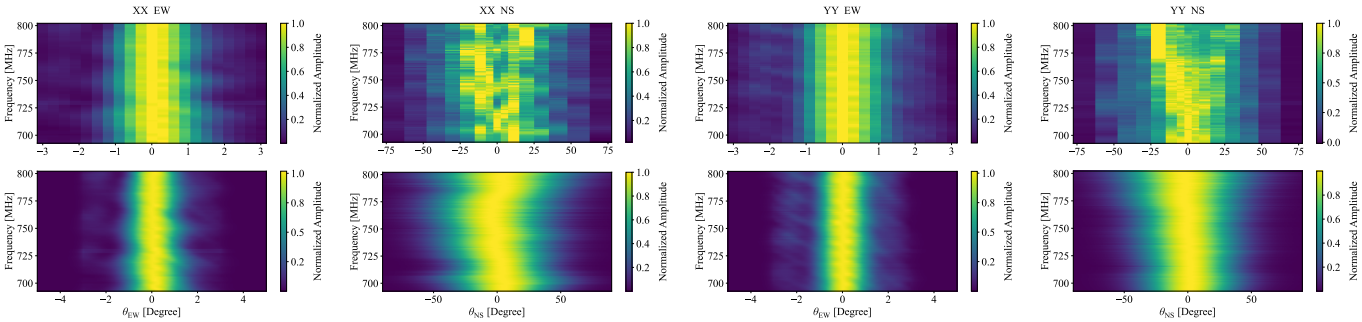

Figure 2. Frequency-dependent evolution of the drone-measured beam response. The top row displays the raw far-field measurements across the observed band. The bottom row shows the resulting models: cubic spline interpolation for the EW direction and Gaussian fits for the NS direction. But the top and bottom panels do not look similar?

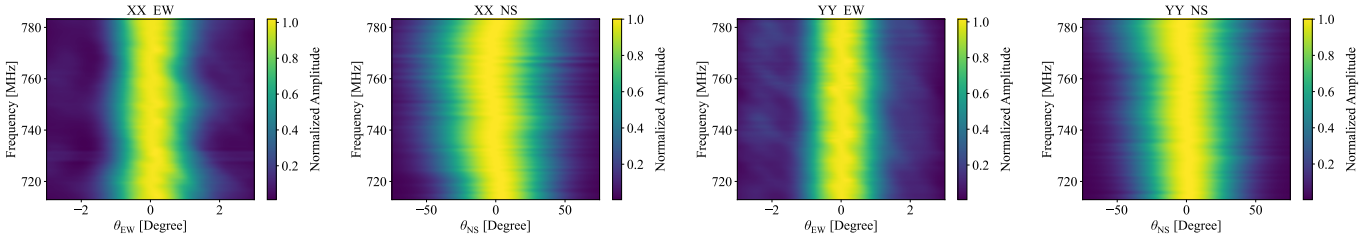

Figure 3. Interpolation of the sparse drone-measured beam profiles to the full frequency resolution (576 channels) of the visibility data. This ensures a consistent beam model for all frequency channels used in the map-making process.

ture is attenuated by the lack of short baselines (the "missing-zero" problem). Notably, systematic artifacts like the Galactic ghost and solar sidelobes are markedly suppressed in these maps, though not entirely eliminated.

For comparison, we have also generated a simulated 21 cm signal using `cora` (J. R. Shaw et al. 2025) within the same frequency band, applying a declination cut at $\delta = -15°$ to match the Tianlai observational footprint.

2.2. *Foreground Subtraction*

The reconstructed sky maps are dominated by intense diffuse foreground emission, which must be accurately characterized and removed to extract the faint cosmological 21 cm signal.

Conventional foreground mitigation techniques, such as polynomial fitting (X. Wang et al. 2006), Principal Component Analysis (PCA; E. R. Switzer et al. 2013), Fast Independent Component Analysis (FastICA;

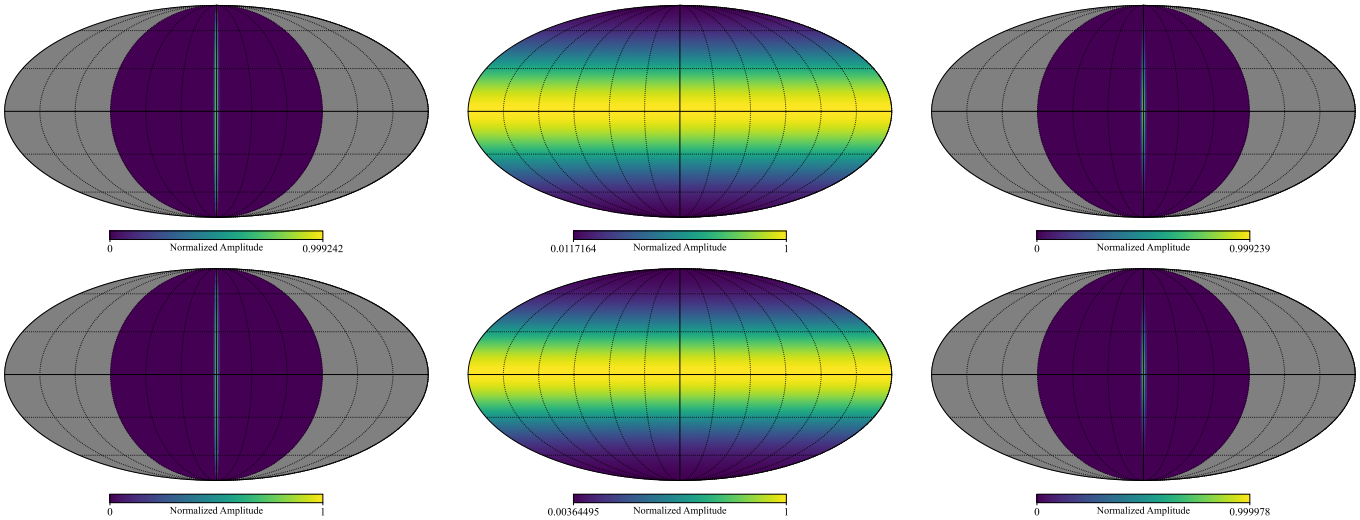

Figure 4. Two-dimensional primary beam model projected onto the celestial sphere. The EW (left) and NS (middle) profiles are displayed in HEALPix Mollview projection at the central frequency (748.11 MHz). The final 2D model (right), constructed assuming separability, is used for wide-field map reconstruction. Top and bottom rows correspond to the XX and YY polarizations, respectively.

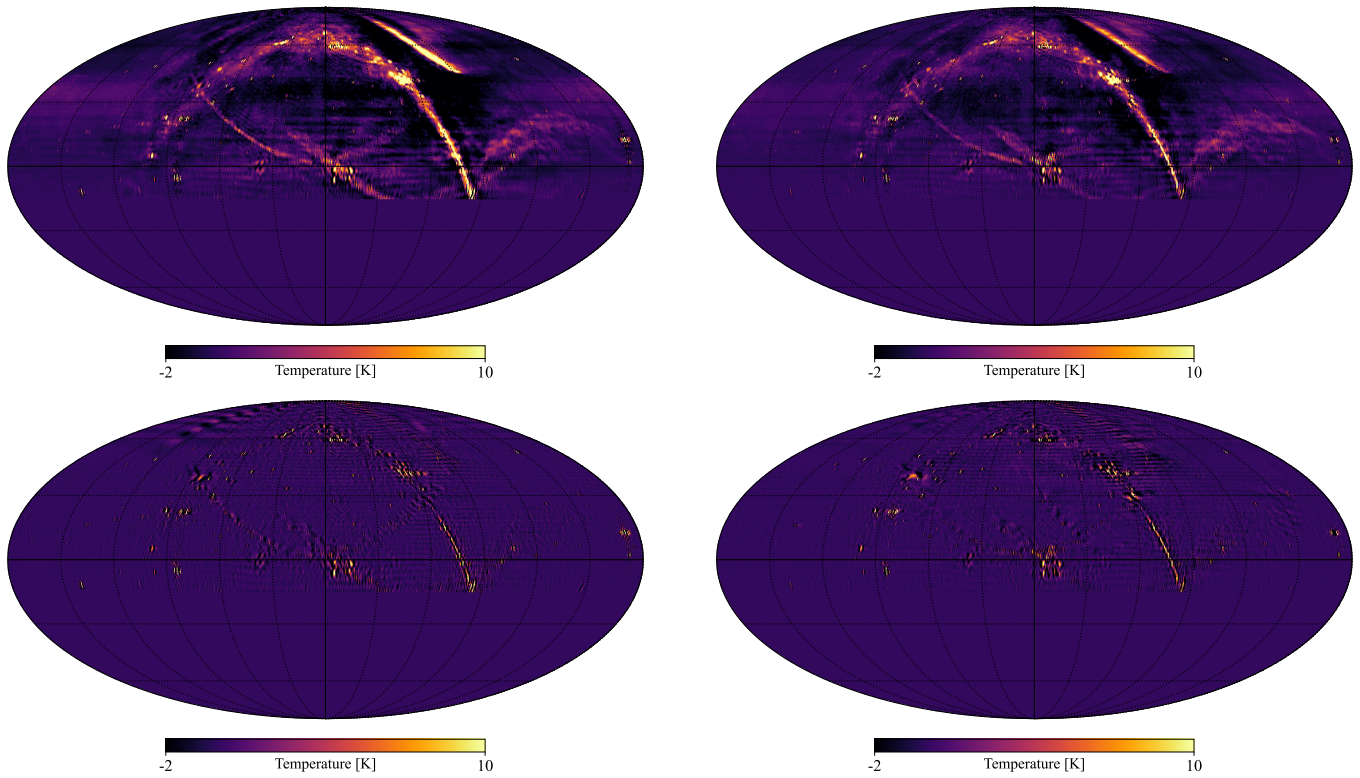

Figure 5. Reconstructed sky maps for the XX (left) and YY (right) polarizations at 748.11 MHz. The top row shows maps using the all-baseline set, while the bottom row utilizes only the cross-cylinder baseline subset. The maps are dominated by Galactic diffuse emission and systematic artifacts (e.g., solar sidelobes and Galactic ghosts). The temperature scale is restricted to $[-2, 10]$ K to highlight faint structures.

