





Matching cosmic shear analysis in harmonic and real space

Andy Park ^{1,2}★, Sukhdeep Singh,^{1,2} Xiangchong Li ^{1,2}, Rachel Mandelbaum ^{1,2} and Tianqing Zhang ^{1,2,3}

¹*McWilliams Center for Cosmology and Astrophysics, Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15213, USA*

²*NSF AI Planning Institute, Carnegie Mellon University, Pittsburgh, PA 15213, USA*

³*Department of Physics and Astronomy and PITT PACC, University of Pittsburgh, Pittsburgh, PA 15260, USA*

Accepted 2025 May 9. Received 2025 May 2; in original form 2024 April 10

ABSTRACT

Recent cosmic shear analyses have exhibited discrepancies of up to 1σ between the inferred cosmological parameters when analysing summary statistics in real space versus harmonic space. In this paper, we demonstrate the consistent measurement and analysis of cosmic shear two-point functions in harmonic and real space using the *i*MASTER algorithm. This algorithm provides a unified prescription to model the survey window effects and scale cuts in both real space (due to observational systematics) and harmonic space (due to model limitations), resulting in a matching estimation of the cosmic shear power spectrum from both harmonic and real space estimators. We show that the *i*MASTER algorithm gives matching results using measurements from the HSC Y1 mock shape catalogs in both real and harmonic space, resulting in matching inferences of $S_8 = \sigma_8(\Omega_m/0.3)^{0.5}$. This method provides an unbiased estimate of the cosmic shear power spectrum, and S_8 inference that has a correlation coefficient of 0.997 between analyses using measurements in real space and harmonic space when S_8 is the only free parameter. We observe the mean difference between the two inferred S_8 values to be 0.0004 across noise-free mock realizations, far below the observed difference of 0.042 for the published HSC Y1 analyses and well below the statistical uncertainties. While the notation employed in this paper is specific to photometric galaxy surveys, the methods are equally applicable and can be extended to spectroscopic galaxy surveys, intensity mapping, and CMB surveys.

Key words: gravitational lensing: weak – cosmology: observations.

1 INTRODUCTION

Weak gravitational lensing (WL) is a powerful tool for investigating the large-scale structure (LSS) of the Universe. It involves the coherent deflection of light rays from distant source objects by the foreground matter distribution. Accurate WL measurements yield crucial cosmological insights into the history of structure formation, the distribution of dark matter, and the nature of dark energy. The primary advantage of this method lies in its ability to directly study the properties and distribution of dark matter through sensitivity to the matter density field along the line of sight (Hu 2001; Huterer 2010; Weinberg et al. 2013). Cosmic shear measurements, which analyse the coherent distortion pattern in the observed shapes of distant source galaxies, stand out as an effective means to constrain structure growth and refine the cosmological model.

The 2020s will be the most exciting decade for observational cosmology to date, due to dramatic increases in data volume as detectors, telescopes, and computers become ever more powerful. The upcoming generations of imaging surveys (known as Stage IV surveys) like the Vera C. Rubin Observatory Legacy Survey of Space and Time (LSST; LSST Science Collaboration 2009; Ivezić et al. 2019), *Euclid* (Laureijs et al. 2011), and the *Nancy Grace Roman* Space Telescope High Latitude Imaging Survey (Akeson et al. 2019)

will provide percent-level cosmic shear measurements by covering a wider sky coverage and collecting a larger sample of fainter source galaxies compared to previous surveys such as the Subaru Hyper Suprime-Cam (HSC; Aihara et al. 2018) survey, the Dark Energy Survey (DES; Dark Energy Survey Collaboration 2016), and the Kilo-Degree Survey (KiDS; de Jong et al. 2017), and thus require us to revisit many of the assumptions made in previous studies.

Once the data are acquired and processed into shear maps or shear catalogues, the next step is to compute the two-point summary statistics. One question that we need to address at this stage is whether to carry out the cosmic shear analysis in real space or harmonic space; each space has its own advantages and disadvantages, and we introduce the trade-offs in Section 2.3. Analysis in real space corresponds to measuring the two-point correlation functions of the galaxy shear estimates as a function of angular separation (e.g. as in Schneider et al. 2022; Secco et al. 2022; Li et al. 2023), and analysis in harmonic space corresponds to measuring the cosmic shear power spectrum as a function of multipole moments (e.g. as in Doux et al. 2022; Loureiro et al. 2022; Dalal et al. 2023). One major challenge is to account for the survey window. The survey window is a function that captures the effects of the survey geometry and other observational effects that affect the selection of galaxies included in our samples. The survey window complicates the power spectrum analysis as the survey mask couples Fourier modes, which would otherwise be independent. This leads to mode coupling in

* E-mail: chanhyup@andrew.cmu.edu

the power spectrum, which is commonly addressed using pseudo- C_ℓ estimators (e.g. Wandelt, Hivon & Górski 2001; Hikage et al. 2011; Alonso et al. 2019).

Since the coherent distortion of galaxy shapes is caused by a scalar gravitational potential of the matter distribution, we can further separate the lensing signal into curl-free (E mode) and divergence-free (B mode) components. The survey window also introduces leakage of E modes into the B mode component of a given map, and vice versa, an effect that needs to be modelled and corrected using detailed knowledge of the window. The modelling of the survey window in harmonic space was first addressed in Hivon et al. (2002), who introduced the MASTER algorithm to deconvolve the effects of the survey window to reconstruct the power spectrum of the underlying fields for Cosmological Background Microwave (CMB) analysis (Lewis, Challinor & Turok 2001; Wandelt et al. 2001; Brown, Castro & Taylor 2005). Alonso et al. (2019) introduced the now standard NAMASTER software with an implementation of the MASTER algorithm for spin-0 and spin-2 fields that is also used for galaxy clustering and cosmic shear analyses.

Throughout this paper, we define two estimators as *matching* if they yield results that agree within a percent when applied to the same data set, and use the term *biased* to describe cases where systematic effects cause a significant deviation from the expected result. For example, Hamana et al. (2020) and Doux et al. (2021) show how such a disparity in approach produces an apparent mismatch between the analyses in real space and harmonic space when performing cosmological parameter inference. Crucially, if we truly believe certain scales cannot be well modelled, they should be removed regardless of the space in which the analysis is performed.

The survey window also affects the two-point correlation functions in real space, although this effect is typically ignored since the commonly used estimator removes most of the window effects (see appendix D of Singh et al. 2020) for demonstration of residual percent-level window effects on correlation functions).

After measuring the summary statistics, we apply scale cuts on the measurements to restrict our analysis to scales where the model is reliable (normally done in harmonic space where the models are defined) and to remove scales contaminated by observational systematics (normally done in real space to account for physical and observational effects tied to angular separation) before performing cosmological parameter inference.

We can typically model a limited range of scales that is smaller than the range of scales probed by the current data. Thus, we need to apply scale cuts carefully to extract as much information as possible while avoiding biases from applying the model outside its range of validity. Failure to account for the scale cut that was implemented in one space when modelling the other space can lead to using different sets of information in real versus harmonic space analysis, even though the underlying field is the same.

There have been many efforts to introduce a statistic that can separate out the E/B modes. Schneider, van Waerbeke & Mellier (2002) introduced the band power spectra that used the aperture mass dispersion $\langle M_{\text{ap}}^2(\theta) \rangle$ and $\langle M_\times^2(\theta) \rangle$ (Schneider et al. 1998) measured from a linear combination of the two-point correlation functions weighted by specific window functions to perform the E/B mode decomposition. However, calculating the aperture mass dispersion requires estimation of the shear correlation functions down to zero angular separation, which is not possible in practice. To remedy this, they applied apodization in the window function that truncates the integral over all angular distances and reduces the ringing effect caused by the limited range of the two-point correlation functions.

Schneider, Eifler & Krause (2010) introduced the Complete Orthogonal Sets of E/B -Mode Integrals (COSEBI) method to measure pure E and B -mode cosmic shear power spectra from the two-point correlation functions over a finite interval $\theta_{\min} \leq \theta \leq \theta_{\max}$. The method, similar to Schneider et al. (2002), uses a linear combination of binned measured two-point correlation functions weighted by specific window functions in the angular scale. Becker & Rozo (2016) introduced a similar band-power estimator for cosmic shear power spectra from linear combinations of binned two-point correlation functions. Although the aforementioned methods return unbiased two-point statistics estimators, there were still inconsistencies in analysis between the two spaces.

Asgari et al. (2021) used three different two-point statistics (band power spectra, COSEBIs, and two-point correlation functions) on the KiDS data set and compared the measured values of the parameter $S_8 \equiv \sigma_8(\Omega_m/0.3)^{0.5}$, where σ_8 is the standard deviation of matter density fluctuations in spheres of radius 8 Mpc h^{-1} (here $h \equiv H_0/100$) and Ω_m is the fraction of matter density in the Universe with respect to the critical density. Comparing the inferred S_8 from each method provided a useful internal consistency check between the methods; Asgari et al. (2021) found a 0.4σ difference in S_8 between the COSEBIs and the two-point correlation functions analyses. Camacho et al. (2022) did a cosmic shear analysis in real space and harmonic space using the DES-Y1 METACALIBRATION shear catalogue, and also found 0.4σ difference in S_8 . These past results show that offsets at the $\sim 0.5\sigma$ level can arise because the different two-point statistics have different sensitivities to ℓ scales, even for the same data set.

Singh (2021) further improved the MASTER algorithm (to produce *i*MASTER) and generalized the method to account for the scale cuts used in the analysis, after which the power spectra and two-point correlation function estimates match. In this work, we present the first application of the *i*MASTER algorithm to survey mock catalogs with complex geometry.

We measure the cosmic shear two-point functions in harmonic and real space in the HSC Y1 mock shape catalogues (Shirasaki et al. 2019), using the *i*MASTER method to estimate the power spectrum from both harmonic and real space estimators consistently, accounting for both the survey window and the scale cuts used in the analysis in a way that ensures matching results. We quantify the consistency between the two analyses by measuring $\delta S_8 = S_{8|R \rightarrow F} - S_{8|F}$, where $S_{8|R \rightarrow F}$ is the inferred S_8 value going from the real space estimator to the harmonic space estimator, and $S_{8|F}$ using the harmonic space estimator directly.

This paper is organized as follows: In Section 2, we provide an overview of cosmic shear analysis. In Section 3, we describe how two-point estimators are measured in harmonic and real space and outline steps to make a matching two-point estimator that includes the impact of the survey window and scale cuts. Then, in Section 4, we describe the shear catalogue used to test the proposed formalism in this paper and the implementation details of our estimators. The results of the reconstruction from harmonic space and real space with likelihood analysis are shown in Section 5. Finally, in Section 6, we summarize our results and the future outlook.

2 BACKGROUND

In this section, we provide a brief background to weak gravitational lensing. We introduce the history of cosmic shear in Section 2.1, discuss how shear is estimated from the shapes of galaxies in Section 2.2, and describe how shear is used in cosmological analysis in Section 2.3.

2.1 Cosmic shear

Cosmic shear is the coherent distortion of images of distant background galaxies by the foreground Large Scale Structure (LSS) in the Universe (see Kilbinger 2015 for a review). Such distortions are induced by all matter along the line of sight, making cosmic shear a powerful tool for probing the matter distribution, providing us with a unique glimpse into the structure and evolution of the Universe on large scales. The first cosmic shear detections were made in the early 2000s (e.g. Bacon, Refregier & Ellis 2000; Kaiser, Wilson & Luppino 2000; Wittman et al. 2000; Van Waerbeke et al. 2001; Rhodes et al. 2004), and further advancements were made in the subsequent years until now. These improvements included larger volumes of survey data, improved redshift estimation (e.g. Myles et al. 2021; Rau et al. 2023), and refined statistical analysis (e.g. Schrabback et al. 2010). Recent surveys, as well as ongoing ones, have successfully measured cosmic shear and constrained cosmological parameters to high precision. The upcoming imaging surveys like LSST (Ivezić et al. 2019), *Euclid* (Laureijs et al. 2011), and *Nancy Grace Roman* Space Telescope (Akeson et al. 2019) will allow us to make percent-level measurements with unprecedented data volume. As these data sets become increasingly precise, ensuring *matching* results between real-space and harmonic-space analyses is essential to prevent systematic biases in cosmological constraints.

2.2 Shear estimation

The distortion of galaxy shapes by weak gravitational lensing can be well described by a local coordinate transformation. The mapping between the unlensed coordinates (x_u, y_u) and the lensed coordinates (x_l, y_l) is defined according to the distortion matrix:

$$\begin{pmatrix} x_l \\ y_l \end{pmatrix} = \begin{pmatrix} 1 - \gamma_1 - \kappa & -\gamma_2 \\ -\gamma_2 & 1 + \gamma_1 - \kappa \end{pmatrix} \begin{pmatrix} x_u \\ y_u \end{pmatrix}, \quad (1)$$

where γ_1 and γ_2 are the two lensing shear components and κ is the lensing convergence. Positive (negative) γ_1 stretches the image along the x (y) axis direction, and positive (negative) γ_2 stretches the image in the direction at an angle of 45 deg (-45 deg) with the x axis. The convergence κ changes the size and brightness of the unlensed galaxy. Shear negates when rotated by $\pi/2$, making it a spin-2 quantity, whereas convergence is a spin-0 scalar quantity as it is invariant under rotation. We encode the two components of shear into a shear spinor $\gamma = \gamma_1 + i\gamma_2 = |\gamma|e^{2i\phi}$, where i is the imaginary number unit, ϕ is the azimuthal angle, and the factor of 2 arises due to the spin-2 nature of the shear. It is common to factor out $(1 - \kappa)$ from the distortion matrix in equation (1) as this multiplier only affects the size but not the galaxy's shape. Since cosmic shear is based on the measurement of galaxy shapes and not the size, the observable that we can access in practice is not the shear γ , but rather the reduced shear g , which can be written in terms of γ as

$$g = \frac{\gamma}{1 - \kappa}. \quad (2)$$

The reduced shear g has the same spin-2 properties as the shear (i.e. $g = g_1 + ig_2 = |g|e^{2i\phi}$). In the weak lensing limit, where the effect of lensing is on the order of a few percent or less ($\kappa, \gamma \ll 0.02$), $g \simeq \gamma$ is a reasonable approximation to linear order (Shapiro 2009). For the rest of the work, we use the reduced shear g and shear γ interchangeably.