E. Chapman et al. 2012), and Gaussian Process Regression (GPR; F. G. Mertens et al. 2018; P. S. Soares et al. 2022), primarily exploit the spectral smoothness of astrophysical foregrounds (S. Cunnington et al. 2021; E. Chapman & V. Jelić 2019). However, incorporating spatial (angular) information is essential, as instrumental systematics and astrophysical contaminants often exhibit distinct morphological characteristics across different spatial scales. For instance, while large-scale features are dominated by synchrotron and free-free emission, small-scale fluctuations are typically more sensitive to instrumental noise and localized systematics.

Drawing upon the multi-scale PCA (mPCA) algorithm proposed by the MeerKLASS Collaboration (I. P. Carucci et al. 2025), we implement a tailored approach for wide-field spherical maps. Within the mPCA framework, each frequency map is first decomposed into multiple wavelet-filtered components representing distinct spatial scales via the isotropic undecimated wavelet transform (IUWT, or starlet transform) (J.-L. Starck et al. 2007). A temperature map $X(\nu, p)$ at frequency ν pixel p? is expressed as the sum of a coarse-scale component $X_{\rm L}(\nu, p)$ and a set of wavelet scales $W_j(\nu, p)$:

$$X(\nu, p) = X_{\rm L}(\nu, p) + \sum_{j=1}^{j_{\rm max}} W_j(\nu, p). \quad (1)$$

The wavelet coefficients W_j capture features at dyadic scales 2^j, while $X_{\rm L}$ represents the remaining large-scale information. I am unfamiliar with this, does this mean $j = 1$ corresponding to 1 pixel, $j_{\rm max}$ corresponding to $2^{j_{\rm max}}$ pixels? In contrast to the original MeerKLASS implementation, which utilized a single wavelet scale for small-area flat-sky maps, our analysis adopts a multi-scale decomposition to accommodate the wide-field geometry and complex systematic environment of the TCPA dataset.

For spherical datasets, we utilize the isotropic Undecimated Wavelet Transform on the Sphere (UWTS) (J.-L. Starck et al. 2006). The UWTS retains the key advantages of the flat-sky starlet transform—including exact reconstruction, isotropy, and compact support—while operating directly in the spherical harmonic domain. The decomposition is performed using a scaling function $\phi_{l_c}(\theta, \varphi)$ characterized by azimuthal symmetry and a cut-off multipole l_c. This is a bit abstract, maybe could add a little more details with formula?

In this work, we set $l_c = 1024$, matching the maximum angular resolution accessible to the TCPA ($\ell \sim 750$). This yields $j_{\rm max} = 10$, resulting in 10 wavelet scales and one coarse component for each sky map. We then apply PCA to each scale independently. By organizing the data at each scale into a 2D matrix $\mathbf{T}$ of dimension $N_\nu \times N_{\rm pix}$, we perform Singular Value Decomposition (SVD):

$$\mathbf{T} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\rm T}, \quad (2)$$

where the columns of $\mathbf{U}$ represent the principal spectral modes and $\mathbf{\Sigma}$ denotes the singular values s_i. The eigenvalues of the frequency covariance matrix $\mathbf{C} = \frac{\mathbf{T}\mathbf{T}^{\rm T}}{N_{\rm pix}-1}$ are given by $\lambda_i = s_i^2$.

The foreground-cleaned residual for each scale is obtained by subtracting the $N_{\rm FG}$ dominant modes:

$$\mathbf{T}_{\rm res} = \mathbf{T} - \sum_{i=1}^{N_{\rm FG}} s_i \mathbf{u}_i \mathbf{v}_i^{\rm T}, \quad (3)$$

where $N_{\rm FG}$ is determined independently for each scale to optimize the balance between foreground suppression and signal preservation. The final cleaned map is reconstructed by summing the residuals across all scales:

$$T_{\rm clean}(\hat{\boldsymbol{r}}, \nu) = c_{j_{\rm max},{\rm res}}(\hat{\boldsymbol{r}}, \nu) + \sum_{j=1}^{j_{\rm max}} w_{j,{\rm res}}(\hat{\boldsymbol{r}}, \nu). \quad (4)$$

The exact reconstruction property of the UWTS ensures that the decomposition-recomposition process introduces no numerical artifacts.

A key advantage of mPCA over conventional PCA is the ability to adapt the number of removed components to the specific noise and foreground properties of each spatial scale. However, this flexibility necessitates a robust criterion for selecting $N_{\rm FG}$. In this study, we utilize the *fractional variance explained* as the primary metric for determining the optimal $N_{\rm FG}$ at each scale.

Figure 6 illustrates the mean frequency spectra for the XX and YY polarizations. The YY polarization exhibits a markedly smoother spectral profile compared to XX, rendering it more amenable to foreground mitigation. Given that the cosmological 21 cm signal is expected to be unpolarized, we focus our subsequent power spectrum analysis on the YY polarization maps.

2.3. *Spherically Averaged Power Spectrum Estimation*

The spherically averaged power spectrum, $P(k)$, quantifies the statistical variance of 21 cm brightness temperature fluctuations across different spatial scales. It serves as a fundamental observable for extracting cosmological and astrophysical constraints from intensity mapping data. In the flat-sky approximation—commonly applied to narrow-field observations—the power spectrum is typically decomposed into components parallel ($k_\parallel$) and perpendicular ($k_\perp$) to a fixed line-of-sight (LoS) (A. Liu et al. 2016). Assuming statistical isotropy, the power spectrum depends solely on the magnitude of the 3D wavevector $\boldsymbol{k}$, where $k = \sqrt{k_\perp^2 + k_\parallel^2}$.

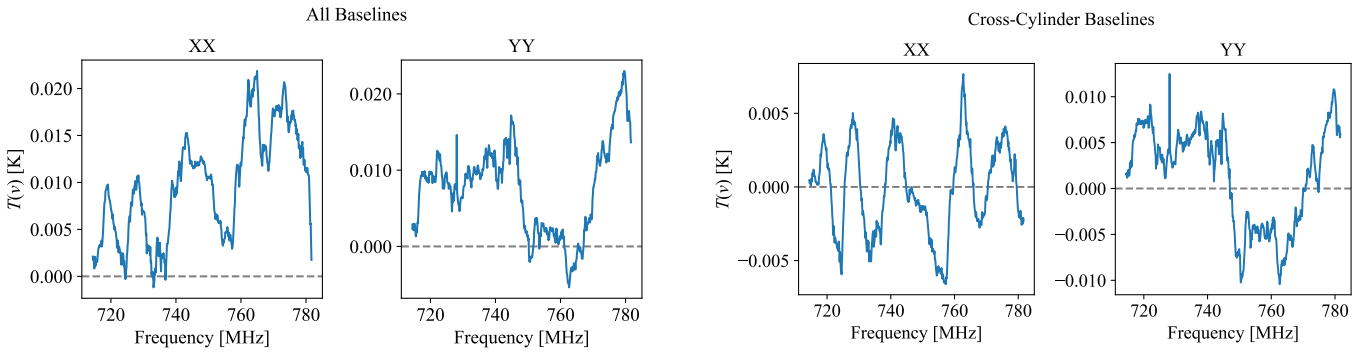

Figure 6. Mean frequency spectra for the reconstructed XX and YY polarization sky maps. The left panel shows the results using the all-baseline set, while the right panel corresponds to the cross-cylinder baseline subset.

Standard Fourier-based estimators, which involve a 3D Fast Fourier Transform (FFT) of a Cartesian image cube, are not directly applicable to our data given the wide-field coverage and the spherical shell geometry spanning the redshift range $z \approx 0.82$–0.99. Instead, we adopt the methodology proposed by K. K. Datta et al. (2007) and S. Pal et al. (2022) to estimate the spherically averaged $P(k)$. We first compute the multi-frequency angular power spectrum (MAPS), $C_\ell(\nu_a, \nu_b)$, defined as:

$$C_\ell(\nu_a, \nu_b) = \langle a_{\ell m}(\nu_a) a^*_{\ell m}(\nu_b) \rangle. \tag{5}$$

Assuming statistical homogeneity over the narrow relative bandwidth ($\Delta\nu \ll \nu_c$), the MAPS depends primarily on the frequency separation $\Delta\nu = |\nu_b - \nu_a|$. We average the C_ℓ values for frequency pairs with consistent $\Delta\nu$ to obtain $C_\ell(\Delta\nu)$. The cylindrically binned power spectrum $P(k_\perp, k_\parallel)$ is then derived via a Fourier transform of $C_\ell(\Delta\nu)$ along the frequency axis:

$$P(k_\perp, k_\parallel) = r^2 r' \int d(\Delta\nu) e^{-i k_\parallel r' \Delta\nu} C_\ell(\Delta\nu), \tag{6}$$

where $k_\perp = \ell / r$. At the reference frequency $\nu_c = 748.05\,\mathrm{MHz}$ Earlier 748.9 MHz was used?, the comoving distance r and its derivative $r' = dr/d\nu$ are $2151.15\,h^{-1}\,\mathrm{Mpc}$ and $4.66\,h^{-1}\,\mathrm{Mpc\,MHz^{-1}}$, respectively, calculated using the WMAP9 cosmology (G. Hinshaw et al. 2013). Finally, the 1D power spectrum $P(k)$ is obtained by spherically averaging the 2D $P(k_\perp, k_\parallel)$ into 20 k-bins with a uniform logarithmic width of $\Delta \log_{10} k \approx 0.1$, spanning the range $k \in [0.002, 0.2]\,h\,\mathrm{Mpc}^{-1}$.

2.4. *Spherical Fourier–Bessel Power Spectrum*

In observational cosmology, spherical coordinates—centered on the observer—provide the most natural basis for analyzing large-scale structure (LSS) fields. In this geometry, the eigenfunctions of the Laplacian operator are the product of spherical Bessel functions and spherical harmonics, $j_\ell(kr)Y_{\ell m}(\hat{\boldsymbol{r}})$, with corresponding eigenvalues $-k^2$. This basis framework intrinsically accounts for sky curvature and the expanding survey volume, avoiding the wide-angle biases inherent in flat-sky approximations. For a statistically homogeneous 3D field $\delta(\boldsymbol{r})$ in a flat universe, the spherical Fourier–Bessel (SFB) decomposition (K. B. Fisher et al. 1995; A. Heavens 2003; P. G. Castro et al. 2005) is defined as:

$$\delta(\boldsymbol{r}) = \int dk \sum_{\ell m} \left[\sqrt{\frac{2}{\pi}} k j_\ell(kr) Y_{\ell m}(\hat{\boldsymbol{r}}) \right] \tilde{\delta}_{\ell m}(k), \tag{7}$$

with the inverse relation:

$$\tilde{\delta}_{\ell m}(k) = \int d^3 r \left[\sqrt{\frac{2}{\pi}} k j_\ell(kr) Y^*_{\ell m}(\hat{\boldsymbol{r}}) \right] \delta(\boldsymbol{r}). \tag{8}$$

The SFB power spectrum $C_\ell(k)$ is defined via the covariance:

$$\langle \tilde{\delta}_{\ell m}(k) \tilde{\delta}^*_{\ell' m'}(k') \rangle = C_\ell(k, k') \delta_{\ell\ell'} \delta_{mm'} \tag{9}$$

$$= C_\ell(k) \delta^D(k - k') \delta_{\ell\ell'} \delta_{mm'}. \tag{10}$$

Practical surveys are confined to finite volumes, in which case the radial modes have a discrete set of values $k_{n\ell}$. For a survey volume defined by a spherical shell with boundaries $r_{\min}$ and $r_{\max}$, the SFB decomposition of the temperature field $T(\boldsymbol{r})$ is:

$$T_{n\ell m} = \int d^2\hat{\boldsymbol{r}}\, Y^*_{\ell m}(\hat{\boldsymbol{r}}) \int_{r_{\min}}^{r_{\max}} dr\, r^2 g_{n\ell}(r) T(\boldsymbol{r}), \tag{11}$$

where the radial basis functions $g_{n\ell}(r)$ are linear combinations of spherical Bessel functions of the first (j_ℓ) and second (y_ℓ) kind:

$$g_{n\ell}(r) = c_{n\ell} j_\ell(k_{n\ell} r) + d_{n\ell} y_\ell(k_{n\ell} r). \tag{12}$$

how is the k_{nl} is defined here? These functions satisfy the orthonormality relation:

$$\int_{r_{\min}}^{r_{\max}} dr\, r^2 g_{n\ell}(r) g_{n'\ell}(r) = \delta_{nn'}. \tag{13}$$

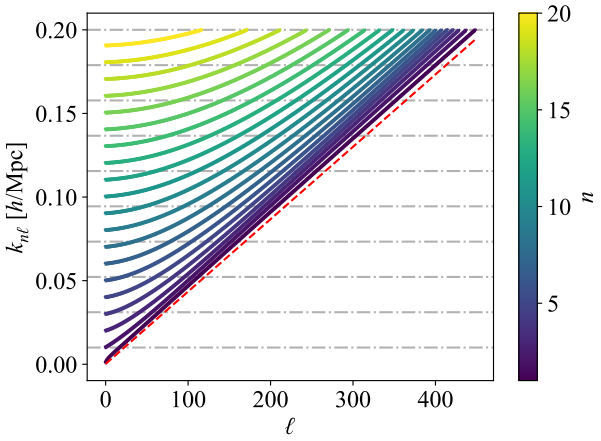

Figure 7. Distribution of the discrete SFB modes $k_{n\ell}$ as a function of angular multipole ℓ and radial index n, calculated for the redshift range $z \in [0.8, 1.0]$. The dashed red line indicates the boundary formed by the $n = 1$ modes, representing the largest radial scales accessible to the survey. This boundary is well-approximated by the geometric relation $k \approx (\ell + 1/2)/r_{\rm max}$.

The coefficients $c_{n\ell}$ and $d_{n\ell}$ are determined by the boundary conditions at $r_{\rm min}$ and $r_{\rm max}$. In this analysis, we adopt the potential boundary condition what is this?. The resulting discrete radial eigenvalues $k_{n\ell}$ are coupled to both the radial index n and the angular multipole ℓ, as illustrated in Figure 7 It is unclear how is obtained?. For the TCPA frequency band, the comoving boundaries are $r_{\rm min} = 1997.83\,h^{-1}\,$Mpc and $r_{\rm max} = 2310.81\,h^{-1}\,$Mpc.

The discrete modes $k_{n\ell}$ represent the magnitude of the wavevector $\boldsymbol{k}$, analogous to k in the Cartesian Fourier basis. The angular transform is performed via the HEALPix framework, while the radial integration is implemented using the trapezoidal rule. The observed (pseudo) SFB power spectrum, $\hat{C}_{\ell nn'}$, is then estimated as:

$$\hat{C}_{\ell nn'} = \frac{1}{2\ell+1}\sum_m T_{n\ell m} T^*_{n'\ell m}. \tag{14}$$

Assuming statistical isotropy, we average over the m-modes and focus on the diagonal components ($n = n'$). The pseudo-power spectrum extends the standard angular power spectrum C_ℓ to three dimensions.

To mitigate the effects of an incomplete survey volume, window function deconvolution is required to recover an unbiased estimate of the underlying power spectrum. Under the assumption of a separable observational window function $W(\boldsymbol{r}) = \phi(r)M(\hat{\boldsymbol{r}})$, where $\phi(r)$ and $M(\hat{\boldsymbol{r}})$ represent the radial and angular selection functions respectively, the observed SFB power spectrum $C^{\rm obs}_{\ell nn'}$ is related to the true spectrum $C_{LNN'}$ via a mixing matrix $\mathcal{M}$ (H. S. Grasshorn Gebhardt & O. Doré 2021):

$$C^{\rm obs}_{\ell nn'} = \sum_{LNN'} \mathcal{M}^{LNN'}_{\ell nn'} C_{LNN'}, \tag{15}$$

where the mixing matrix is defined as:

$$\begin{aligned}\mathcal{M}^{LNN'}_{\ell nn'} = &\frac{2L+1}{4\pi}\sum_{L_1}\begin{pmatrix}\ell & L & L_1\\ 0 & 0 & 0\end{pmatrix}^2 \sum_{M_1}|W_{L_1M_1}|^2 \\ &\times \int dr\, r^2 g_{n\ell}(r) g_{NL}(r)\phi(r) \\ &\times \int dr'\, r'^2 g_{n'\ell}(r') g_{N'L}(r')\phi(r').\end{aligned} \tag{16}$$

Here, $W_{L_1M_1}$ are the spherical harmonic coefficients of the angular mask $M(\hat{\boldsymbol{r}})$, and the term in parentheses is the Wigner $3j$ symbol. For our survey, we set $\phi(r) = 1$ and $M(\hat{\boldsymbol{r}}) = 1$ within the colatitude range $\theta < 105°$, and zero otherwise. Deconvolution aims to retrieve the true spectrum $C_{LNN'}$ from the observed $C^{\rm obs}_{\ell nn'}$, though this inversion is often numerically ill-conditioned.

To suppress the variance of poorly constrained modes and compress the data, we group the SFB pseudo-power spectrum into bandpowers. The binning is performed over both ℓ and n modes to obtain the bandpower-binned SFB power spectrum:

$$B^{\rm obs}_{LNN'} = \sum_{\ell nn'} \widetilde{w}^{\ell nn'}_{LNN'} C^{\rm obs}_{\ell nn'}, \tag{17}$$

where the weights $\widetilde{w}$ average neighboring modes $(\ell nn') \sim (LNN')$ and are normalized such that $\sum_{\ell nn'} \widetilde{w}^{\ell nn'}_{LNN'} = 1$. In matrix notation, the relationship between the observed bandpowers $B^{\rm obs}$ and the underlying spectrum C is:

$$B^{\rm obs} = \widetilde{w}\mathcal{M}C = \mathcal{N}B, \tag{18}$$

where $\mathcal{N} = \widetilde{w}\mathcal{M}v$ is the bandpower mixing matrix and v is the Moore–Penrose inverse of $\widetilde{w}$. The bin widths $\Delta\ell$ and Δn are estimated based on the survey sky fraction $f_{\rm sky}$ and volume fraction $f_{\rm vol}$:

$$\Delta\ell \approx \frac{1}{f_{\rm sky}}, \quad \Delta n \approx \frac{f_{\rm sky}}{f_{\rm vol}}. \tag{19}$$

For our survey geometry, this yields bin widths of $\Delta n = 1$ and $\Delta\ell \in \{1, 2\}$, depending on the target angular resolution.

In the limit of statistical homogeneity and isotropy, the SFB power spectrum is related to the traditional 3D power spectrum $P(k)$ by:

$$C_\ell(k, k') = \delta^D(k - k')P(k). \tag{20}$$

Incorporating observational effects—including the radial selection function $\phi(r)$, the linear growth factor $D(r)$, scale-dependent bias $b(k,r)$, and redshift-space distortions (RSD)—the SFB power spectrum can be modeled via the Limber approximation (H. S. Grasshorn Gebhardt & O. Doré 2021):

$$C_\ell(k,k') \approx P(k) e^{-\sigma_u^2 k^2} \delta^D(k-k') \times [\phi(r) D(r) b(k,r) \mathcal{F}_{\rm RSD}(\ell,k,r)]^2_{r=\frac{\ell+1/2}{k}} , \quad (21)$$

where $e^{-\sigma_u^2 k^2}$ accounts for the Gaussian Finger-of-God (FoG) effect with σ_u being the pairwise velocity dispersion in units of length, and $\mathcal{F}_{\rm RSD}$ is the RSD correction factor. This highlights the capacity of the SFB framework to encapsulate complex cosmological and astrophysical signatures—including FoG effects, redshift evolution, and the Kaiser effect—without necessitating the flat-sky approximation.

In the plane-parallel limit, the SFB power spectrum can be approximately mapped to the clustering wedge $P(k,\mu)$ (R. Y. Wen et al. 2025):

$$C_{\ell nn} \approx P\left(k = k_{n\ell}, \mu = \frac{k_{\parallel,n\ell}}{k_{n\ell}}, r = r_{{\rm eff},n\ell}\right), \quad (22)$$

where the effective distance $r_{{\rm eff},n\ell}$ is defined for each (n,ℓ) mode based on the radial selection function and survey boundaries. This mapping allows the SFB analysis to leverage theoretical frameworks developed for Cartesian clustering wedges while maintaining the advantages of a spherical basis.