In cosmic shear, we use galaxy ellipticity to quantify the spin-2 aspect of galaxy shape and measure weak lensing shear distortion. In practice, we define an ellipticity spinor $e = e_1 + ie_2$ where the two components of the spinor follow the same definition as the two shear

components. There exist many methods in the literature for using the light profile of a galaxy to define the ellipticity, e.g. using moments or derivatives of the galaxy's light intensity profile in harmonic space (Zhang 2008; Bernstein & Armstrong 2014), decompose light intensity profile onto basis functions (Refregier & Bacon 2003; Massey & Refregier 2005; Li et al. 2018), or parametric galaxy's light intensity profile model (Zuntz et al. 2013; Fenech Conti et al. 2017). We then obtain the linear shear responses of the ellipticities (Huff & Mandelbaum 2017; Li & Mandelbaum 2023) to infer the shear from an ensemble of galaxies with ellipticity measurements, and there are many ways to correct noise bias, selection bias and detection bias (Sheldon & Huff 2017; Sheldon et al. 2020; Li, Li & Massey 2022; Li & Mandelbaum 2023).

2.3 Cosmic shear analysis

The outcome of the measurements described in Section 2.2 is a shear catalogue. The shear catalogue is then compressed using a summary statistic. Two common summary statistics are to measure the shear-shear power spectrum C_ℓ in harmonic space and the two-point correlation functions (TPCFs) $\xi_{\pm}(\theta)$ in real space. Higher-order statistics can also be used to extract the non-Gaussian nature of the cosmic density field not encoded by the two-point summary statistics (e.g. Zaldarriaga & Scoccimarro 2003). Both spaces have their own advantages and disadvantages. It is often easier to estimate the TPCFs (Landy & Szalay 1993) as our data are in real space, and we have a better understanding of observational systematics in real space. In contrast, Gaussian and non-Gaussian modes and linear and non-linear scales are clearly separated in harmonic space (Hamimeche & Lewis 2008; Alonso et al. 2019). In harmonic space analysis, individual modes are significantly less correlated compared to real space analysis, making the covariance matrix estimation easier. Our theoretical model of the summary statistics is also computed in harmonic space. For this reason, there has been a large body of literature aimed at developing efficient power spectrum estimators, which have been widely employed in the Cosmic Microwave Background (CMB) two-point measurements (e.g. Lewis et al. 2001; Wandelt et al. 2001; Brown et al. 2005). While both approaches should ideally yield *matching* cosmological constraints, previous studies have found discrepancies between real space and harmonic space analyses when applying different scale cuts or improperly handling the survey function (Hamana et al. 2020; Doux et al. 2021). These mismatches highlight the importance of developing methods that correctly account for scale-dependent effects to avoid introducing systematic biases into cosmological inference.

The shear field is a spin-2 field, and we can further decompose the lensing signal into the curl-free (E mode) and divergence-free (B mode) components. We use the Limber approximation (Limber 1953) to compute the model power spectra, which is valid for scales used in this analysis (Dalal et al. 2023).

Following this assumption, the cosmic shear power spectra for flat spatial geometry can be related to the matter power spectrum as:

$$C_{ij}^{EE}(\ell) = \int_0^{x_H} d\chi \frac{q^i(\chi)q^j(\chi)}{\chi^2} P_m \left(k = \frac{\ell + 1/2}{\chi}, z(\chi) \right), \quad (3)$$

where i and j are tomographic bins, χ is the comoving distance, χ_H is the comoving horizon distance, and P_m is the non-linear matter power spectrum. The lensing efficiency, $q^i(\chi)$ is defined as:

$$q^i(\chi) = \frac{3}{2} \Omega_m \left(\frac{H_0}{c} \right)^2 \frac{\chi}{a(\chi)} \int_{\chi}^{x_H} d\chi' n^i(\chi') \frac{\chi - \chi'}{\chi'}, \quad (4)$$

where H_0 is the Hubble constant today, a is the scale factor, Ω_m is the matter density parameter, and $n^i(\chi)$ is the redshift distribution of source galaxies in the i th tomographic bin. Since the lensing is caused by the scalar gravitational potential of the lens, the B mode is zero, i.e. $C_{ij}^{BB}(\ell) = 0$. However, E mode leakage and some systematics can introduce B modes (e.g. Alonso et al. 2019).

The cosmic shear TPCFs are expressed as a transformation of E and B modes of the power spectrum. With the flat sky approximation, the transformation is the Hankel transform:

$$\xi_{\pm}^{ij}(\theta) = \frac{1}{2\pi} \int d\ell \ell J_{0/4}(\theta\ell) (C_{ij}^{EE}(\ell) \pm C_{ij}^{BB}(\ell)), \quad (5)$$

where $J_{0/4}$ are the 0th/4th-order Bessel functions of the first kind. In the curved sky limit, the cosmic shear TPCFs cannot be written as the Hankel transform of the power spectrum, and should be replaced by the spherical transform

$$\xi_{\pm}^{ij}(\theta) = \sum_{\ell} \frac{2\ell+1}{4\pi} (C_{ij}^{EE}(\ell) \pm C_{ij}^{BB}(\ell)) {}_2d_{\ell,\pm 2}(\theta) \quad (6)$$

$$= H_{\theta,\ell}^{\pm} (C_{ij}^{EE}(\ell) \pm C_{ij}^{BB}(\ell)) \quad (7)$$

where ${}_2d_{\ell,\pm 2}$ is the Wigner-d matrix and we use $H_{\theta,\ell}^{\pm}$ to denote the spherical transform operator. The summation is over all ℓ and consequently computing ξ_{\pm} to an arbitrary accuracy can be computationally expensive. Here, we have added the B mode power spectrum for completeness, and equations (6) and (7) apply to any spin-2 field. Hereafter, we will drop the subscript for tomographic bin pairs and write the multipole modes in the subscript instead.

3 METHOD

In this section, we describe the measurement of the cosmic shear power spectrum (Section 3.1.2) and two-point correlation functions (TPCFs; Section 3.2) from the shear catalogue. We discuss the methods to account for survey window and scale cuts, and how to consistently reconstruct the power spectrum from the TPCFs and vice versa in Sections 3.3 and 3.4. The motivation behind our method is to ensure *matching* cosmic shear measurements between real-space and harmonic space analyses. Since each space has distinct advantages and challenges, properly accounting for survey masks, pixelization effects, and scale cuts is essential to avoid introducing systematic biases into cosmological inferences.

3.1 Power spectra

In this section, we describe how the power spectra are measured in practice. In Section 3.1.1, we describe the idealized case where we have a full-sky shear map. In Section 3.1.2, we describe the biases that arise in power spectrum estimation due to partial sky and non-uniform coverage and introduce the formalism to correct it.

3.1.1 Full sky shear map

For a full-sky uniform shear map, where the number density does not vary across the sky, the spin-2 continuous shear field can be decomposed into two scalar fields using spherical harmonics as

$$(\gamma_1 + i\gamma_2)(\mathbf{n}) = - \sum_{\ell m} [E_{\ell m} + iB_{\ell m}] {}_{\pm 2}Y_{\ell m}(\mathbf{n}), \quad (8)$$

where \mathbf{n} is the sky position, $E_{\ell m}$ and $B_{\ell m}$ are the curl-free and divergence-free components of the shear field mentioned in Section 2.3, and ${}_s Y_{\ell m}$ are the spin-weighted spherical harmonics. For two

shear fields, the cosmic shear power spectrum is measured by taking the average of the product of two spherical harmonic coefficients $\psi_{\ell m}$ and $\phi_{\ell m}$ where $\psi, \phi \in \{E, B\}$ as

$$\langle \psi_{\ell m} \phi_{\ell m}^* \rangle = \delta_{\ell\ell'} \delta_{mm'} \hat{C}_{\ell}^{\psi\phi}. \quad (9)$$

Throughout the paper, we will use $\langle \cdot \rangle$ to denote the ensemble average, \bar{x} to represent the sample mean of quantity x , and \hat{x} to represent the measurement from data. In this work, we only consider the auto power spectra C_{ℓ}^{EE} and C_{ℓ}^{BB} . For a full sky uniform coverage map, the optimal estimator of the cosmic shear power spectrum is then:

$$\hat{C}_{\ell}^{EE} = \frac{1}{2\ell+1} \sum_m E_{\ell m} E_{\ell m}^* \quad (10)$$

$$\hat{C}_{\ell}^{BB} = \frac{1}{2\ell+1} \sum_m B_{\ell m} B_{\ell m}^*, \quad (11)$$

where the power spectrum is averaged over the m values for each multipole mode ℓ . For notational convenience, we assume the two shear fields to be in the same redshift bin and omit explicit expressions for the tomographic redshift bin pairs for the rest of the paper. In the idealized case of a full sky, uniform survey, the power spectrum estimators are *matching* between real and harmonic space because there is no survey mask or scale-dependent selection effects. However, real observations introduce non-trivial window functions and observational systematics that must be accounted for to ensure unbiased results.

3.1.2 Masked shear map

Equations (10) and (11) are for an idealized case – they do not account for the survey geometry or for the scale cuts used in real space to remove scales that may be contaminated by observational systematics. In practice, we have non-uniform coverage of the sky, due to the limited exposure time, finite survey area, bad pixels, and regions masked due to bright stars. Once we obtain a galaxy shear catalogue, we pixelize the shear value onto a finite grid. When using a pixelized map, the measurements can be affected by the pixel area and the number of pixels as the pixelized signal is the average within each pixel of the underlying true signal. The measured shear value at pixel p is defined as the sum of the shear values of each galaxy in pixel p divided by the mean number of galaxies:

$$\hat{\gamma}(\mathbf{n}_p) = \frac{1}{\bar{N}_g} \sum_{i \in p} \hat{\gamma}_i, \quad (12)$$

where \bar{N}_g is the mean number of galaxies per pixel given by the ratio of total galaxy number to the number of pixels within the survey mask. The theoretical model for the pixelized shear field is given by

$$\gamma(\mathbf{n}_p) = \int \Pi(\mathbf{n} - \mathbf{n}_p) W^{\gamma}(\mathbf{n}) \gamma(\mathbf{n}) d\mathbf{n}, \quad (13)$$

where \mathbf{n}_p is the sky position of the p -th pixel, Π is the pixel window function which is equal to the inverse of the pixel area for galaxies within the pixel and zero outside, and the weight $W^{\gamma}(\mathbf{n})$ (hereinafter the survey window) is defined by

$$W^{\gamma}(\mathbf{n}) = \frac{\hat{N}_g(\mathbf{n})}{\bar{N}_g}, \quad (14)$$

where $\hat{N}_g(\mathbf{n})$ is the observed number of galaxies at position \mathbf{n} . Note that \bar{N}_g is a normalization choice we made and as long as normalization choice is treated consistently between data and model (including covariance), it does not have any impact on the analysis (its impact cancels out in the likelihood function). This definition

includes clustering and shot noise in the survey window. Using a constant \bar{N}_g in the denominator of equation (12) instead of the actual number of galaxies per pixel provides a more stable estimator of the shear. Dividing by the observed number of galaxies in each pixel would increase the noise in low-density pixels, as these estimators are highly sensitive to stochastic fluctuations in regions with fewer galaxies. Instead, the constant \bar{N}_g acts as an implicit weighting factor; it reduces the relative contribution of noisy, low-density pixels while maintaining the contribution from high-density, low-noise pixels. Since this weighting is consistently included in the model, the final analysis remains unbiased. A more optimal choice of the window includes inverse variance weighting (the FKP weighting), which depends on the signal and noise power spectrum (Feldman, Kaiser & Peacock 1994; Hamilton 1997), but for this work, we adopt the weighting scheme of equation (14). In summary, equation (12) is the measured value of shear at pixel p , and equation (13) is the theory shear value at pixel p that accounts for the survey window and the pixel window function.

In equation (13), the pixelized shear field relates to the true underlying shear field via two operations in real space: multiplication of the survey window W^γ followed by convolution of the pixel window function Π . Multiplication by the survey window W^γ in real space is equivalent to convolving the shear harmonic coefficients in harmonic space. Measuring the power spectrum using the weighted shear field leads to a biased estimation as the survey window couples different ℓ modes and introduces E mode contamination in B modes and vice versa. Similarly, the convolution of the pixel window Π and the weighted shear in real space corresponds to the harmonic coefficients of shear being multiplied by the Fourier transform of Π in the harmonic space. This results in multiplying the power spectrum measurements in harmonic space by the square of the Fourier transform of Π . Since the pixel window function is localized in real space, the Fourier transform of Π is unity for large scales $\ell \lesssim \pi/\Delta\theta$ and decreases for small scales $\ell \gtrsim \pi/\Delta\theta$, where $\Delta\theta$ is the angular resolution (Jeong et al. 2014). This suppresses measurements of the power spectrum on smaller scales; we correct for this effect in the measurement by deconvolving the pixel window function in harmonic space. The measured power spectrum of this weighted map results in the pseudo- C_ℓ estimator (D_ℓ) and is related to the true power spectrum via a coupling matrix (Hivon et al. 2002) as

$$\hat{D}_\ell^{EE} = C_\ell^{(\Pi)} (M_{\ell\ell'}^+ C_{\ell'}^{EE} + M_{\ell\ell'}^- C_{\ell'}^{BB}) \quad (15)$$

$$\hat{D}_\ell^{BB} = C_\ell^{(\Pi)} (M_{\ell\ell'}^+ C_{\ell'}^{BB} + M_{\ell\ell'}^- C_{\ell'}^{EE}), \quad (16)$$

where $C_\ell^{(\Pi)}$ is the angular power spectrum of the spin-0 quantity Π and $M_{\ell\ell'}^\pm$ is the coupling matrix. The spin-2 field coupling matrix is defined as

$$M_{\ell\ell'}^\pm = \frac{2\ell'+1}{4\pi} \sum_{\ell''} C_{\ell''}^{(W^\gamma)} (2\ell''+1) \left(\frac{1 \pm (-1)^{\ell+\ell'+\ell''}}{2} \right) \times \begin{pmatrix} \ell & \ell' & \ell'' \\ 2 & -2 & 0 \end{pmatrix} \begin{pmatrix} \ell & \ell' & \ell'' \\ 2 & -2 & 0 \end{pmatrix} \quad (17)$$

where $C_{\ell''}^{(W^\gamma)}$ is the angular power spectrum of the survey window and $\begin{pmatrix} \ell & \ell' & \ell'' \\ m & m' & m'' \end{pmatrix}$ is the Wigner 3- j symbol (or Clebsch–Gordan coefficients). The $M_{\ell\ell'}^- C_{\ell'}^{EE/BB}$ terms correspond to the leakage introduced by the survey window. Throughout the paper, we use the Einstein summation convention to imply summation over repeated indices.

Additionally, one must account for the noise bias in the estimated power spectrum. The shape noise generates this bias due to the

intrinsic ellipticities of galaxies. For a windowed field, the noise is effectively multiplied by the window given by $\sqrt{W^\gamma(\mathbf{n})}$ (see appendix B of Singh et al. 2020) for derivation of the window and the noise effects in pseudo- C_ℓ estimators). We analytically estimate the constant shape noise power spectrum as:

$$D_\ell^{(N)} = M_{\ell\ell'}^N N_{\ell'} = \sum_{\ell'} M_{\ell\ell'}^N \frac{\sigma_\ell^2}{\bar{n}_{\text{eff}}}, \quad (18)$$

where σ_ℓ is the shape noise, \bar{n}_{eff} is the effective number of galaxies in inverse steradians, and $M_{\ell\ell'}^N$ is the coupling matrix from using the window $\sqrt{W^\gamma(\mathbf{n})}$. We use the superscript N to denote the coupling matrix for the noise in the shear field as the signal and noise have different windows. We will drop $C_\ell^{(\Pi)}$ and N_ℓ to simplify our notation, though we do correct for the pixel window effect and the additive noise bias on the measurements.

In summary, this estimator introduces an artificial mixing of E - and B -modes due to the survey window. If uncorrected, this leads to systematic errors in cosmological parameter estimation.

3.1.3 Accounting for scale cuts

The pseudo- C_ℓ measurements in equations (15) and (16) are not matching with traditional TPCFs measurements, and the two two-point estimators are weakly correlated (e.g. Hamana et al. 2020; Doux et al. 2021). This is because TPCFs are measured over a limited θ range in real space, but this is not explicitly modeled in the pseudo- C_ℓ measurements. Therefore, the two measurements effectively use different sets of information that are probed by different ℓ ranges. As shown in Singh (2021), to make a matching two-point estimator in harmonic space with the TPCFs in real space, we need to account for the finite range of θ used in real space. Applying scale cuts in real space corresponds, in practice, to multiplying the TPCFs by a θ -dependent function denoted as $\nu(\theta)$. Mathematically, this operation is represented as $\hat{\xi}_\pm(\theta) = \xi_\pm(\theta)\nu(\theta)$, where the scale cut $\nu(\theta)$ is typically a step function defined over a finite range of scales.

Just as the product of the survey window in the shear map transformed into a convolution of the shear harmonic coefficients in the harmonic space, taking the spherical transform of TPCFs weighted by the scale cut $\nu(\theta)$ results in a convolution of power spectra in harmonic space. Since the pseudo- C_ℓ measurement in equations (15) and (16) do not include this effect, we manually convolve the measured \hat{D}_ℓ with a new coupling matrix, $M_{\ell\ell'}^{\pm,\nu}$, treating $\nu(\theta)$ as the window in equation (17), and define F_ℓ as the convolved pseudo- C_ℓ . We name this new power spectrum as the consistent pseudo- C_ℓ , a power spectrum obtained by convolving the pseudo- C_ℓ with a coupling matrix induced by imposing scale cuts in real space as (Singh 2021)

$$\begin{aligned} \hat{F}_\ell^{EE} &= M_{\ell\ell'}^{+,\nu} \hat{D}_{\ell'}^{EE} + M_{\ell\ell'}^{-,\nu} \hat{D}_{\ell'}^{BB} \\ &= (M_{\ell\ell'}^{+,\nu} M_{\ell'\ell''}^{+,w} + M_{\ell\ell'}^{-,\nu} M_{\ell'\ell''}^{-,w}) C_{\ell''}^{EE} \\ &\quad + (M_{\ell\ell'}^{+,\nu} M_{\ell'\ell''}^{-,w} + M_{\ell\ell'}^{-,\nu} M_{\ell'\ell''}^{+,w}) C_{\ell''}^{BB} \\ &\equiv \mathcal{M}_{\ell\ell'}^+ C_{\ell'}^{EE} + \mathcal{M}_{\ell\ell'}^- C_{\ell'}^{BB} \end{aligned} \quad (19)$$

and similarly

$$\hat{F}_\ell^{BB} = \mathcal{M}_{\ell\ell'}^+ C_{\ell'}^{BB} + \mathcal{M}_{\ell\ell'}^- C_{\ell'}^{EE}, \quad (20)$$

where

$$\mathcal{M}_{\ell\ell'}^\pm \equiv (M_{\ell\ell'}^{+,\nu} M_{\ell'\ell''}^{\pm,w} + M_{\ell\ell'}^{-,\nu} M_{\ell'\ell''}^{\mp,w}). \quad (21)$$

We distinguish the coupling matrices using superscripts; superscript w refers to the coupling matrix for the survey window W^γ as in

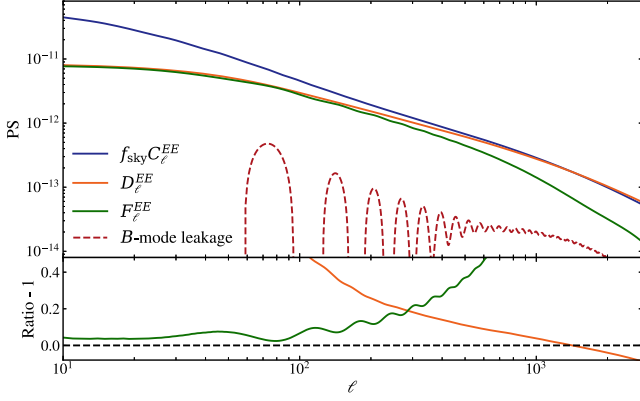


Figure 1. In the upper panel, we compare three different E -mode auto power spectra for the lowest tomographic bin (0.3, 0.6] of the fiducial cosmology defined in Section 4.2, using the HSC Y1 mask: $f_{\text{sky}} C_{\ell}^{EE}$ (blue) with normalization convention of the pseudo- C_{ℓ} that includes a factor of f_{sky} , D_{ℓ}^{EE} (orange), and F_{ℓ}^{EE} (green). The B -mode leakage of the form $M^{-sc} M^{-sw} C_{\ell}^{EE}$ in equation (21) is plotted as a dotted red line. We find that B -mode leakage contributes ~ 5 per cent to the E -mode consistent pseudo- C_{ℓ} . In the lower panel, we plot the ratios between the power spectrum and pseudo- C_{ℓ} (orange) and between pseudo- C_{ℓ} and consistent pseudo- C_{ℓ} (green). We find that information is lost on large scales when measuring the power spectrum on a weighted map, and power is recovered on small scales. Power is further lost on small scales when we model the impact of the scale cut used in real space in harmonic space.

equation (17) and ν refers to the coupling matrix with the scale cut function $\nu(\theta)$ acting as the window. The coupling matrix \mathcal{M} is a linear combination of the product of the two coupling matrices, representing the combined impact of the two effects on the observed power spectrum. In short, equations (19) and (20) take into account both the survey window and the scale cut used in the real space. See Appendix B for the derivation of F_{ℓ} from two-point correlation functions.

In Fig. 1, we compare three different E -mode model power spectra of the lowest redshift bin used in this work: the power spectrum C_{ℓ}^{EE} (blue), pseudo- C_{ℓ} D_{ℓ}^{EE} (orange), and consistent pseudo- C_{ℓ} F_{ℓ}^{EE} (green), and illustrate how each coupling matrix transforms one power spectrum to the other. We define the redshift bin used for this figure in Section 4.2. Going from the power spectrum to pseudo- C_{ℓ} , we find that power is greatly suppressed on large scales (small ℓ values) by an order of magnitude and is recovered on small scales (large ℓ values) as can be seen from the lower panel ratio plot. This is because the weighted maps exclude (include) many realizations of small (large) ℓ -modes. Comparing the orange and green curves, having a scale cut $\nu(\theta)$ in real space introduces further coupling between modes in harmonic space and has ~ 5 per cent effect on large scales and has $\gg 10$ per cent effect for small scales for our adopted scale cut $10 < \theta < 360$ [arcmin] (see Section 4.4 for details). Failure to account for this in harmonic space can bias power spectrum estimation on small scales. Scale cuts $\nu(\theta)$ are generally chosen to remove data points with significant evidence for observational systematic errors and/or model limitations. A smoother scale cut, as shown in equation (40) and illustrated in Fig. 2, is more optimal as it leads to convolution over fewer scales in ℓ . We also show the contribution of B -mode leakage into the E -mode of the consistent pseudo- C_{ℓ} , which leads to errors of order a few percent. Reconstruction of power spectra without correction for such B -mode leakage can lead to biased results. Implementation details to produce Fig. 1 are discussed in Sections 4.3 and 4.4.

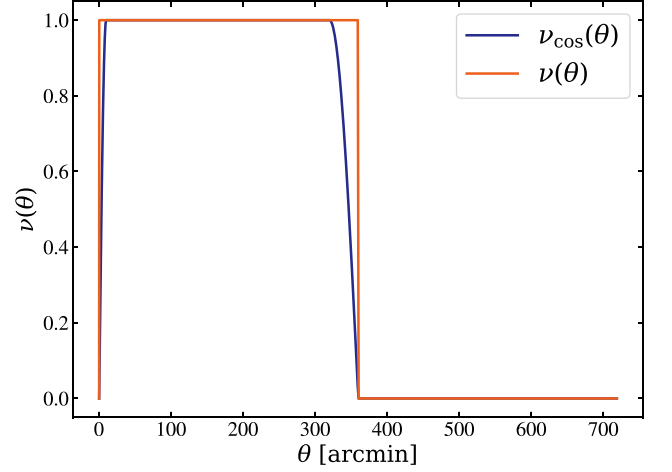


Figure 2. The scale cut imposed on the TPCFs measurements over angular scale. $\nu_{\text{cos}}(\theta)$ (blue line) is the tapered version of a flat top-hat (orange line) scale cut using equation (40) with $\theta_{\text{cut,min}} = 10$ arcmin and $\theta_{\text{cut,max}} = 320$ arcmin. The tapered version of the top-hat drops to zero more smoothly. For both scale cuts, the absolute minimum and maximum θ values are 0.25 and 360 arcmin.

3.1.4 Binning in ℓ

In general, each coupling matrix $M_{\ell\ell'}$ is a rectangular matrix of order $O(\ell_{\text{max}}^2)$ where $\ell_{\text{max}} \sim 5000$ – 7000 or even higher. In order to reduce the dimensionality and effects of noise, we bin both the measurement F_{ℓ} and the product of coupling matrices $\mathcal{M}_{\ell\ell'}$. We use the notation convention in Singh (2021) and define the binned coupling matrix as

$$\begin{aligned} F_{\ell_b}^{EE} &= \mathcal{B}_{\ell_b, \ell}^F F_{\ell}^{EE} \\ &= \mathcal{B}_{\ell_b, \ell}^F \mathcal{M}_{\ell\ell'}^+ \left(\mathcal{B}_{\ell, \ell_b}^{C^{EE}} \right)^{-1} \mathcal{B}_{\ell_b, \ell'}^{C^{EE}} C_{\ell'}^{EE} \\ &\quad + \mathcal{B}_{\ell_b, \ell}^F \mathcal{M}_{\ell\ell'}^- \left(\mathcal{B}_{\ell, \ell_b}^{C^{BB}} \right)^{-1} \mathcal{B}_{\ell_b, \ell'}^{C^{BB}} C_{\ell'}^{BB} \\ &= \mathcal{M}_{\ell_b \ell_b'}^{+EE} C_{\ell_b}^{EE} + \mathcal{M}_{\ell_b \ell_b'}^{-BB} C_{\ell_b}^{BB} \end{aligned} \quad (22)$$

and similarly

$$F_{\ell_b}^{BB} = \mathcal{M}_{\ell_b \ell_b'}^{+BB} C_{\ell_b}^{BB} + \mathcal{M}_{\ell_b \ell_b'}^{-EE} C_{\ell_b}^{EE}, \quad (23)$$

where \mathcal{B} is the binning operator defined in this work as

$$\mathcal{B}_{\ell_b, \ell} = \begin{cases} \frac{2\ell+1}{(2\ell_b+1)\Delta\ell}, & \ell \in b \\ 0, & \text{otherwise,} \end{cases} \quad (24)$$

where $\ell \in b$ is true when ℓ belongs to the bin centered on ℓ_b and $\Delta\ell$ is the bin size. The binned coupling matrices $\mathcal{M}_{\ell_b \ell_b'}$ with the superscript, e.g. $+EE$, refer to the power spectra used in defining the \mathcal{B} operator. The same binning operator can be used for F_{ℓ} and C_{ℓ} as long as $\ell_{\text{max}} \sim \ell'_{\text{max}}$, so the coupling matrix is not too wide. We assume that the noise is estimated and subtracted from the measurements before binning. The binning operator in equation (24) is essentially weighting each ℓ mode by the effective number of modes. The choice of the binning operator is flexible and the *i*MASTER algorithm works with any binning operator as long as the same binning definition is used for all quantities. See Singh (2021) for more optimal binning choices.