A primary advantage of the SFB representation is its inherent ability to localize systematic effects. For instance, foreground residuals—characterized by smooth radial profiles—are predominantly confined to the lowest radial modes ($n = 1$). This localization enables the selective mitigation of contaminated modes while preserving the cosmological signal at other scales, a feature we exploit in our subsequent analysis.

3. RESULTS

3.1. *UWTS Foreground Mitigation*

To analyze the observational data from the Tianlai Cylinder Pathfinder Array, we first mitigate the dominant astrophysical foregrounds. We employ the mPCA method described in Sec. 2.2, which involves decomposing the maps into 10 wavelet scales and one coarse component using the UWTS algorithm, followed by independent PCA cleaning of each component. The number of removed foreground modes, $N_{\rm FG}$, is determined based on the fractional variance explained

$$R = \frac{\sum_{i=1}^{N_{\rm FG}} \lambda_i}{\sum_{i=1}^{N_\nu} \lambda_i}, \quad (23)$$

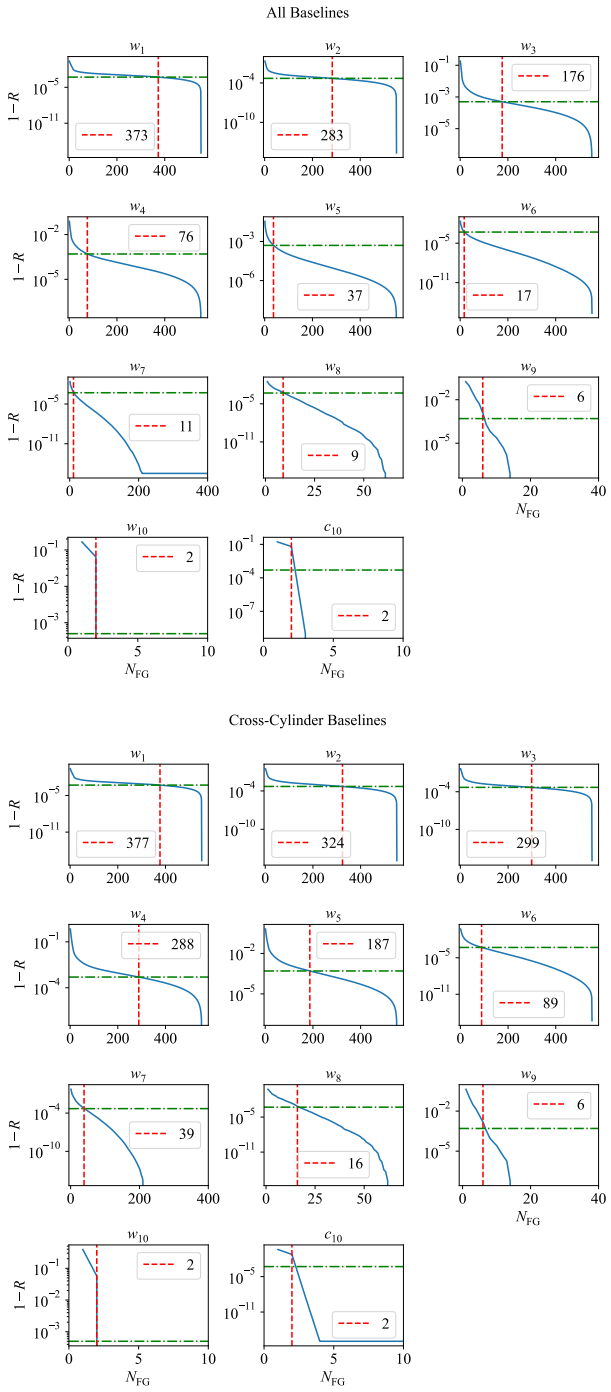

Figure 8. Fractional retained variance $(1-R)$ as a function of PCA mode index for the all-baseline set (Top) and the cross-cylinder baseline subset (Bottom). The horizontal dashed-dotted line indicates our chosen cleaning threshold of 0.0005. The vertical dashed line marks the resulting number of modes ($N_{\rm FG}$) subtracted at each UWTS scale, balancing foreground removal against potential signal loss.

where $\lambda_i$denotes the i-th eigenvalue of the PCA decomposition.

For a given threshold R, the required number of modes $N_{\rm FG}$ varies between the all-baseline set and the cross-cylinder baseline subset. To illustrate this dependence, Figure 8 shows the retained variance $(1-R)$ as a function of the PCA mode index for both baseline configurations. We adopt a threshold of $1-R=0.0005$, indicated by the green dashed-dotted horizontal lines, with the corresponding $N_{\rm FG}$ marked by red dashed vertical lines. This choice is motivated by the fact that foregrounds are approximately 4–5 orders of magnitude brighter than the expected 21 cm signal; a value of 0.0005 provides a conservative balance, avoiding the potential over-subtraction associated with more aggressive thresholds (e.g., 10^{-5}).

The eigenvalue spectra exhibit a strong scale dependence. For large-scale components (w_7–c_{10}), the fractional variance explained drops precipitously within the first few modes before reaching a stable plateau. This indicates that the foreground power at these scales is highly concentrated in a small number of dominant spectral modes. In contrast, for smaller-scale components (w_1–w_6), the decline in variance is more gradual, necessitating the removal of a larger number of modes to achieve equivalent suppression. Notably, for the cross-cylinder baseline subset, the required $N_{\rm FG}$ to reach the same $1-R$ threshold is generally higher than for the all-baseline set. This likely reflects a more complex foreground residual structure or a different noise floor in the cross-cylinder baseline subset.

Figure 9 displays the 11 UWTS scale maps generated using the all-baseline set at the central frequency of 748.11 MHz Here the reference frequency changed again?. The maps reveal distinct foreground morphologies across scales: small and medium scales (w_1–w_5) are dominated by granular structures and point-source-like features, while larger scales (w_6, w_7) show more diffuse contours. At the coarsest scales (w_8–c_{10}), the emission transitions into smooth, large-scale asymmetries. The map amplitudes exhibit a systematic decrease towards coarser scales, reflecting the concentration of fluctuation power at higher multipoles. The c_{10} component, which corresponds to the angular monopole or global mean ($\ell \approx 0$), maintains a nearly constant value across the sky and exhibits minimal frequency dependence.

Figure 10 presents the residual maps for each of the 11 UWTS scales following independent PCA cleaning using the all-baseline set. The spatial features systematically transition from fine-grained textures at small scales (w_1–w_3) to intermediate patches (w_4–w_6), and finally to diffuse asymmetries at the coarsest scales (w_7–c_{10}). Compared to the original maps (Figure 9), the dominant foreground structures are significantly suppressed, revealing a more irregular morphology with substantially reduced amplitudes.

We note that the residuals in the w_1 to w_3 scales exhibit non-uniform features along the longitudinal direction and near the central region, likely reflecting the point spread function (PSF) response and residual point-source contamination. At intermediate scales (w_4–w_6), a marked increase in variance is observed in sky regions corresponding to daytime observations compared to nighttime data. This underscores the difficulty of removing complex, time-varying artifacts induced by solar transits during the day.

The reconstructed foreground-subtracted maps for both the all-baseline set and the cross-cylinder baseline subset are presented in Figure 11. Comparing these with the raw sky maps (Figure 5) highlights a dramatic suppression of astrophysical foregrounds. In the all-baseline reconstruction, intense Galactic plane emission, bright point sources, and extended diffuse structures are effectively mitigated. The resulting maps exhibit a predominantly stochastic texture consistent with instrumental noise and low-level residual systematics, with only minimal large-scale coherent emission remaining near the Galactic plane. Notably, the "ghost" aliased artifacts of the Galaxy and the arc-like features from solar transits are significantly suppressed, demonstrating that the mPCA-UWTS framework robustly handles both compact and diffuse foreground contamination.

The cross-cylinder baseline subset yields a more homogeneous, noise-dominated morphology, further reducing the large-scale patchiness seen in the all-baseline results. This underscores the superior foreground isolation of the cross-cylinder baseline subset, though it comes at the cost of reduced sensitivity to the largest angular scales. While the all-baseline set preserves broader spatial information, it requires more stringent control of residual Galactic emission for power spectrum estimation. In both cases, a persistent brightness enhancement in daytime regions remains, reflecting the extreme intensity of solar interference and the ongoing challenge of its complete removal.

3.2. *Power Spectrum for Mock Data*

We utilize the publicly available `SuperFab` code (H. S. Grasshorn Gebhardt & O. Doré 2021) to compute the SFB power spectrum, which we adapted for 21 cm intensity mapping analysis.

Fig.12 is unfamiliar and a bit hard to understand, maybe we could add a plot before it, just showing $C_{\ell nn}$ for a few specific n values?

To characterize the SFB power spectrum of the 21 cm signal, we first analyze simulated maps containing only

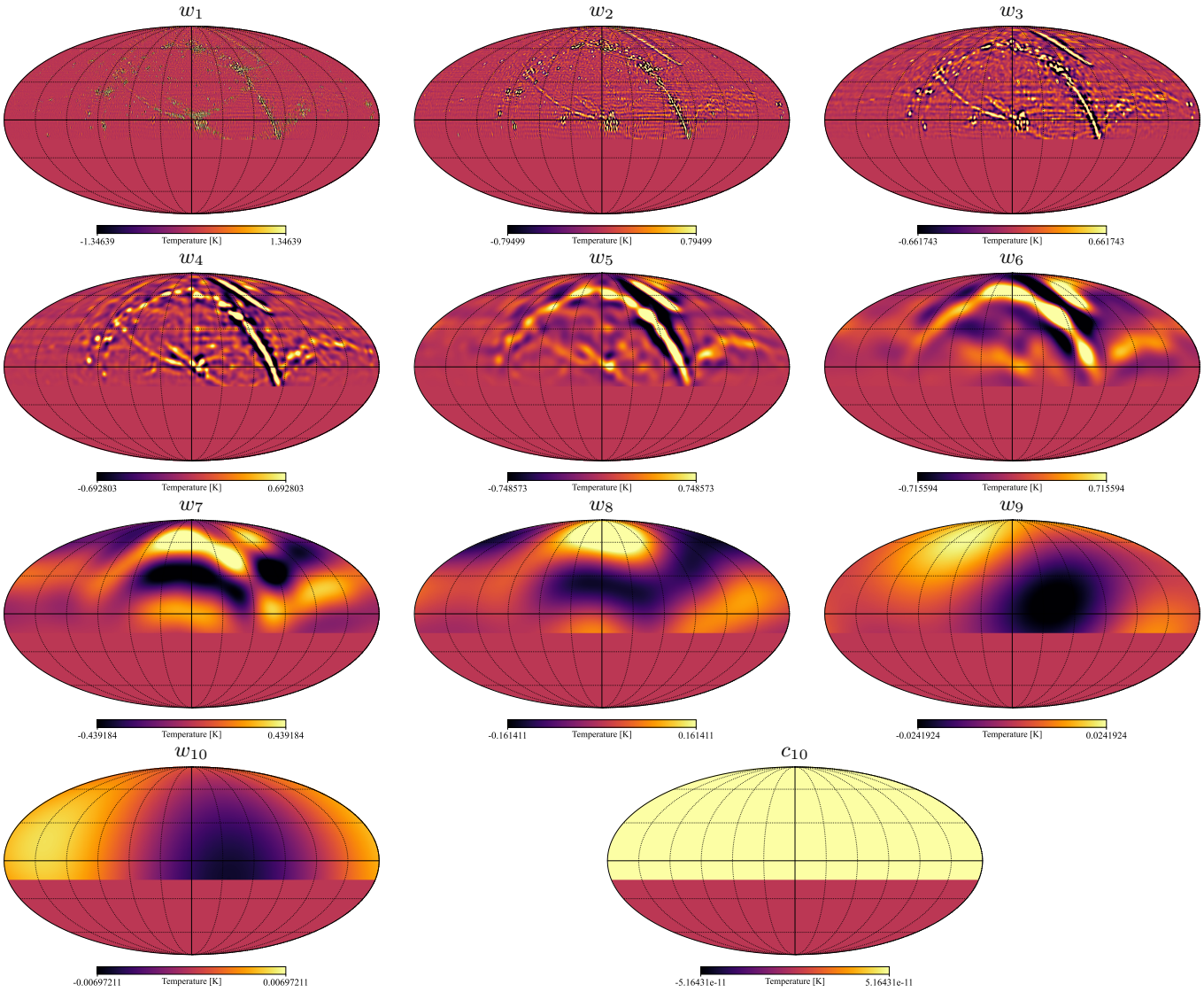

Figure 9. Multiscale decomposition of the TCPA sky map (all-baseline set) using the isotropic UWTS. The 11 scale maps (w_1–c_{10}) are displayed at the central frequency of 748.11 MHz. The plotting range for each map is restricted to $[-2\sigma, 2\sigma]$ to visualize structures across diverse angular scales.

the cosmological signal. The resulting SFB power spectra $C_{\ell nn}$, prior to window deconvolution or bandpower binning, are shown in Figure 12. For comparison, we also plot the k-binned $C_{\ell nn}$ (black dashed line) and the spherically averaged power spectrum $P(k)$ (black solid line). Most $C_{\ell nn}$ values are clustered between $\sim 10^{-5}$ and $\sim 10^{-3}\,\mathrm{K^2\,Mpc^3}\,h^{-3}$. We exclude the $\ell = 0$ modes, as they are sensitive to the global mean and are dominated by numerical noise in our implementation. We also verified that the UWTS decomposition and reconstruction process introduces errors below $\sim 10^{-6}\,\mathrm{K^2\,Mpc^3}\,h^{-3}$, confirming the signal integrity is preserved.

The k-binned $C_{\ell nn}$ and $P(k)$ display oscillatory patterns at low k, with the k-binned $C_{\ell nn}$ exhibiting a characteristic peak, for $P(k)$ at $k \approx 0.023\,h\,\mathrm{Mpc}^{-1}$, slightly shifted from the $C_{\ell nn}$ peak at $\approx 0.012\,h\,\mathrm{Mpc}^{-1}$, likely due to window function effects and the discrete nature of the SFB modes. This peak is consistent with the predicted peak in the matter power spectrum. Is this really due to the matter power, or due to the response of the array? While $P(k)$ follows the overall trend of $C_{\ell nn}$, it exhibits amplitude oscillations due to the uneven binning. At low k, $P(k)$ is systematically lower than $C_{\ell nn}$, reflecting the breakdown of the flat-sky approximation.

3.3. *Power Spectrum for Observational Data*

Figure 13 displays the SFB power spectrum for the foreground-subtracted observational data. The observational results contrast sharply with the simulations: for the all-baseline set, the observed $C_{\ell nn}$ modes span

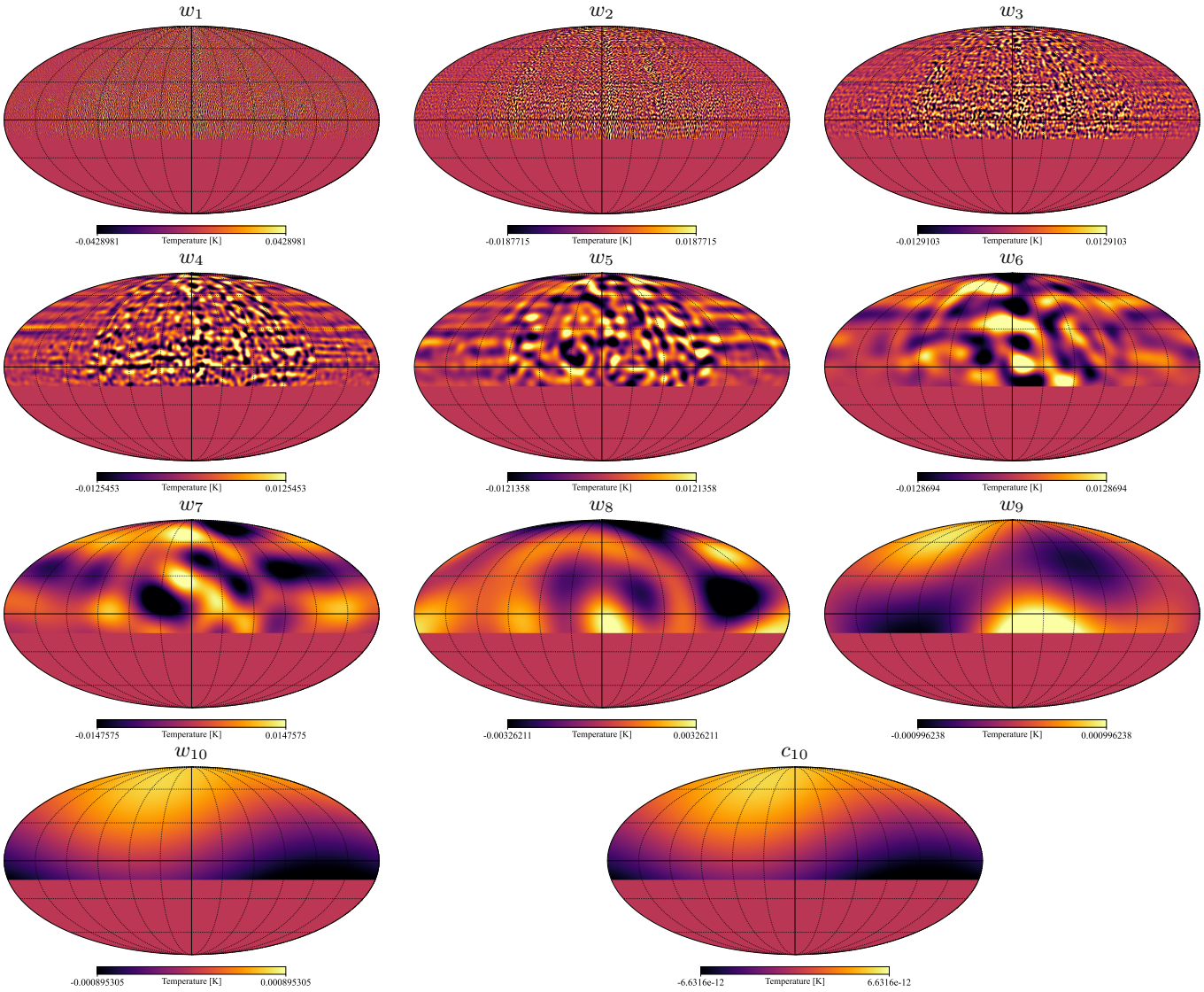

Figure 10. Residual maps for each of the 11 UWTS scales after independent PCA foreground removal (all-baseline set). The maps are shown at the central frequency (748.11 MHz) with a plotting range of $[-2\sigma, 2\sigma]$. The dominant foreground structures seen in Figure 9 are significantly suppressed, revealing scale-dependent systematic residuals and potential noise fluctuations.

a dynamic range from $\sim 10^{-6}$ to $10^3\,\mathrm{K}^2\,\mathrm{Mpc}^3\,h^{-3}$, reflecting significant residual foregrounds and instrumental noise. In the cross-cylinder baseline subset, the overall amplitude is significantly lower, with most $C_{\ell nn}$ values falling below $10^{-4}\,\mathrm{K}^2\,\mathrm{Mpc}^3\,h^{-3}$. The traditional $P(k)$ estimator yields results comparable in magnitude to the k-binned $C_{\ell nn}$ across the entire k-range but exhibits prominent oscillations at $k < 0.08\,h\,\mathrm{Mpc}^{-1}$. Are these oscillations, or merely that there is large noise at small k?