Since binning data leads to loss of information, we generally cannot invert the binning operator. However, given a good model for the data, we can define the inverse of the binning operator as

(Singh 2021)

$$(\mathcal{B}_{\ell, \ell_b}^{CEE})^{-1} = \begin{cases} \frac{C_{\ell}^{EE}}{C_{\ell_b}^{EE}} & \text{if } \ell \in b \\ 0 & \text{otherwise} \end{cases}, \quad (25)$$

where $C_{\ell_b}^{EE} = \mathcal{B}_{\ell, \ell_b}^{CEE} C_{\ell}^{EE}$ is obtained after binning the model C_{ℓ}^{EE} . The same inverse binning operator can be used for different power spectra as long as the two power spectra have the same slope.

From $F_{\ell_b}^{EE}$ and $F_{\ell_b}^{BB}$ with equations (22) and (23), we can reconstruct the underlying E/B -mode power spectrum as

$$\hat{C}_{\ell_b}^{EE} = (\mathcal{M}_{\ell_b \ell_b'}^{+EE})^{-1} \hat{F}_{\ell_b}^{EE} - (\mathcal{M}_{\ell_b \ell_b'}^{+EE})^{-1} \mathcal{M}_{\ell_b \ell_b''}^{-BB} C_{\ell_b''}^{BB} \quad (26)$$

$$\hat{C}_{\ell_b}^{BB} = (\mathcal{M}_{\ell_b \ell_b'}^{+BB})^{-1} \hat{F}_{\ell_b}^{BB} - (\mathcal{M}_{\ell_b \ell_b'}^{+BB})^{-1} \mathcal{M}_{\ell_b \ell_b''}^{-EE} C_{\ell_b''}^{EE}. \quad (27)$$

In equations (26) and (27), we also distinguished between the quantities measured from the data, denoted with $\hat{\cdot}$, and the ones estimated from theory. We are using the model $C_{\ell_b}^{EE/BB}$ to subtract out the leakage contribution. If the model is not well known and the noise is sub-dominant, we can use an iterative method where the initial estimates of $\hat{C}_{\ell_b}^{EE/BB}$ in equations (26) and (27) can be used to replace the model $C_{\ell_b}^{EE/BB}$ in the next iteration (see discussion in section 4.2.1 of Singh 2021). Hereafter, we will assume that the B -mode model power spectrum is zero, i.e. $C_{\ell_b}^{BB} = 0$, and consider only the E -mode theory to be non-zero. However, in practice, we still measure the pseudo- B -mode power spectrum induced by E -modes leaking into them.

3.2 Correlation function

In cosmic shear, the TPCFs measure the correlation of galaxy shapes given some angular separation θ . The TPCFs can be measured from the shear catalogue as

$$\hat{\xi}_{\pm}(\theta) = \frac{\sum_{ij} \mathcal{W}(\mathbf{n}_i) \mathcal{W}(\mathbf{n}_j) \hat{\gamma}_1(\mathbf{n}_i) \hat{\gamma}_1(\mathbf{n}_j)}{\sum_{ij} \mathcal{W}(\mathbf{n}_i) \mathcal{W}(\mathbf{n}_j)} \pm \frac{\sum_{ij} \mathcal{W}(\mathbf{n}_i) \mathcal{W}(\mathbf{n}_j) \hat{\gamma}_2(\mathbf{n}_i) \hat{\gamma}_2(\mathbf{n}_j)}{\sum_{ij} \mathcal{W}(\mathbf{n}_i) \mathcal{W}(\mathbf{n}_j)}, \quad (28)$$

where the summation is over every galaxy pair (i, j) with their angular separation $\theta = |\mathbf{n}_i - \mathbf{n}_j|$, \mathcal{W} is the galaxy shape weight, and $\hat{\gamma}_1, \hat{\gamma}_2$ are the two measured shear components.

Traditional TPCFs estimators return TPCFs that are average weighted by the number of pairs at each angular bin. However, the number of galaxies in each angular bin can be noisy with many bins, and the TPCFs estimator defined in equation (28) does not fully remove the effects of the survey window. This is due to the effects of separately binning the numerator and denominator in equation (28). This way, the window effect is accounted for imperfectly, which leads to residual percent-level window effects when using traditional TPCFs estimators (see appendix D of Singh et al. 2020 for demonstration of residual percent-level window effects on correlation functions). Rather than trying to use the above estimator and also identify and apply those additional window corrections, we take the alternate approach with a modified estimator from Singh (2021) that is matching with the pseudo- C_{ℓ} defined in equations (15) and (16), and is motivated by the weighting scheme in equation (14). Using an estimator that explicitly accounts for the survey window ensures *matching* results with harmonic space analyses.

In this new TPCFs estimator, we divide the measurement from equation (28) by the mean number of galaxies in each angular bin based on the survey window instead of the measured number of pairs,

and use $\xi_W^{(\nu)}(\theta)$ to denote the survey window correlation function. Effectively, multiplying the TPCFs measurements with $\xi_W^{(\nu)}(\theta)$ in real space corresponds to convolving the power spectrum with the coupling matrix in harmonic space, where the Hankel transform of $\xi_W^{(\nu)}(\theta)$ is the angular power spectrum of the survey window in equation (17). We define the survey window correlation function as

$$\xi_W^{(\nu)}(\theta) = \frac{\sum_{ij} \mathcal{W}(\mathbf{n}_i) \mathcal{W}(\mathbf{n}_j)}{f_{\text{sky}} N_{\text{source}}^2 \int_{\theta-\Delta\theta/2}^{\theta+\Delta\theta/2} 2\pi \sin \theta' d\theta'}, \quad (29)$$

where f_{sky} is the fraction of the sky covered, N_{source} is the number of source galaxies in the tomographic bin, and $\Delta\theta$ is the width of the angular bins. The denominator in equation (29) corresponds to the expected number of pairs in each bin if we have a full sky homogeneous survey window, rescaled by f_{sky} to bring the normalization closer to the survey expectation (see fig. D1 of Singh et al. 2020 for an illustration of the survey window function). The numerator is based on weighted sums of pairs and the denominator is based on unweighted sums. However, as long as we use a self-consistent definition of the window in the measurements, model, and covariance, this inconsistency does not affect the results. Therefore, our new binned TPCFs are defined as (Singh 2021)

$$\hat{\xi}_{\pm, w}(\theta) = \hat{\xi}_{\pm}(\theta) \xi_W^{(\nu)}(\theta). \quad (30)$$

In general, if possible, one should keep the data and model in the same space to avoid the complexities of transforming to the Fourier counterpart. Since the theory directly predicts the power spectrum in harmonic space, we reconstruct the cosmic shear power spectrum from the TPCFs measurements.

3.3 Going from the TPCFs to the power spectrum

In this section, we describe how to consistently go from measured TPCFs in real space described in Section 3.2 to power spectrum in harmonic space. This transformation includes both the survey window and the scale cut used in both real and harmonic space.

As briefly mentioned in Section 3.1.2, in practice, we measure the correlation function over a limited θ range as

$$\hat{\xi}_{\pm, w, \text{cut}}(\theta) = \hat{\xi}_{\pm, w}(\theta) \nu(\theta) = \hat{\xi}_{\pm}(\theta) \xi_W^{(\nu)}(\theta) \nu(\theta), \quad (31)$$

where $\xi_W^{(\nu)}(\theta)$ is the correlation function of the window and $\nu(\theta)$ is the scale cut function, usually a top-hat function defined over $\theta \in [\theta_{\min}, \theta_{\max}]$. Multiplying the correlation function with the survey window correlation function and scale cut in real space results in convolution in harmonic space as (Singh 2021),

$$(H_{\ell, \theta}^{\pm})^{-1} \hat{\xi}_{\pm, w, \text{cut}}(\theta) = \mathcal{Q}_{\ell \ell'}^{+} C_{\ell'}^{EE} \pm \mathcal{Q}_{\ell \ell'}^{-} C_{\ell'}^{BB} = F_{\ell}^{\pm} \quad (32)$$

where $\mathcal{Q}_{\ell \ell'}$ is the coupling matrix with window power spectrum computed by taking the Fourier transform of the product $\xi_W^{(\nu)}(\theta) \nu(\theta)$ in equation (17). Note that equation (32) is the generalization of equation (7). In summary, to account for both the survey window and the scale cut multiplied in real space, one can convolve the power spectrum once with $\mathcal{Q}_{\ell \ell'}^{\pm}$, or convolve the power spectrum twice with $\mathcal{M}_{\ell \ell'}$ in equation (21).

Here, we assume that $\hat{\xi}_{+}$ and $\hat{\xi}_{-}$ are measured over the same θ range. Note that when the θ range is small, its corresponding coupling matrix can be rather broad, requiring a larger ℓ range to make the power spectrum unbiased. To reduce the complexity of the coupling matrix, one can apodize the scale cut $\nu(\theta)$. Singh (2021) shows a simple apodizing scheme where one can multiply the power spectrum of $\nu(\theta)$ with a cosine function that goes from one to zero over some ℓ range and transform back to obtain a new scale cut $\nu'(\theta)$

that brings some power from θ outside of its initial range and softens the edges of the top hat. In practical applications, the apodizing scheme will depend on the measurements being performed and the scales being used.

We can further bin quantities similar to equations (22) and (23) and define a binned inverse spherical transform as

$$(H_{\theta_b, \ell_b}^\pm)^{-1} = \mathcal{B}_{\ell_b, \ell}^\pm (H_{\theta, \ell}^\pm)^{-1} (\mathcal{B}_{\theta_b, \theta}^\pm)^{-1}, \quad (33)$$

where the inverse binning operator $(\mathcal{B}_{\theta_b, \theta}^\pm)^{-1}$ is defined similar to equation (25) using the ratio $\xi_\pm(\theta)/\xi_\pm(\theta_b)$. We can obtain the E/B -mode of F_ℓ as

$$F_{\ell_b}^{EE} = \frac{1}{2}(F_{\ell_b}^+ + F_{\ell_b}^-) \quad (34)$$

$$F_{\ell_b}^{BB} = \frac{1}{2}(F_{\ell_b}^+ - F_{\ell_b}^-). \quad (35)$$

Then, using equations (26) and (27), we can reconstruct the underlying cosmic shear power spectrum. By accounting for the survey window and scale cuts, our method ensures that the reconstructed power spectrum is *matching*, allowing for consistent cosmic shear analysis.

3.4 Going from the power spectrum to the TPCFs

Sometimes it is desired to transform the model defined in harmonic space into real space to compare with the data, as many cosmological models written in harmonic space have well-defined scale cuts. Similar to the issues we addressed in the case of TPCFs, multiplying the power spectrum with a scale cut in harmonic space should result in a convolution on the TPCFs in real space as

$$\xi_{\pm, \ell\text{-cut}} = H_{\theta, \ell}^\pm C_\ell b_\ell = b_{\theta\theta'} \xi_\pm(\theta'), \quad (36)$$

where b_ℓ is the truncation function in harmonic space and $b_{\theta\theta'}$ is coupling matrix defined as

$$b_{\theta\theta'} = \sum_\ell b_\ell \frac{2\ell + 1}{4\pi} {}_2Y_{\ell, 2}(\theta) {}_2Y_{\ell, 2}(\theta'). \quad (37)$$

Similar to how we manually convolved the pseudo- C_ℓ to account for the scale cut used in real space, we can simply convolve the measured TPCFs with $b_{\theta\theta'}$ to impose the scale cut used in harmonic space before running the inference chain. By properly handling scale-dependent effects, we ensure that real space and harmonic space measurements provide *matching* constraints on cosmological parameters.

4 DATA AND IMPLEMENTATION

In this section, we briefly introduce the HSC survey (Section 4.1) and describe the mock catalogues used to test our new analysis methods (Section 4.2). We discuss the details of the cosmic shear power spectrum measurements; we present the implementation details of pseudo power spectra in Section 4.3 and TPCFs in Section 4.4. We additionally describe the covariance matrix estimation in Section 4.5.

4.1 Hyper Suprime-Cam Survey

The Hyper Suprime-Cam Subaru Strategic Programme is a comprehensive imaging survey conducted using the 8.2 m Subaru Telescope and the Hyper Suprime-Cam wide-field camera (see Aihara et al. 2018 for survey design). The HSC survey consists of three layers: Wide, Deep, and UltraDeep. Each layer of the survey is observed in five broad-band filters (*grizy*) with a point source 5σ depth of

$r \sim 26$ mag. The combination of depth and excellent seeing (typical *i*-band seeing of ~ 0.59 arcseconds) allows HSC to measure cosmic shear signals up to higher redshifts with lower shape noise than other Stage III imaging surveys.