To quantify the noise level and potential pipeline-induced biases, we performed a suite of simulations using the `tlpipe` package. We generated zero-signal noise visibilities assuming a 20-day observation with a system temperature of 90 K (J. Li et al. 2020), utilizing the exact TCPA array configuration. These noise realizations were first processed through the map-making pipeline to generate noise maps. To account for the coupling between foregrounds and noise during subtraction, these noise maps were added to the observational sky maps (prior to foreground cleaning), and the combined maps were then passed through the mPCA-UWTS foreground subtraction module. In the mPCA stage, we subtracted the same number of modes per scale (N_{FG}) as used for the primary data (see Figure 8) to ensure the noise residuals correctly reflect any mode-mixing or signal loss effects. The resulting noise power spectra are plotted as pink dash-dotted lines in Figure 13. This simulation-based approach provides a more robust characterization of the effective noise floor and systematic artifacts than

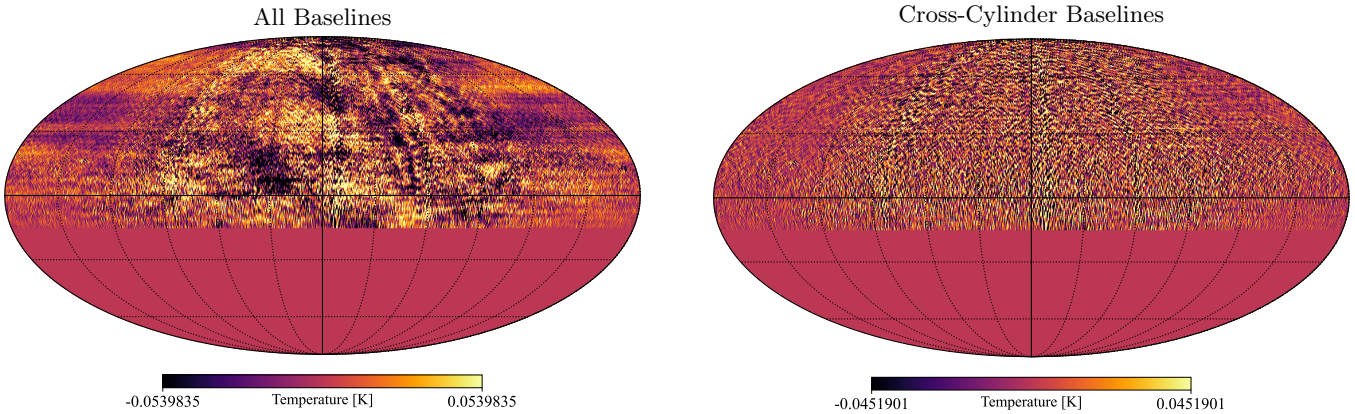

Figure 11. Final foreground-subtracted residual sky maps at the central frequency for the all-baseline set (left) and the cross-cylinder baseline subset (right). These maps are reconstructed by inverse-transforming the cleaned UWTS scale maps. The plotting range is set to $[-2\sigma, 2\sigma]$ to emphasize the distribution of residual power.

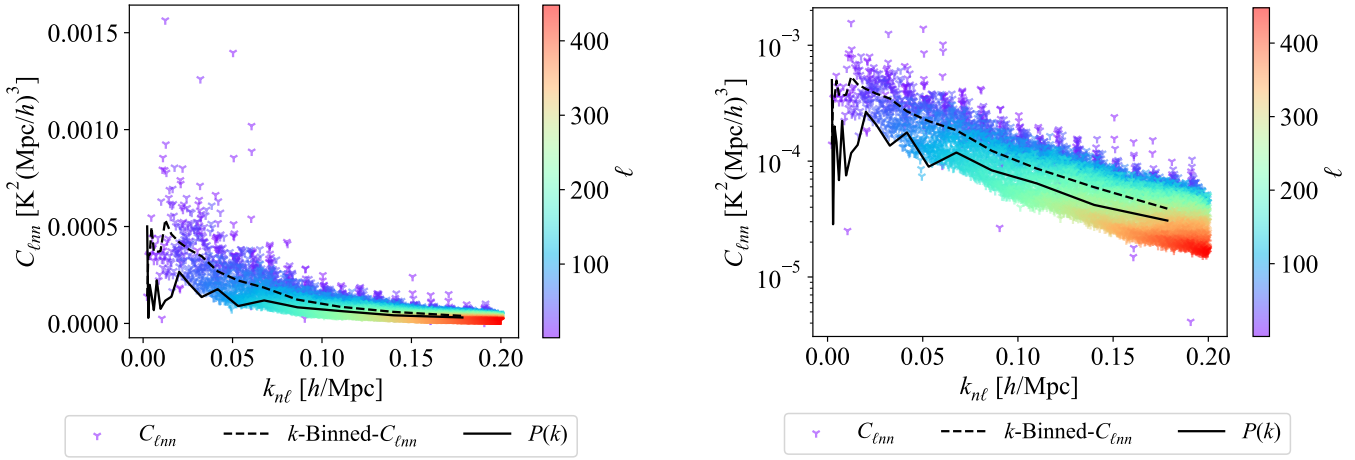

Figure 12. Three-dimensional SFB power spectrum of the simulated 21 cm signal prior to window function deconvolution and bandpower binning. Individual multipoles ℓ are distinguished by color. The results are presented on both linear (left) and logarithmic (right) scales to illustrate the dynamic range and the characteristic k-dependence of the signal.

simplified analytical models such as `21cmSense` (S. Murray et al. 2024).

The observed power in the foreground-subtracted data is generally an order of magnitude higher than the noise level, indicating the presence of significant residual foregrounds and systematic artifacts. While a more aggressive PCA cleaning could further reduce these residuals, it would also increase the risk of signal loss. Given the current sensitivity of the Tianlai Cylinder Pathfinder Array and the limited 20-day integration time, we adopt a conservative subtraction strategy. Future work will involve analyzing larger datasets with enhanced foreground mitigation techniques.

We quantitatively assess the impact of two critical post-processing steps: window function deconvolution and bandpower binning, as defined in Eq. (15) and Eq. (17), respectively. Since $\Delta n = 1$ for our survey, no bandpower binning is applied in the radial direction; hereafter, we refer to this process as angular multipole (ℓ) binning. Their effects are illustrated in Figure 14. In the simulated data, window deconvolution results in a systematic increase in the $C_{\ell nn}$ amplitude, while ℓ-binning introduces minimal changes. However, deconvolution also increases the dispersion for low-ℓ modes and occasionally yields unphysical negative values (down to $\sim -10^{-3}\,\mathrm{K}^2\,\mathrm{Mpc}^3\,h^{-3}$ for $\ell \leq 9$). The deconvolved spectra exhibit a broader distribution, whereas ℓ-binning leads to a more compact representation. Applying deconvolution after ℓ-binning mitigates these unphysical values while maintaining the elevated power level and a tighter distribution.

For the observational data, the all-baseline set shows a similar increase in the k-binned deconvolved power (black dashed line). However, the dispersion is significantly larger, with substantial unphysical values appearing across a wide range of multipoles, resulting in an

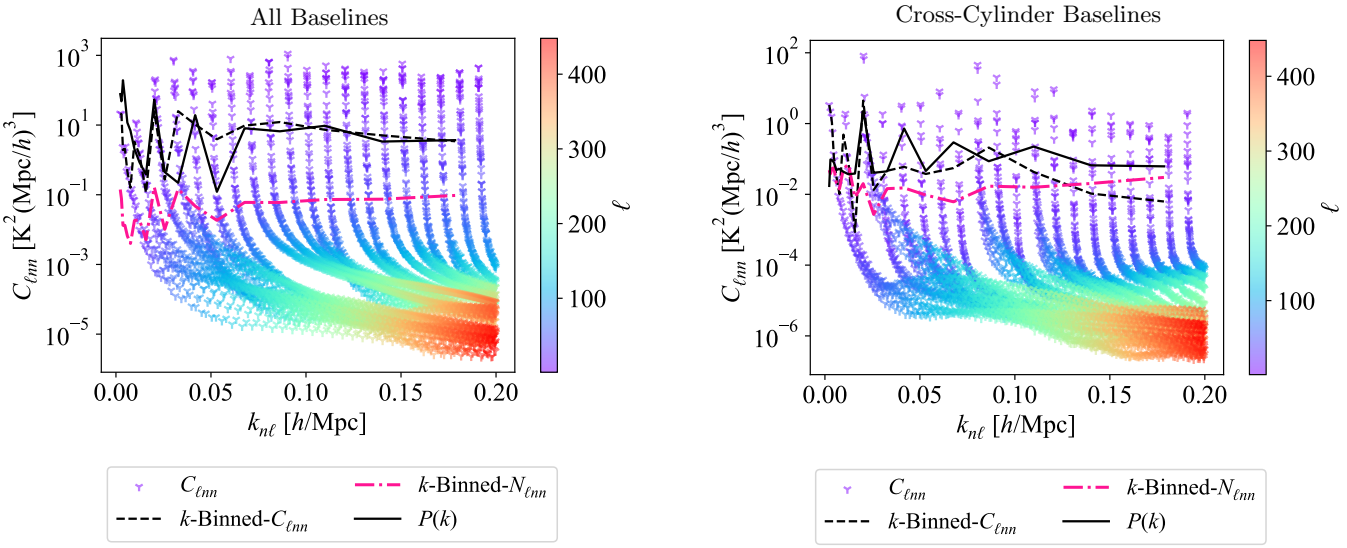

Figure 13. SFB power spectra for the foreground-subtracted all-baseline set (left) and cross-cylinder baseline subset (right) before window deconvolution and binning. The raw $C_{\ell nn}$ modes (colored markers) exhibit significant scatter due to residual foregrounds and noise. The pink dash-dotted line represents the k-binned power spectrum of the processed noise simulations, providing an empirical estimate of the noise floor.

irregular spectral shape. In the cross-cylinder baseline subset, window deconvolution also increases the $C_{\ell nn}$ amplitude, though the effect is less pronounced. The occurrence of unphysical values is reduced compared to the all-baseline set, and ℓ-binning further suppresses these artifacts.

4. DISCUSSION

The mPCA-UWTS foreground subtraction strategy, unlike traditional methods that rely primarily on frequency-domain correlations, captures both spatial and spectral foreground signatures. This multi-scale approach is critical for wide-field observations like those from the TCPA, where foreground structures exhibit significant spatial heterogeneity. Scale-dependent eigenvalue distributions reveal that large-scale emission (w_7–c_{10}) is dominated by a few coherent modes, consistent with spectrally smooth Galactic synchrotron emission. In contrast, small-scale fluctuations require more principal components ($N_{\rm FG}$) to account for their variance, as instrumental noise, point-source confusion, and frequency-dependent beam chromaticity become more prominent and less spectrally correlated. These beam effects introduce fine-grained spectral structures that are not perfectly coherent across the band, posing a challenge for standard PCA techniques.

Notably, the cross-cylinder baseline configuration consistently requires more modes than the all-baseline set at small scales (e.g., 283 vs. 324 for w_2). This suggests that the longer baselines sample spatial frequencies where point-source residuals and noise properties differ significantly, potentially leading to greater mixing between foregrounds and the cosmological signal. Compared to Gaussian Process Regression (GPR), which typically models foregrounds using flexible covariance kernels, our mPCA-UWTS method offers superior computational scalability for large datasets while achieving comparable separation performance. However, GPR may excel in modeling localized spectral features like beam ripples, suggesting that a hybrid approach—using mPCA for dominant foreground removal followed by GPR for residual systematics—could be a promising future direction.

The primary theoretical advantage of the SFB framework lies in its exact treatment of spherical sky geometry, eliminating the need for the flat-sky approximation used in conventional $P(k)$ estimators. Standard Cartesian $P(k)$ methods implicitly assume a fixed line-of-sight (LoS) direction across the survey volume, which breaks down for wide-field observations. In contrast, the SFB basis functions naturally account for sky curvature and the radial variation of the comoving distance. As shown in Figure 7, the discrete radial modes $k_{n\ell}$ form a non-uniform grid that adapts to the survey geometry, providing a more faithful representation of the 3D field than the uniform FFT grid.

Furthermore, the SFB formalism enables a clean separation of angular and radial modes through the quantum numbers (ℓ, n), offering diagnostic capabilities inaccessible to Cartesian Fourier analysis. This allows for the targeted identification and mitigation of systematic contaminants based on their characteristic scale signatures, avoiding the "power bleeding" across modes

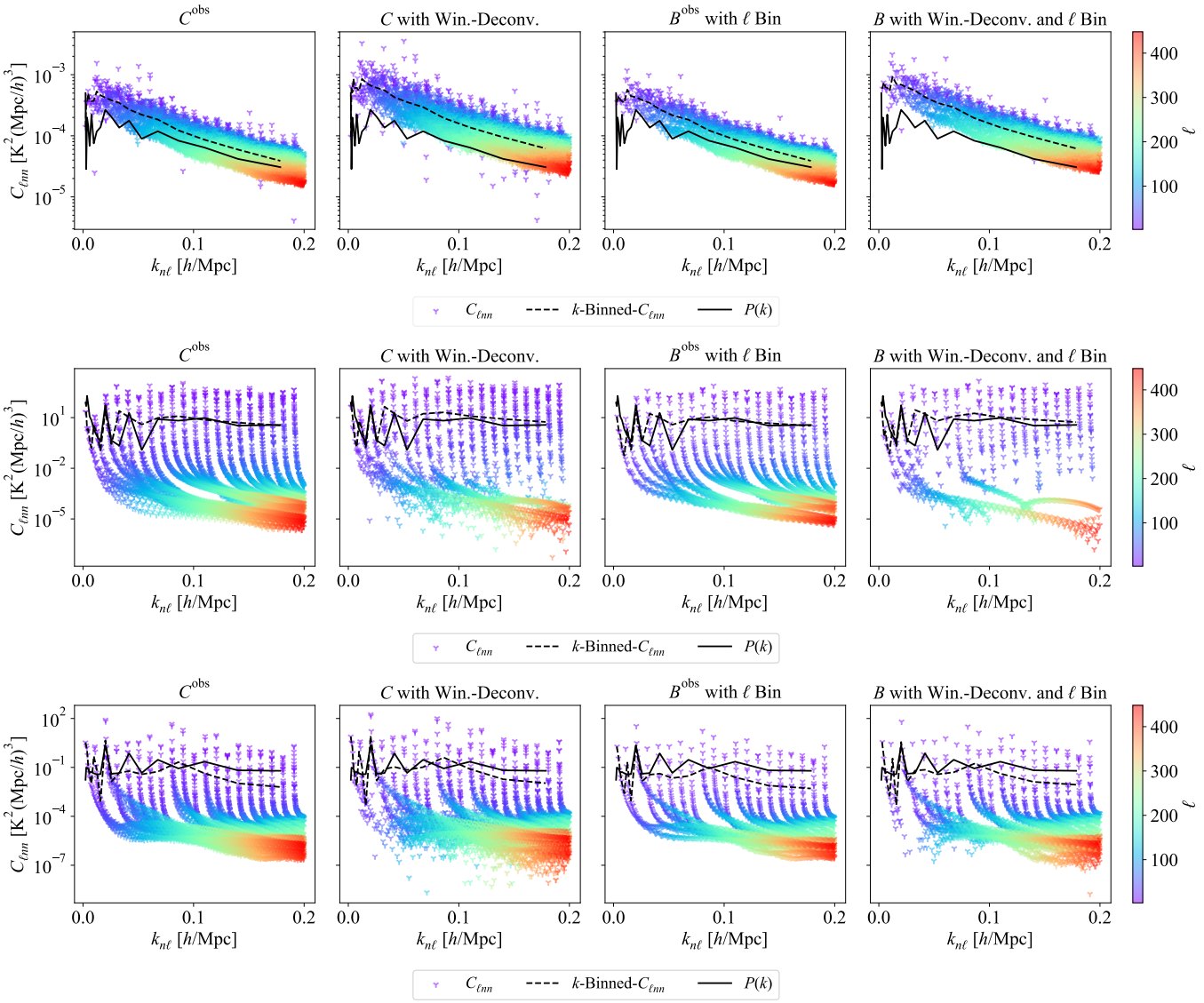

Figure 14. Impact of window deconvolution and bandpower binning on the SFB power spectrum. *Top*: Simulated 21 cm signal. *Middle*: Foreground-subtracted all-baseline data. *Bottom*: Foreground-subtracted cross-cylinder subset. In each panel, we compare the raw $C_{\ell nn}$ modes (colored markers), the k-binned spectrum (black dashed line), and the spherically averaged power spectrum $P(k)$ (black solid line). Empty bins or excluded data points represent regions with negative values or insufficient mode density.

that occurs in Cartesian k-space binning. The relatively flat or declining trend of $C_{\ell nn}$ with increasing n (for a fixed ℓ) reflects a signal dominated by noise or spectrally uncorrelated residuals at high radial modes. The traditional $P(k)$ estimator often exhibits oscillations at low k due to the cylindrical binning scheme, which can obscure real spectral features; the SFB power spectrum is less sensitive to such user-defined binning choices. The comparison between the all-baseline set and the cross-cylinder baseline subset highlights the fundamental trade-off between sensitivity and systematic control. The all-baseline set retains maximum sensitivity to large-scale modes, capturing the full range of angular scales but requiring more aggressive foreground subtraction. Conversely, the cross-cylinder baseline subset—which excludes intra-cylinder baselines—filters out much of the large-scale diffuse emission, yielding power spectra with amplitudes suppressed by 1–2 orders of magnitude. While this results in a cleaner probe closer to the expected 21 cm signal amplitude ($z \approx 0.8$–1.0), it comes at the cost of reduced sensitivity to the large-scale fluctuations necessary for BAO detection.

Finally, we find that window function deconvolution is essential for unbiased power spectrum recovery, though it introduces stability challenges. In simulations, deconvolution correctly accounts for the fractional sky cover-

age ($f_{\rm sky} \approx 62.94\%$) but increases scatter at low ℓ and occasionally yields unphysical negative values due to the inversion of ill-conditioned mixing matrices. We demonstrate that applying deconvolution after angular multipole (ℓ) binning significantly stabilizes the inversion, as binning reduces the effective noise level in each bandpower. In the observational data, however, the presence of strong foreground residuals leads to more frequent unphysical values, particularly in the all-baseline set. The more favorable performance of the cross-cylinder subset confirms that the feasibility of deconvolution depends critically on the quality of the preceding foreground subtraction.

5. SUMMARY

In this study, we have presented the inaugural application of the Spherical Fourier–Bessel (SFB) power spectrum framework to observational 21 cm intensity mapping data from the Tianlai Cylinder Pathfinder Array (TCPA). To mitigate the dominant astrophysical foregrounds, we implemented a multi-scale subtraction strategy—mPCA-UWTS—which leverages the isotropic Undecimated Wavelet Transform on the Sphere (UWTS) to decompose sky maps into distinct spatial scales, followed by an independent Principal Component Analysis (PCA) within each wavelet domain. This approach effectively isolates complex foreground signatures while preserving the large-scale cosmological signal. Our results demonstrate that the SFB framework provides a mathematically rigorous treatment of the curved sky, offering a robust alternative to Cartesian $P(k)$ estimators for future wide-field 21 cm surveys.

The primary conclusions of this work are summarized as follows:

- The mPCA-UWTS algorithm provides a highly flexible and computationally efficient approach to foreground removal. By performing scale-dependent mode subtraction, the method successfully isolates both diffuse Galactic emission and compact sources, achieving a reconstruction that is superior to standard spectral-only PCA techniques.

- The SFB power spectrum, $C_{\ell nn}$, naturally accounts for the spherical geometry of the sky, eliminating the need for flat-sky approximations. It provides a clean separation of angular (ℓ) and radial (n) modes, thereby avoiding the artificial power-spectrum oscillations often introduced by the cylindrical binning schemes used in Cartesian Fourier analysis.

- Baseline selection serves as a critical hardware-level systematic control. We find that the cross-cylinder baseline configuration—which excludes intra-cylinder correlations—yields a power spectrum with significantly lower residual contamination, providing a cleaner (though less sensitive) probe of the $z \approx 0.8$ Universe.

- Proper window function deconvolution is essential for an unbiased power spectrum estimation. We demonstrated that a hybrid post-processing strategy—applying deconvolution after angular multipole (ℓ) binning—stabilizes the matrix inversion process and successfully mitigates the occurrence of unphysical negative power values.