The HSC Year 1 shear catalogue (HSC-Y1) is based on data from the HSC S16A internal data release, consisting of data taken between 2014 March to 2016 April with about 90 nights. A number of cuts were made to construct a robust shape catalogue for weak lensing analysis, including a conservative magnitude cut of $i < 24.5$, removing bad pixels, and masking regions of the sky around bright stars. The HSC-Y1 catalogue covers a total area of 136.9 deg^2 that consists of six disjoint fields: XMM, GAMA09H, GAMA15H, HECTOMAP, VVDS, and WIDE12H (see fig. 1 in Mandelbaum et al. 2018 for more details). The shapes of galaxy are estimated on the *i*-band coadded images using the re-Gaussianization shape measurement method (Hirata & Seljak 2003). The method defines the two components of the galaxy ellipticity as

$$(e_1, e_2) = \frac{1 - (b/a)^2}{1 + (b/a)^2} (\cos 2\phi, \sin 2\phi), \quad (38)$$

where b/a is the observed minor-to-major axis ratio and ϕ is the position angle of the major axis with respect to the equatorial coordinate system. The two components of the shear are then estimated using the measured ellipticity as

$$\hat{\gamma}_\alpha = \frac{1}{2\mathcal{R}} \langle e_\alpha \rangle, \quad (39)$$

where $\alpha = 1, 2$ are the indices for the two spinor components and the shear responsivity \mathcal{R} is the response of an ensemble of galaxies ellipticity to a small shear distortion (Kaiser, Squires & Broadhurst 1995; Bernstein & Jarvis 2002).

4.2 HSC-Y1 mock shape catalogues

In this work, we use the HSC-Y1 mock shape catalogues and verify that our new analysis method produces unbiased measurement of the underlying true tomographic cosmic shear power spectrum. These mock catalogues are generated following the method described in Shirasaki et al. (2019) using the full-sky ray-tracing N -body simulations (Takahashi et al. 2017). In short, real HSC galaxies are populated to the simulated lensing field according to their measured positions on the transverse plane, and their redshift is randomly drawn from the estimated MLZ photometric redshift posterior, where MLZ is a machine-learning code based on a self-organizing map for photometric redshift (photo- z) inference. Therefore, the mock catalogs have the same survey geometry as the real HSC Y1 data. The six disjoint patches with holes in area coverage due to masking bright stars make it challenging to model the survey window W^γ in equation (14), making the HSC-Y1 survey geometry a realistically challenging testbed for validating our new analysis method. As described in Shirasaki et al. (2019), the shear catalogue galaxies are divided into four tomographic redshift bins, in the intervals $(0.3, 0.6]$, $(0.6, 0.9]$, $(0.9, 1.2]$, and $(1.2, 1.5]$, using the ‘best’ photo- z estimation by the MLZ algorithm.

The mock catalogues have galaxy positions in RA and Dec., true shear γ_1 and γ_2 . We use the true shear value for our $F_\ell^{EE/BB}$ and ξ_\pm measurements and use them to estimate the covariance matrix as described in Section 4.5. Since we use the true shear values for our measurements, the noise power spectrum N_ℓ in equation (18) is set to zero.

The mock catalogs adopt a standard flat lambda cold dark matter (Λ CDM) cosmological model with CDM density parameter

$\Omega_c = 0.233$, the baryon density $\Omega_b = 0.046$, the matter density $\Omega_m = \Omega_c + \Omega_b = 0.279$, the cosmological constant $\Omega_\Lambda = 0.721$, the Hubble parameter $h = 0.7$, the amplitude of density fluctuations $\sigma_8 = 0.82$, and the spectral index $n_s = 0.97$. We refer to this as the fiducial model.

Before testing our method on the HSC-Y1 mock shape catalogue, we first run our method on Gaussian mocks. We generate a set of Gaussian mocks using the HEALPY (Górski et al. 2005) package with the fiducial model power spectra and resolution parameter $N_{\text{side}} = 8192$ corresponding to a pixel size of ~ 0.43 arcmin. Each Gaussian mock shares the same footprint as the HSC-Y1 mock shape catalogue, allowing us to test our method for providing an unbiased reconstruction of the fiducial model.

4.3 Power spectra from pseudo power spectra

We use the pseudo- C_ℓ method described in Section 3.1 to reconstruct the tomographic cosmic shear power spectra from the HSC Y1 shape mock catalogs. We measure the power spectra and mode coupling matrices to larger ℓ than we wish to use for our analysis to correctly account for mode coupling between smaller scale modes and to account for the E/B -mode leakage. The shear maps of each tomographic bin of the mocks were constructed using HEALPY (Górski et al. 2005) following equation (12) with a resolution parameter $N_{\text{side}} = 4096$, corresponding to a pixel size of ~ 0.86 arcmin. We measure the tomographic full-sky pseudo power spectra up to $\ell_{\text{max}} = 8000$, and the same ℓ_{max} was used to generate the coupling matrices $M_{\ell\ell'}^v$ and $M_{\ell\ell'}^w$ in equation (17). Before reconstructing the power spectra from the pseudo power spectra, we correct for the pixel window function by dividing the power spectra in equations (15) and (16) by the angular power spectrum of the pixel window function; $C_\ell^{(N)} \simeq \text{sinc}(\ell\Delta\theta/2\pi)$ is the pixel window function of HEALPIX (Jeong et al. 2014), where $\Delta\theta$ is the angular resolution. We then convolve the pseudo power spectra with a coupling matrix derived from the scale cut (equations 19 and 20) to estimate the consistent pseudo- C_ℓ and bin each spectrum and coupling matrices as described by the binning operator in equation (24). Equations (26) and (27) are then used to reconstruct the EE and BB spectra for each tomographic bin pair.

We reconstruct EE and BB spectra in 20 linearly spaced multipole bins between $\ell_{\text{min}} = 20$ and $\ell_{\text{max}} = 5980$, and restrict to a narrower range and adapt the HSC Y1 analysis multiple ranges $300 < \ell < 1800$ (Hikage et al. 2019) when doing cosmological inference. Thus, this gives us 7 multipole bins that we use for our cosmological inference, from the 20 that we originally measure. For each mock, we have 10 spectra, four auto- and six cross-power spectra, and our data vector then consists of $10 \times 7 = 70$ data points.

4.4 Power spectra from TPCFs

We use the public software TREECORR (Jarvis, Bernstein & Jain 2004)¹ to measure TPCFs. We use the TPCF method described in Section 3.2 to reconstruct the tomographic cosmic shear power spectra from the TPCF measurements. For each tomographic bin pair, we measure the TPCFs in 2400 logarithmically spaced angular bins between 0.25 and 720 arcmin. While $\nu(\theta)$ is typically chosen to be a top-hat function (default scale cut used for TREECORR), it can lead to convolution over rather large scales in ℓ . We use the scale-cut

function that drops more smoothly from one to zero (see section 4.2.2 of Singh 2021)

$$\nu_{\cos}(\theta, \theta_{\text{cut,min}}, \theta_{\text{cut,max}} | \theta_{\text{min}}, \theta_{\text{max}}) = \begin{cases} 0, & \theta \leq \theta_{\text{min}} \text{ and } \theta \geq \theta_{\text{max}} \\ \cos\left(\frac{\pi}{2} \frac{\theta - \theta_{\text{cut,min}}}{\theta_{\text{cut,max}} - \theta_{\text{cut,min}}}\right), & \theta_{\text{cut,min}} < \theta < \theta_{\text{cut,max}} \\ 1, & \theta_{\text{cut,min}} \leq \theta \leq \theta_{\text{cut,max}} \\ \cos\left(\frac{\pi}{2} \frac{\theta - \theta_{\text{cut,max}}}{\theta_{\text{max}} - \theta_{\text{cut,max}}}\right), & \theta_{\text{cut,max}} < \theta < \theta_{\text{max}} \end{cases}, \quad (40)$$

where $\nu_{\cos}(\theta)$ uses a cosine function to smoothly truncate the power spectra to zero between $\theta_{\text{max}} - \theta_{\text{cut,max}}$ and $\theta_{\text{cut,min}} - \theta_{\text{min}}$ with $\{\theta_{\text{min}}, \theta_{\text{cut,min}}, \theta_{\text{cut,max}}, \theta_{\text{max}}\} = \{0.25 \text{ arcmin}, 10 \text{ arcmin}, 320 \text{ arcmin}, 360 \text{ arcmin}\}$ for the analysis. We first adopted the range of scales used in HSC Y1 and extended the range to reduce the complexity of the coupling matrix. A larger separation between $\theta_{\text{cut,max}}$ and $\theta_{\text{cut,min}}$ will lead to narrower convolution in ℓ space. The values of $\theta_{\text{cut,min}}$ and $\theta_{\text{cut,max}}$ were chosen as the Fourier transform of $\nu_{\cos}(\theta)$ drops to zero smoothly. We do not optimize $\nu_{\cos}(\theta)$ in this work. In Fig. 2, we compare the modified version of a top-hat scale cut with the aforementioned parameter values. The measured TPCFs are then multiplied with $\nu_{\cos}(\theta)$.

Once we have our TPCFs, we multiply by the survey window correlation function to match with the power spectra measurements, resulting in measurements as described in equation (31). After binning the measurements and the inverse spherical transform operator, equations (34) and (35) are used to get the pseudo- D_ℓ , in which we reconstruct the power spectra using equations (26) and (27). We reconstruct the EE and BB spectra in the same 20 linearly equal ℓ bins defined in 4.3, resulting in 70 data points for the data vector for each mock.

4.5 Covariance matrix

We estimate the covariance matrix of our estimated cosmic shear power spectrum using mock catalogues. We do so by measuring the cosmic shear from 571 realizations of the mocks and utilize these 571 to compute the covariance matrix.

The lensing covariance matrix can be decomposed into three terms: the Gaussian, connected non-Gaussian, and the super-sample covariance (Barreira, Krause & Schmidt 2018). The non-Gaussian terms are not modelled as the covariance matrix estimation is done on Gaussian mocks. The Gaussian covariance term of the cosmic shear power spectrum can be further decomposed into three terms: the auto-term of cosmic shear, the cross-term of cosmic shear and shape noise spectra, and the auto-term of the shape noise spectra as

$$\text{Cov}^{(G)}(\hat{C}_{\ell_1}^{ij}, \hat{C}_{\ell_2}^{mn}) = \text{Cov}_{SS}^{(G)} + \text{Cov}_{SN}^{(G)} + \text{Cov}_{NN}^{(G)}, \quad (41)$$

where G denotes the Gaussian part of the covariance, i, j, m, n are the tomographic bins, S denotes the cosmic shear, and N denotes the shape noise. Since we make the measurements on noiseless data, our estimation of the covariance from noiseless measurements only estimates the auto-term of the cosmic shear in equation (41). Neglecting the shape noise contribution to the covariance can lead to adding more weights to small scales, which are generally more biased from the model.

To get the full Gaussian covariance, we model the cross-term between cosmic shear and shape noise spectra and the auto-term of the shape noise spectra as

¹<https://github.com/rmjarvis/TreeCorr>

$$\text{Cov}_{SN}^{(G)}(\hat{C}_{\ell_1}^{ij}, \hat{C}_{\ell_2}^{mn}) = \frac{\delta_{\ell_1 \ell_2}}{f_{\text{sky}}(2\ell_1 + 1)\Delta\ell_1} \left[\frac{\sigma_e^2}{\bar{n}_{\text{eff}}^i} \left(C_{\ell_1}^{jn} \delta_{im} + C_{\ell_1}^{jm} \delta_{in} \right) + \frac{\sigma_e^2}{\bar{n}_{\text{eff}}^j} \left(C_{\ell_1}^{im} \delta_{jn} + C_{\ell_1}^{in} \delta_{jm} \right) \right] \quad (42)$$

$$\text{Cov}_{NN}^{(G)}(\hat{C}_{\ell_1}^{ij}, \hat{C}_{\ell_2}^{mn}) = \frac{\delta_{\ell_1 \ell_2}}{f_{\text{sky}}(2\ell_1 + 1)\Delta\ell_1} \frac{\sigma_e^4}{\bar{n}_{\text{eff}}^i \bar{n}_{\text{eff}}^j} (\delta_{im} \delta_{jn} + \delta_{in} \delta_{jm}), \quad (43)$$

where $\sigma_e = 0.25$ is the RMS ellipticity of the source galaxies and we use values of \bar{n}_{eff}^i from table 1 column $n_{g,\text{eff}}^{(C13)}$ in Hikage et al. (2019). The Kronecker deltas δ ensure that the Gaussian covariance is non-vanishing when ℓ_1 and ℓ_2 belong in the same ℓ bin and the shape noise spectra contribute only to matching tomographic bins. Note that our analysis uses the true lensing shear values in the mock catalogs, so there is no shape noise in the data vectors. However, we use the above covariance (including the cross-term between cosmic shear and shape noise and the auto-term of the shape noise) so that the relative weight given to the bins in ℓ in our analysis is similar to what would be used in a realistic case with real data. By applying this covariance to the data vectors without shape noise, we obtain a precise estimate of any biases due to our method of power spectrum reconstruction with a realistic weighting across ℓ bins.

5 RESULT

In this section, we discuss the result of our analysis of HSC Y1 mock shape catalogs described in Section 4.2. In Section 5.1, we show the result of power spectrum reconstruction from consistent pseudo- C_ℓ and TPCFs measurements. In Section 5.2, we describe the outcome of fitting a cosmological model to the reconstructed power spectrum and performing cosmological parameter inference. In Section 5.3, we discuss the scalability of the *i*MASTER algorithm.

5.1 Cosmic shear power spectrum reconstruction

Here, we present the reconstruction of the cosmic shear power spectrum from two-point function measurements in both real and harmonic space over multipole ranges specified in Sections 4.3 and 4.4.

Before running our method on the HSC mocks, we first test the accuracy of the reconstruction method on Gaussian simulations with an HSC footprint as described in Section 4.2. Our analysis method on the simple Gaussian simulations gives unbiased reconstructions of the power spectrum (one from the consistent pseudo- C_ℓ and the other from TPCFs), and the two reconstructions agree within 1 per cent for $300 < \ell < 1800$.