These results establish the SFB formalism as a powerful and scalable diagnostic tool for the next generation of 21 cm intensity mapping experiments. Unlike Cartesian Fourier methods that require uniform gridding and often introduce aliasing artifacts in wide-field surveys, the SFB basis naturally accommodates the spherical geometry with a computational complexity that scales linearly with the number of frequency channels, $O(N_{\rm freq})$, for each angular multipole ℓ. This inherent modularity allows for highly parallelized processing across different spatial scales, making it exceptionally well-suited for the massive data volumes expected from the Square Kilometre Array (SKA) and the full Tianlai array. By providing a rigorous treatment of LoS curvature and wide-angle effects within a computationally tractable framework, the SFB approach ensures that the next generation of 21 cm surveys can fully leverage their large-volume coverage for precision cosmology.

ACKNOWLEDGMENTS

This work is supported by the National SKA Program of China (Nos. 2022SKA0110100 and 2022SKA0110101), the National Natural Science Foundation of China (NSFC) International (Regional) Cooperation and Exchange Project (No. 12361141814), the NSFC (No. 12303004(ZuoSF), 12203061(LiJX).), the Specialized Research Fund for State Key Laboratory of Radio Astronomy and Technology, and the National Astronomical Observatories, Chinese Academy of Science (No. E5ZB0901). The reduction of the Tianlai observational data was carried out at National Supercomputer Center in Tianjin, and the calculations were performed on Tianhe new generation supercomputer.

Software: astropy (Astropy Collaboration et al. 2013, 2018, 2022), cora (J. R. Shaw et al. 2025), superfab

(H. S. Grasshorn Gebhardt & O. Doré 2021), tlpipe (S. Zuo et al. 2021)

REFERENCES

Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33, doi: 10.1051/0004-6361/201322068

Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, AJ, 156, 123, doi: 10.3847/1538-3881/aabc4f

Astropy Collaboration, Price-Whelan, A. M., Lim, P. L., et al. 2022, ApJ, 935, 167, doi: 10.3847/1538-4357/ac7c74

Binney, J., & Quinn, T. 1991, MNRAS, 249, 678, doi: 10.1093/mnras/249.4.678

Bowman, J. D., Rogers, A. E. E., Monsalve, R. A., Mozdzen, T. J., & Mahesh, N. 2018, Nature, 555, 67, doi: 10.1038/nature25792

Carucci, I. P., Irfan, M. O., & Bobin, J. 2020, MNRAS, 499, 304, doi: 10.1093/mnras/staa2854

Carucci, I. P., Bernal, J. L., Cunnington, S., et al. 2025, A&A, 703, A222, doi: 10.1051/0004-6361/202453461

Castro, P. G., Heavens, A. F., & Kitching, T. D. 2005, PhRvD, 72, 023516, doi: 10.1103/PhysRevD.72.023516

Chapman, E., & Jelić, V. 2019, arXiv e-prints, arXiv:1909.12369, doi: 10.48550/arXiv.1909.12369

Chapman, E., Abdalla, F. B., Harker, G., et al. 2012, MNRAS, 423, 2518, doi: 10.1111/j.1365-2966.2012.21065.x

Chen, X. 2012, in International Journal of Modern Physics Conference Series, Vol. 12, International Journal of Modern Physics Conference Series, 256–263, doi: 10.1142/S2010194512006459

Cunnington, S., Irfan, M. O., Carucci, I. P., Pourtsidou, A., & Bobin, J. 2021, MNRAS, 504, 208, doi: 10.1093/mnras/stab856

Datta, K. K., Choudhury, T. R., & Bharadwaj, S. 2007, MNRAS, 378, 119, doi: 10.1111/j.1365-2966.2007.11747.x

Eastwood, M. W., Anderson, M. M., Monroe, R. M., et al. 2018, AJ, 156, 32, doi: 10.3847/1538-3881/aac721

Fisher, K. B., Lahav, O., Hoffman, Y., Lynden-Bell, D., & Zaroubi, S. 1995, MNRAS, 272, 885, doi: 10.1093/mnras/272.4.885

Glasscock, K. A., Bull, P., Burba, J., Garsden, H., & Wilensky, M. J. 2024, RAS Techniques and Instruments, 3, 607, doi: 10.1093/rasti/rzae041

Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, ApJ, 622, 759, doi: 10.1086/427976

Grasshorn Gebhardt, H. S., & Doré, O. 2021, PhRvD, 104, 123548, doi: 10.1103/PhysRevD.104.123548

Greig, B., Mesinger, A., Koopmans, L. V. E., et al. 2021, MNRAS, 501, 1, doi: 10.1093/mnras/staa3593

Heavens, A. 2003, MNRAS, 343, 1327, doi: 10.1046/j.1365-8711.2003.06780.x

Heavens, A. F., & Taylor, A. N. 1995, MNRAS, 275, 483, doi: 10.1093/mnras/275.2.483

HERA Collaboration, Abdurashidova, Z., Adams, T., et al. 2023, ApJ, 945, 124, doi: 10.3847/1538-4357/acaf50

Hinshaw, G., Larson, D., Komatsu, E., et al. 2013, ApJS, 208, 19, doi: 10.1088/0067-0049/208/2/19

Khek, B., Gebhardt, H. G., & Doré, O. 2024, PhRvD, 110, 063524, doi: 10.1103/PhysRevD.110.063524

Lahav, O. 1993, in Cosmic Velocity Fields, ed. F. Bouchet & M. Lachieze-Rey, Vol. 9, 205, doi: 10.48550/arXiv.astro-ph/9309030

Lanusse, F., Rassat, A., & Starck, J.-L. 2015, A&A, 578, A10, doi: 10.1051/0004-6361/201424456

Leistedt, B., Rassat, A., Réfrégier, A., & Starck, J.-L. 2012, A&A, 540, A60, doi: 10.1051/0004-6361/201118463

Li, J., Zuo, S., Wu, F., et al. 2020, Science China Physics, Mechanics, and Astronomy, 63, 129862, doi: 10.1007/s11433-020-1594-8

Li, J., Suo, N., Xu, S., et al. 2025, arXiv e-prints, arXiv:2508.01413, doi: 10.48550/arXiv.2508.01413

Liu, A., & Shaw, J. R. 2020, PASP, 132, 062001, doi: 10.1088/1538-3873/ab5bfd

Liu, A., Zhang, Y., & Parsons, A. R. 2016, ApJ, 833, 242, doi: 10.3847/1538-4357/833/2/242

Mertens, F. G., Ghosh, A., & Koopmans, L. V. E. 2018, MNRAS, 478, 3640, doi: 10.1093/mnras/sty1207

Mozdzen, T. J., Bowman, J. D., Monsalve, R. A., & Rogers, A. E. E. 2017, MNRAS, 464, 4995, doi: 10.1093/mnras/stw2696

Murray, S., Pober, J., & Kolopanis, M. 2024, The Journal of Open Source Software, 9, 6501, doi: 10.21105/joss.06501

Nicola, A., Refregier, A., Amara, A., & Paranjape, A. 2014, PhRvD, 90, 063515, doi: 10.1103/PhysRevD.90.063515

Padmanabhan, H. 2024, arXiv e-prints, arXiv:2411.08113, doi: 10.48550/arXiv.2411.08113

Pal, S., Elahi, K. M. A., Bharadwaj, S., et al. 2022, MNRAS, 516, 2851, doi: 10.1093/mnras/stac2419

Samushia, L. 2019, arXiv e-prints, arXiv:1906.05866, doi: 10.48550/arXiv.1906.05866

Semenzato, F., Bertacca, D., & Raccanelli, A. 2025, PhRvD, 111, 023522, doi: 10.1103/PhysRevD.111.023522

Shaw, J. R., Gray, L., Foreman, S., et al. 2025, cora, v25.10.0 Zenodo, doi: 10.5281/zenodo.17438047
Shaw, J. R., Sigurdson, K., Pen, U.-L., Stebbins, A., & Sitwell, M. 2014, ApJ, 781, 57, doi: 10.1088/0004-637X/781/2/57
Singh, S., Jishnu, N. T., Subrahmanyan, R., et al. 2022, Nature Astronomy, 6, 607, doi: 10.1038/s41550-022-01610-5
Soares, P. S., Watkinson, C. A., Cunnington, S., & Pourtsidou, A. 2022, MNRAS, 510, 5872, doi: 10.1093/mnras/stab2594
Starck, J.-L., Fadili, J., & Murtagh, F. 2007, IEEE Transactions on Image Processing, 16, 297, doi: 10.1109/TIP.2006.887733
Starck, J.-L., Moudden, Y., Abrial, P., & Nguyen, M. 2006, A&A, 446, 1191, doi: 10.1051/0004-6361:20053246
Sun, S., Li, J., Wu, F., et al. 2022, Research in Astronomy and Astrophysics, 22, 065020, doi: 10.1088/1674-4527/ac684d
Switzer, E. R., Masui, K. W., Bandura, K., et al. 2013, MNRAS, 434, L46, doi: 10.1093/mnrasl/slt074
Wang, X., Tegmark, M., Santos, M. G., & Knox, L. 2006, ApJ, 650, 529, doi: 10.1086/506597
Wen, R. Y., Gebhardt, H. S. G., Heinrich, C., & Doré, O. 2025, PhRvD, 112, 063518, doi: 10.1103/hqc3-tbxc
Wu, F., Li, J., Zuo, S., et al. 2021, MNRAS, 506, 3455, doi: 10.1093/mnras/stab1802
Xu, Y., Wang, X., & Chen, X. 2015, ApJ, 798, 40, doi: 10.1088/0004-637X/798/1/40
Yu, K., Wu, F., Zuo, S., et al. 2023, Research in Astronomy and Astrophysics, 23, 105008, doi: 10.1088/1674-4527/acf032
Zhang, J., Ansari, R., Chen, X., et al. 2016a, MNRAS, 461, 1950, doi: 10.1093/mnras/stw1458
Zhang, J., Zuo, S.-F., Ansari, R., et al. 2016b, Research in Astronomy and Astrophysics, 16, 158, doi: 10.1088/1674-4527/16/10/158
Zuo, S., Pen, U.-L., Wu, F., et al. 2019, AJ, 157, 34, doi: 10.3847/1538-3881/aaf4c0
Zuo, S., Li, J., Li, Y., et al. 2021, Astronomy and Computing, 34, 100439, doi: 10.1016/j.ascom.2020.100439