Running the method ignoring the second term in equations (22) and (23), or ignoring the leakage of the B mode when constructing F_ℓ , gives a biased reconstruction of the power spectrum and loss of the signal by 15–20 per cent for our survey window and scale cuts. Testing our method on Gaussian mocks, we conclude that our method gives unbiased reconstruction of the power spectrum on a survey geometry as complex as HSC Y1, and that B -mode leakage correction is needed given the scale cuts imposed in HSC.

In Fig. 3, we show the mean of the tomographic cosmic shear power spectrum reconstructed from both consistent pseudo- C_ℓ (blue dots) and TPCFs measurements (orange dots) performed on HSC mocks with error bars. We show the auto-power spectra for the lowest redshift bin and show the ratio of the reconstruction to the forward-modelled theory for the rest. The two reconstructions agree

within ± 1 per cent with each other and within ± 2 per cent with the forward-modelled theory in the ℓ ranges used in the analysis. However, the two reconstructions start to deviate from the fiducial model at higher ℓ values; fixing ℓ_{max} when measuring pseudo- C_ℓ power spectra and the coupling matrices can miss some of the power leakage across scales, biasing the reconstructed power spectra. This effect is largest for higher ℓ as those scales miss the most leakage. These biases can be reduced by increasing ℓ_{max} , with increased computational cost. There may also be some bias from the imperfect power spectra model used in the reconstruction, and these biases can be reduced by decreasing the size of the bin (see discussion in Singh 2021).

In Fig. 4, we show the comparison of the integrated signal-to-noise ratio (S/N) for tomographic cosmic shear up to some ℓ value. The S/N is defined as

$$S/N(q) = \sqrt{\hat{C}_\ell(q) \Sigma^{-1}(q) \hat{C}_\ell^T(q)}, \quad (44)$$

where $q \in \{\text{PS}, \text{TPCF} \rightarrow \text{PS}\}$ represents the direction of the analysis and Σ is the covariance matrix estimated from the mocks. The two integrated S/N of the reconstructions are in agreement within ± 1 per cent. The S/N of the reconstructed power spectrum from the consistent pseudo- C_ℓ in individual redshift bins are 4.86, 9.32, 11.68, and 10.09 from the lowest to highest redshift bins, respectively. The S/N in these redshift bins are consistent with those of the HSC Y1 power spectrum and correlation function analyses to within 20 per cent. The S/N from mocks is expected to be less than that of real data, as real data contains noise that artificially biases the measured signal to noise ratio to higher values. To quantify the covariance of two reconstructions, we compute the correlation coefficient as

$$r(x) = \frac{\text{Cov}(x_R, x_F)}{\text{Cov}(x_R, x_R)^{1/2} \text{Cov}(x_F, x_F)^{1/2}}, \quad (45)$$

with $x = \hat{C}_\ell$. The subscripts R and F stand for the reconstruction from TPCFs (real space) and consistent pseudo- C_ℓ (harmonic space), respectively. We find that the correlation coefficient range is $r(\hat{C}_\ell) \in [0.91, 1.0]$, thus the two reconstructions have the same information and are matching. For a detailed review, refer to Appendix A.

5.2 Likelihood inference

Once we obtain measurements of the cosmic shear power spectrum using the method described in Sections 4.3 and 4.4, the next stage of the analysis is the likelihood analysis. In this step, the tomographic cosmic shear power spectrum data vector is compared with a theoretical data vector computed using a forward model based on cosmological parameters.

To obtain a single parameter quantifying the fidelity of power spectrum reconstruction, we constrain the cosmological parameter S_8 by comparing the reconstructed power spectrum data vector measured in Sections 4.3 and 4.4 to model-predicted power spectrum. A higher-dimensional parameter inference is possible with this new estimator and is left for future work; we deliberately simplify in this case so the interpretation in terms of biases in the cosmic shear power spectrum amplitude is unambiguous. When computing the theory cosmic shear power spectrum in equation (3), we use the publicly available code CAMB (Lewis & Challinor 2011)² to compute the linear part of the matter power spectrum and use PYHALOFIT³ to compute the non-linear matter power spectrum. To properly account for the shell

²<https://github.com/cmbant/CAMB>

³<https://github.com/git-sunao/pyhalofit>

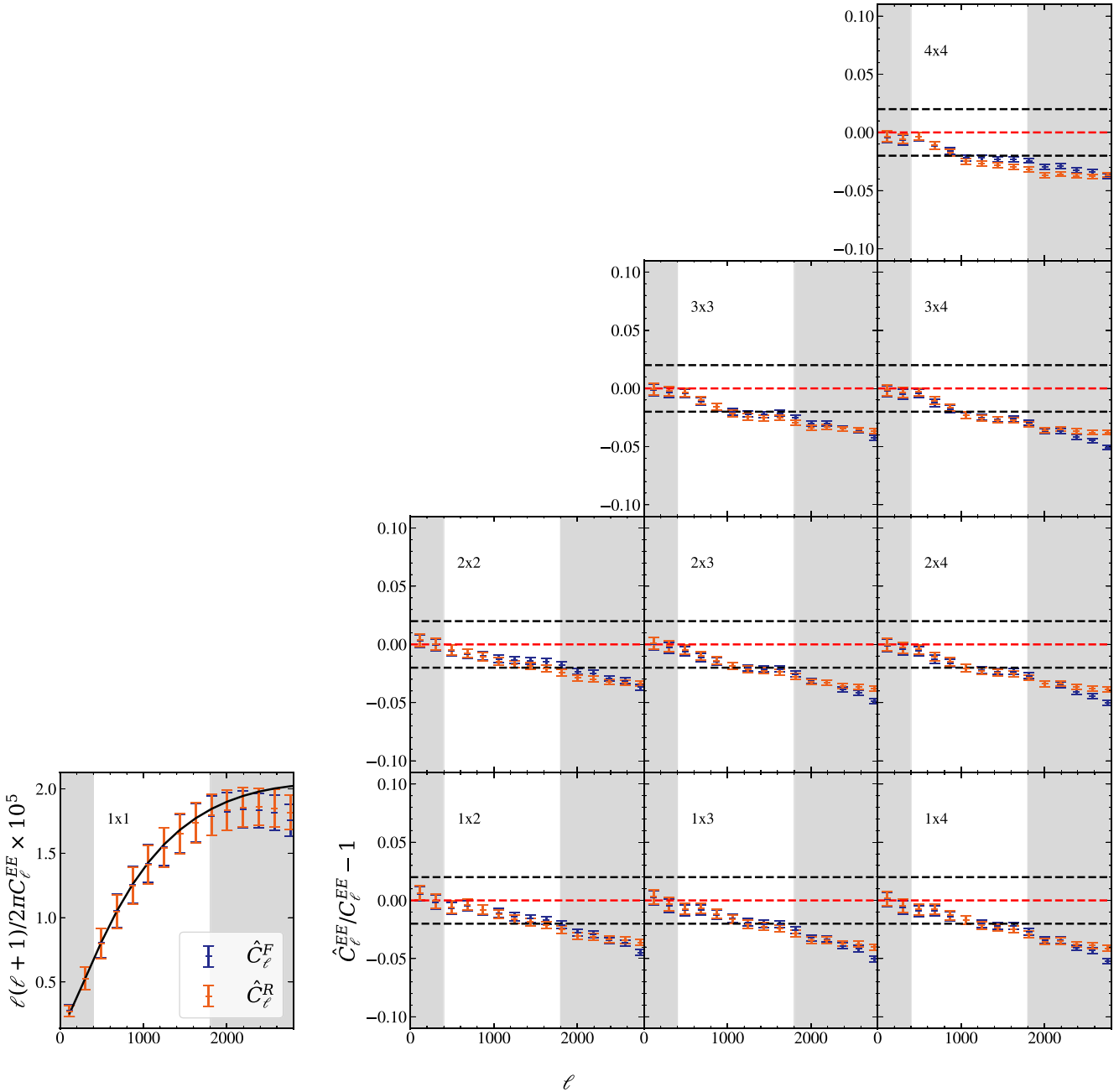


Figure 3. Comparison of the reconstructed tomographic cosmic shear power spectra of EE mode from pseudo power spectra and TPCFs measurements with error bars in the lower left panel. The colored points show the average reconstructed power spectrum from 571 mock realizations from pseudo power spectra (blue dots) and two-point correlation functions (orange dots) for the lowest redshift bin, as indicated by the 1×1 label. The analytical forward-modelled prediction (solid black line) is also shown for comparison. The scales outside of $400 < \ell < 1800$ (shaded regions) are excluded from the inference. The other panels show the ratio of each reconstruction to the forward-modelled theory, with error bars. The error bars are 1σ uncertainties in the ratio estimated using mock catalogs. Reconstructions of power spectra from *i*MASTER give matching results to well within 1 per cent and they give unbiased results to within 2 per cent. Black dotted lines represent the 2 per cent bias region. The scales outside of the region $400 < \ell < 1800$ (shaded regions) are excluded from the cosmological analysis.

thickness effect and finite resolution effect in the full-sky ray-tracing simulations of the mock catalogue, we modify the matter power spectrum as described in Shirasaki et al. (2019). We then use the Limber approximation following equation (3) to get the tomographic cosmic shear power spectra with the matter power spectrum. After obtaining the tomographic cosmic shear power spectra, we forward model them using the methods described in Sections 3.1 and 3.2 to incorporate the effects of the survey window and scale cut.

For each realization of the mock, we define the χ^2 of our model in terms of the Gaussian log-likelihood as:

$$\begin{aligned} \chi^2(q|\Theta) &= -2 \log \mathcal{L}(\hat{C}_{\ell_b}(q)|\Theta) \\ &= (\hat{C}_{\ell_b}(q) - \tilde{C}_{\ell_b}(q|\Theta))^T \Sigma^{-1} (\hat{C}_{\ell_b}(q) - \tilde{C}_{\ell_b}(q|\Theta)), \end{aligned} \quad (46)$$

where χ^2 comes from the concatenated tomographic cosmic shear power spectra data vector and $q \in \{\text{PS}, \text{TPCF} \rightarrow \text{PS}\}$. \hat{C}_{ℓ_b} is the reconstructed power spectrum from the mocks and \tilde{C}_{ℓ_b} is obtained

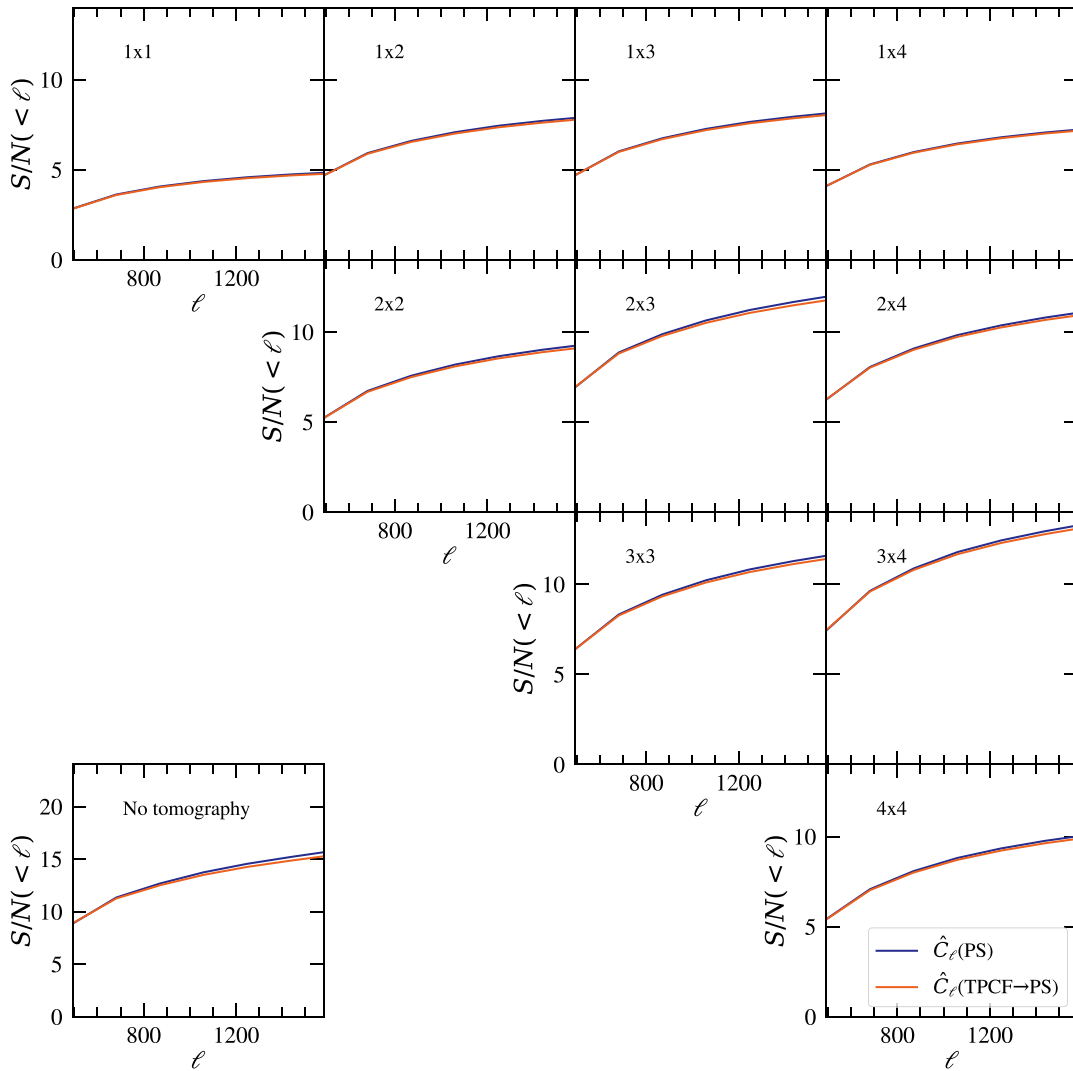


Figure 4. The integrated S/N of the reconstructed auto and cross power spectra from consistent pseudo- C_ℓ (blue) and two-point correlation function (orange) measurements as a function of ℓ . The left-bottom panel shows the non-tomographic S/N of the reconstructed power spectrum. The two integrated S/N of the reconstructions are in agreement within ± 1 per cent. The S/N in these redshift bins are consistent with those of the HSC Y1 auto-power spectrum to within 20 per cent.

after forward modelling the model cosmic shear power spectrum. We ignore the B -mode power spectrum during the inference as our reconstructed and model B -mode are both zero, i.e. $C_\ell^{BB} = \hat{C}_\ell^{BB} = 0$.

Motivated by fig. 17 of Hamana et al. (2020), we use a simple grid-based parameter estimation method to infer S_8 , starting from both real and harmonic space measurements and reconstructing the power spectrum. We pre-compute the model power spectrum from a grid space of parameters of $\Theta = \{S_8\}$ and $S_8 \in [0.4, 1.2]$ with a bin size of $\Delta S_8 = 0.008$. Then, for each mock catalogue, we find the value of S_8 on the grid that maximizes the log-likelihood or minimizes χ^2 as:

$$S_8^i(q) = \arg \min_{S_8 \in [0.4, 1.2]} \chi_i^2(q|S_8), \quad (47)$$

where S_8^i is the inferred S_8 value using the i th mock catalogue. We compare the inferred S_8 values using the reconstructed cosmic shear power spectrum from consistent pseudo- C_ℓ and TPCFs measurements on the same 571 mock catalogues in the scatter plot in Fig. 5.

We use equation (45) to quantify the covariance of the two inferred S_8 values, and we find that $r(S_8) = 0.994$, which confirms that the two S_8 values are strongly correlated. Hamana et al. (2020) did a similar test in which they inferred two parameters (S_8 and Ω_m) separately from TPCFs and power spectrum measurements. The difference between their analysis and our analysis is that we reconstructed the power spectrum from the TPCF measurements when running the inference. They find only weak correlations between the inferred cosmological parameters from the two analyses: $r(S_8) = 0.51$ and $r(\Omega_m) = 0.17$. The reason for such disparity between the two analyses is due to the different multipole ranges probed in the analyses; they showed that the scale cut $\nu(\theta)$ used in the HSC Y1 analysis has a large contribution from scales with $\ell < 300$, which were excluded in the pseudo- C_ℓ analysis, leading to the two analyses using different and complementary information. A direct comparison of the two $r(S_8)$ values is challenging as Hamana et al. (2020) performed a multidimensional parameter inference while marginalizing over nuisance parameters. We show that we get a

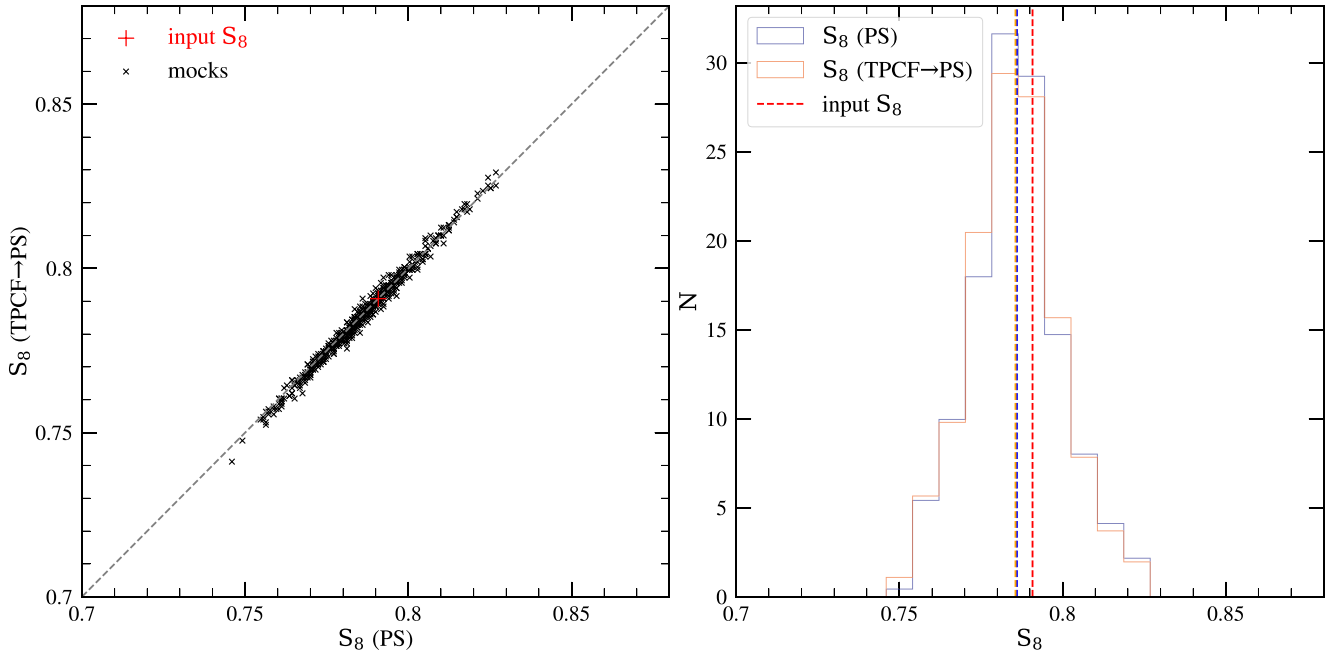


Figure 5. Left panel: scatter plot showing the inferred S_8 values from consistent pseudo- C_ℓ analysis in the horizontal axis and from two-point correlation functions analysis in the vertical axis. The points are from minimizing χ^2 for 571 mock catalogues. The red + denotes the value of S_8 adopted in generating the mock catalogues. The inferred S_8 values between the two independent analyses are very consistent, with the correlation coefficient $r = 0.994$. While the comparison is conceptually similar to fig. 17 of Hamana et al. 2020, which highlights the inconsistency between different summary statistics, we emphasize that their analysis was performed on noisy mock data, whereas our analysis uses noiseless simulations. As such, the results are not directly comparable. Right panel: we show the distribution of S_8 from the mocks for each axes. The dotted vertical lines represent the mean of each distribution, with the red dotted vertical line denoting the fiducial S_8 value. The two distributions have very similar means, with $\Delta\langle S_8 \rangle = 0.0004$.

consistent parameter inference once we account for this scale cut used in real space in the consistent pseudo- C_ℓ measurement.

The right panel in Fig. 5 shows the distribution of inferred S_8 values from reconstruction using consistent pseudo- C_ℓ (blue) and TPCFs (orange) measurements, with the red dashed line showing the fiducial value of S_8 . The mean of each distribution is 0.7860 and 0.7856 with standard deviations of 0.0137 and 0.0143, respectively. The difference between the mean and median values of the inferred values of S_8 , δS_8 , are 0.0004 and 0, respectively. This difference is less than the observed difference of -0.042 for HSC Y1 by two orders of magnitude. The means of both inferred S_8 values are consistent with the fiducial value to well within 0.7 per cent. Our results suggest that the method as implemented here can successfully provide a self-consistent representation of the weak lensing power spectrum starting from pseudo- C_ℓ in harmonic space or from the two-point correlation functions in real space, and residual subpercent biases (consistent across both methods) may be due to factors outside of this method. One possible source of residual biases is the imperfect correction for simulation resolution effects in the mock catalogues.

5.3 Scalability of *i*MASTER

One of the advantages of using *i*MASTER method is that it speeds up the computation during sampling by reducing the complexity of the coupling matrix to $O(N_{\text{bin}}^2)$ instead of $O(\ell_{\text{max}}^2)$ by binning quantities, where $N_{\text{bin}} \sim O(10)$. Assuming that the survey window is fixed, we only have to compute the coupling matrices $M_{\ell\ell'}^{\pm,w}$ and $M_{\ell\ell'}^{\pm,v}$ once, and they are used to get the consistent pseudo- C_ℓ . Since survey window estimation may have uncertainties, coupling matrices may need to be recomputed for multiple windows (see Singh 2021;

Karim et al. 2023). Since we propose to convolve the pseudo- C_ℓ with a coupling matrix from scale-cuts in real space, the complexity becomes $O(N_{\text{bin}}^3)$ from the matrix multiplication.

6 CONCLUSION

In this paper, we presented the first practical application of the *i*MASTER algorithm (Singh 2021) to mock catalogues with a realistic cosmological density field and a realistically complex survey window (the HSC Y1 mock shape catalogues). In doing so, we reviewed the formalism for the measurement of cosmic shear two-point functions and how to correctly account for survey window and scale cuts in harmonic and real space to obtain a *matching* power spectrum.

In Section 3, we discussed the impact of the survey window and scale cut on the various estimators. Applying the survey window and scale cuts in real space generally involves multiplying the underlying maps by some weights, which are equivalent to convolutions in harmonic space. Since measurements of the power spectrum with non-uniform coverage of the sky do not account for the scale cut used in real space, we manually convolved the pseudo- C_ℓ to obtain the consistent pseudo- C_ℓ . We showed that this greatly affects the pseudo- C_ℓ measurements in all scales. We also showed that leakage of the form $M^{-,sc} M^{-,sw} C_\ell^{EE}$, equivalent to B -mode leakage of F_ℓ^{EE} , contributes 5 per cent to the F_ℓ^{EE} measurements. We also derived a new two-point correlation function (TPCF) estimator that is matching with the process of estimating the pseudo- C_ℓ . In particular, the new estimator normalizes the measurements by the expected number of galaxy pairs instead of observed number of galaxy pairs, which is traditionally done in the literature. This new procedure changes the

two-point correlation function measurements by ~ 2 per cent and allows for self-consistent modelling in the two spaces.

We also showed methods to reconstruct the E/B mode power spectra from pseudo power spectra and two-point correlation function measurements. In this paper, we assumed that ξ_{\pm} are measured over the same range of scales. If one wishes to use different scales for ξ_{+} and ξ_{-} , one must generalize equations (34) and (35).

In Section 4, we discussed the HSC Y1 shape mock catalogs used for the analysis and the implementation details of the cosmic shear power spectrum measurements. The reconstruction of the power spectra from TPCFs measurements can be challenging if the range of scales at which TPCFs are measured is limited, leading to a complex coupling matrix. We introduced a new scale-cut that drops to zero smoothly; this smoothed scale-cut would lead to a narrower convolution in ℓ space and reduce the complexity of the coupling matrix. However, we do not optimize the scale-cut in this work.

In Section 5, we showed the result of applying the i MASTER algorithm using the HSC Y1 shape mock catalogue. Repeating the scale cuts used in both harmonic and real space in the actual HSC Y1 analysis, i MASTER algorithm was able to reconstruct the cosmic shear power spectrum from $F_{\ell}^{E/B}$ and ξ_{\pm} that match within ± 1 per cent with each other. We also showed that the two integrated S/N of the reconstructions are matching within ± 1 per cent. We compared the inferred S_8 values using likelihood inference with the reconstructed cosmic shear power spectrum. We found that the two inferred S_8 values from 571 mocks to be strongly correlated with a correlation coefficient of 0.994.

This algorithm has more capabilities than existing algorithms (e.g. the MASTER and NAMASTER algorithms) as it enables direct comparisons with unbinned theoretical models calculated at effective bin centres. Additionally, our algorithm is efficient and consistent in that it allows for more precise extraction of information for a specific set of scales by effectively removing the mode mixing in the consistent pseudo- C_{ℓ} estimator, and we expect it will be useful for cosmic shear data analysis.

The i MASTER implementation provides options to speed up computations after the initial setup (computing coupling matrices). The computational complexity is reduced from $O(\ell_{\max}^2)$ to $O(N_{\text{bin}}^2)$, and similar scaling for computation of covariance. This could be important for a large cross-correlation analysis or memory management during a large analysis, such as an LSST-like 3×2 analysis. This method provides a systematic means of integrating the scale cut applied in one space into another. This will be of particular importance for analysing forthcoming data with improved constraining power. Future works should test the performance of this method using more complicated simulations and real data. Here we list targets of future work:

- (i) Testing the performance of this method using real cosmic shear data from existing surveys (e.g. HSC).
- (ii) Extending this method to cosmological analyses that include scale cuts and survey windows (e.g. galaxy clustering, galaxy–galaxy lensing, CMB lensing, and 3×2 pt)
- (iii) Understanding the impact of window biases on the reconstruction of the power spectrum. Window biases lead to both additive and multiplicative biases in the power spectrum.

The future work outlined here will get this very promising method ready for direction application to Stage-IV surveys.

ACKNOWLEDGEMENTS

This work was performed using the Vera cluster at the McWilliams Center for Cosmology at Carnegie Mellon University, operated by the Pittsburgh Supercomputing Center facility. The authors would like to thank Scott Dodelson, Yuuki Omori, Rupert Croft, and Michael Troxel for many implementation discussions of the method and feedback on this work and thank Axel Guinot for thorough insightful comments on the draft of this paper. We thank Shirasaki Masato for making the HSC Y1 shape mock catalogue public.

We thank the maintainers of NUMPY (Harris et al. 2020), JAX (Bradbury et al. 2018), MATPLOTLIB (Hunter 2007), and HEALPIX (Górski et al. 2005) for their excellent open-source software.

AP was partially supported by a McWilliams Center for Cosmology seed grant (PI: Singh). SS was supported by the McWilliams postdoctoral fellowship at CMU. XL, RM, and TZ were supported in part by a grant from the Simons Foundation (Simons Investigator in Astrophysics, Award ID 620789) and in part by the Department of Energy grant DE-SC0010118.

DATA AVAILABILITY

This work used the public HSC Y1 mock shape catalogue, which can be found at <https://hsc-release.mtk.nao.ac.jp/doc/index.php/s16a-shape-catalog-data-products-pdr2/>. No new data were generated in support of this work. The software used for the computations of Wigner-3j and coupling matrices from the survey window and the scale cut is available at <https://github.com/sukhdeep2/SkyLens/tree/skylens>.

REFERENCES

- Aihara H. et al., 2018, *PASJ*, 70, S4
Akeson R. et al., 2019, preprint (arXiv:1902.05569)
Alonso D., Sanchez J., Slosar A., *LSST Dark Energy Science Collaboration*, 2019, *MNRAS*, 484, 4127
Asgari M. et al., 2021, *A&A*, 645, A104
Bacon D. J., Refregier A. R., Ellis R. S., 2000, *MNRAS*, 318, 625
Barreira A., Krause E., Schmidt F., 2018, *J. Cosmology Astropart. Phys.*, 2018, 053
Becker M. R., Rozo E., 2016, *MNRAS*, 457, 304
Bernstein G. M., Armstrong R., 2014, *MNRAS*, 438, 1880
Bernstein G. M., Jarvis M., 2002, *AJ*, 123, 583
Bradbury J. et al., 2018, JAX: Composable Transformations of Python+NumPy Programs, version = {0.3.13}. Available at: <http://github.com/google/jax>
Brown M. L., Castro P. G., Taylor A. N., 2005, *MNRAS*, 360, 1262
Camacho H. et al., 2022, *MNRAS*, 516, 5799
Dalal R. et al., 2023, *Phys. Rev. D*, 108, 123519
Dark Energy Survey Collaboration, 2016, *MNRAS*, 460, 1270
Doux C. et al., 2021, *MNRAS*, 503, 3796
Doux C. et al., 2022, *MNRAS*, 515, 1942
Feldman H. A., Kaiser N., Peacock J. A., 1994, *ApJ*, 426, 23
Fenech Conti I., Herbonnet R., Hoekstra H., Merten J., Miller L., Viola M., 2017, *MNRAS*, 467, 1627
Górski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, *ApJ*, 622, 759
Hamana T. et al., 2020, *PASJ*, 72, 16
Hamilton A. J. S., 1997, *MNRAS*, 289, 285
Hamimeche S., Lewis A., 2008, *Phys. Rev. D*, 77, 103013
Harris C. R. et al., 2020, *Nature*, 585, 357
Hikage C., Takada M., Hamana T., Spergel D., 2011, *MNRAS*, 412, 65
Hikage C. et al., 2019, *PASJ*, 71, 43
Hirata C., Seljak U., 2003, *MNRAS*, 343, 459

- Hivon E., Górski K. M., Netterfield C. B., Crill B. P., Prunet S., Hansen F., 2002, *ApJ*, 567, 2
- Hu W., 2001, *Phys. Rev. D*, 65, 023003
- Huff E., Mandelbaum R., 2017, preprint (arXiv:1702.02600)
- Hunter J. D., 2007, *Comput. Sci. Eng.*, 9, 90
- Huterer D., 2010, *Gen. Relativ. Gravit.*, 42, 2177
- Ivezić Ž. et al., 2019, *ApJ*, 873, 111
- Jarvis M., Bernstein G., Jain B., 2004, *MNRAS*, 352, 338
- Jeong D., Chluba J., Dai L., Kamionkowski M., Wang X., 2014, *Phys. Rev. D*, 89, 023003
- de Jong J. T. A. et al., 2017, *A&A*, 604, A134
- Kaiser N., Squires G., Broadhurst T., 1995, *ApJ*, 449, 460
- Kaiser N., Wilson G., Luppino G. A., 2000, *ApJ*, preprint (arXiv:astro-ph/0003338)
- Karim T., Rezaie M., Singh S., Eisenstein D., 2023, *MNRAS*, 525, 311
- Kilbinger M., 2015, *Rep. Prog. Phys.*, 78, 086901
- LSST Science Collaboration, 2009, preprint (arXiv:0912.0201)
- Landy S. D., Szalay A. S., 1993, *ApJ*, 412, 64
- Laureijs R. et al., 2011, preprint (arXiv:1110.3193)
- Lewis A., Challinor A., 2011, Astrophysics Source Code Library, record ascl:1102.026
- Lewis A., Challinor A., Turok N., 2001, *Phys. Rev. D*, 65, 023505
- Li X., Mandelbaum R., 2023, *MNRAS*, 521, 4904
- Li X., Katayama N., Oguri M., More S., 2018, *MNRAS*, 481, 4445
- Li X., Li Y., Massey R., 2022, *MNRAS*, 511, 4850
- Li X. et al., 2023, *Phys. Rev. D*, 108, 123518
- Limber D. N., 1953, *ApJ*, 117, 134
- Loureiro A. et al., 2022, *A&A*, 665, A56
- Mandelbaum R. et al., 2018, *PASJ*, 70, S25
- Massey R., Refregier A., 2005, *MNRAS*, 363, 197
- Myles J. et al., 2021, *MNRAS*, 505, 4249
- Ng K.-W., Liu G.-C., 1999, *Int. J. Mod. Phys. D*, 08, 61
- Rau M. M. et al., 2023, *MNRAS*, 524, 5109
- Refregier A., Bacon D., 2003, *MNRAS*, 338, 48
- Rhodes J., Refregier A., Collins N. R., Gardner J. P., Groth E. J., Hill R. S., 2004, *ApJ*, 605, 29
- Schneider P., van Waerbeke L., Jain B., Kruse G., 1998, *MNRAS*, 296, 873
- Schneider P., van Waerbeke L., Mellier Y., 2002, *A&A*, 389, 729
- Schneider P., Eifler T., Krause E., 2010, *A&A*, 520, A116
- Schneider P. et al., 2022, *A&A*, 664, A77
- Schrabback T. et al., 2010, *A&A*, 516, A63
- Secco L. F. et al., 2022, *Phys. Rev. D*, 105, 023515
- Shapiro C., 2009, *ApJ*, 696, 775
- Sheldon E. S., Huff E. M., 2017, *ApJ*, 841, 24
- Sheldon E. S., Becker M. R., MacCrann N., Jarvis M., 2020, *ApJ*, 902, 138
- Shirasaki M., Hamana T., Takada M., Takahashi R., Miyatake H., 2019, *MNRAS*, 486, 52
- Singh S., 2021, *MNRAS*, 508, 1632
- Singh S., Mandelbaum R., Seljak U., Rodríguez-Torres S., Slosar A., 2020, *MNRAS*, 491, 51
- Takahashi R., Hamana T., Shirasaki M., Namikawa T., Nishimichi T., Osato K., Shiroshima K., 2017, *ApJ*, 850, 24
- Van Waerbeke L. et al., 2001, *A&A*, 374, 757
- Wandelt B. D., Hivon E., Górski K. M., 2001, *Phys. Rev. D*, 64, 083003
- Weinberg D. H., Mortonson M. J., Eisenstein D. J., Hirata C., Riess A. G., Rozo E., 2013, *Phys. Rep.*, 530, 87
- Wittman D. M., Tyson J. A., Kirkman D., Dell'Antonio I., Bernstein G., 2000, *Nature*, 405, 143
- Zaldarriaga M., Scoccimarro R., 2003, *ApJ*, 584, 559
- Zhang J., 2008, *MNRAS*, 383, 113
- Zuntz J., Kacprzak T., Voigt L., Hirsch M., Rowe B., Bridle S., 2013, *MNRAS*, 434, 1604

APPENDIX A: CROSS-CORRELATION BETWEEN TWO RECONSTRUCTIONS

In addition to showing the consistency between the two inferred values of S_8 , we also show the consistency in the reconstruction of the power spectrum in Fig. A1. Each block shows the value of $r(\hat{C}_\ell)$ defined in equation (45). We find that all four blocks have homogeneous correlation coefficients, confirming the consistency between the two reconstructions of the power spectrum.

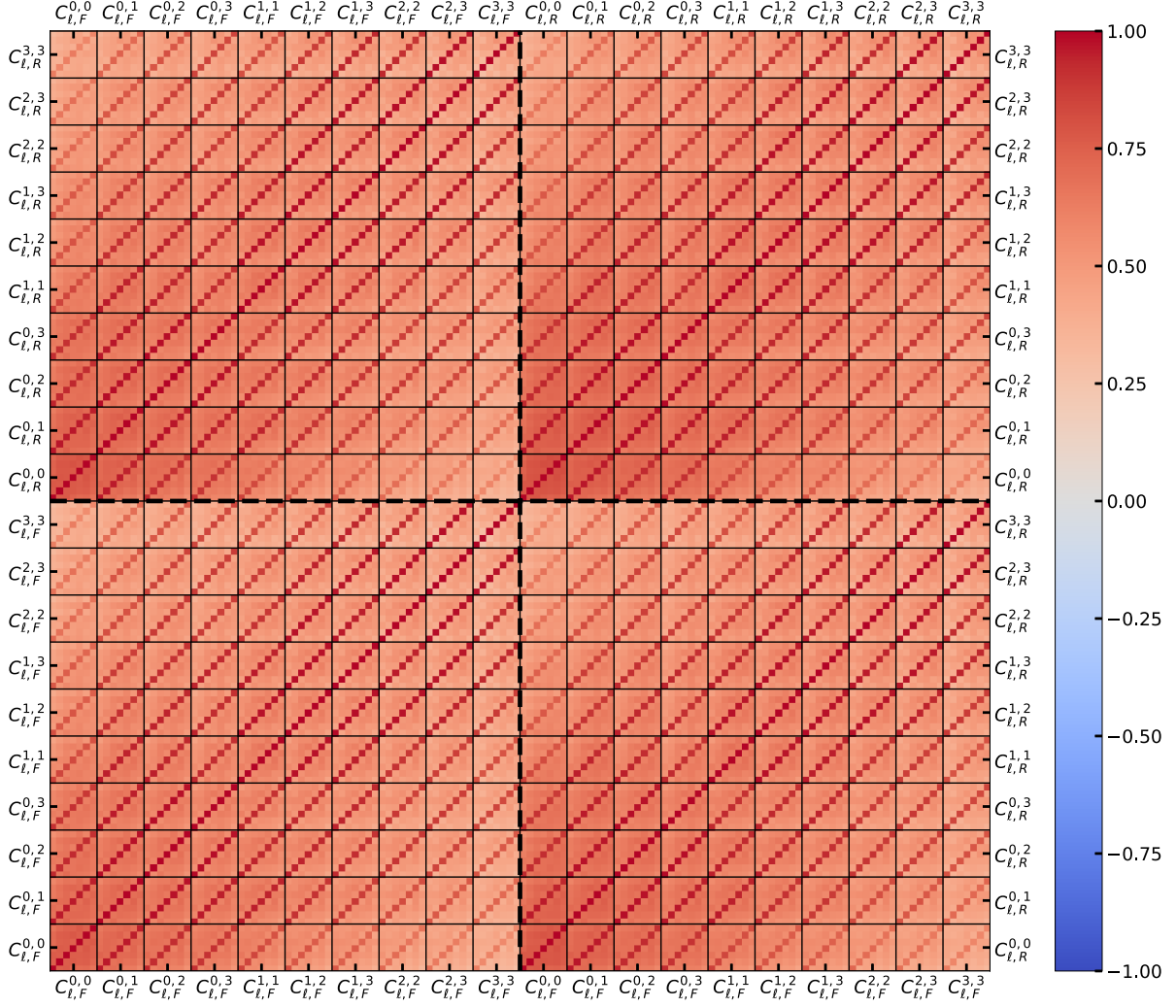


Figure A1. Cross-correlation (normalized covariance) between the two reconstructions where the lower-left and upper-right correlation blocks show the correlation matrix between two tomographic redshift bins within one reconstruction, and the lower-right and upper-left correlation blocks, which are identical, show the cross-correlation between the two reconstructions (using $x = \hat{C}_\ell$ in equation (45)).

APPENDIX B: SCALE CUTS AND SURVEY WINDOW

Here, we show the derivation of the inverse spherical transform of two-point correlation functions measured in the presence of a survey window and a limited range of scales. This leads to the pseudo- D_ℓ described in equations (19) and (20). Applying the inverse of the spherical transform in equation (6) gives

$$F_\ell^{EE/BB} = \frac{1}{2} \int d\theta \, 2\pi \sin\theta \left[{}_2d_{\ell 2}^*(\theta) \xi_+(\theta) \xi_W^{(\gamma)}(\theta) \nu(\theta) \pm {}_2d_{\ell 2}^*(\theta) \xi_-(\theta) \xi_W^{(\gamma)}(\theta) \nu(\theta) \right], \quad (\text{B1})$$

where ${}_s d_{\ell m}$ is the Wigner-d matrix, with superscript * denoting complex conjugate and $s = m = 2$ for spin-2 shear field, $\xi_W^{(\gamma)}(\theta)$ is the survey window correlation function defined in equation (30), and $\nu(\theta)$ is the scale cut. Replacing the Wigner-d matrix with spin-

weighted spherical harmonics (see discussions in Ng & Liu 1999 and the appendix in Singh 2021), we get

$$\begin{aligned} F_\ell^{EE/BB} &= \frac{1}{2} \sqrt{\frac{4\pi}{2\ell+1}} \int d\Omega \left[{}_2Y_{\ell-2}^*(\theta, 0) \xi_+(\theta) \xi_W^{(\gamma)}(\theta) \nu(\theta) \right. \\ &\quad \left. \pm {}_2Y_{\ell 2}^*(\theta, 0) \xi_-(\theta) \xi_W^{(\gamma)}(\theta) \nu(\theta) \right] \quad (\text{B2}) \\ &= \frac{1}{2} \sqrt{\frac{4\pi}{2\ell+1}} \int d\Omega \sum_{\ell'\ell''} \sqrt{\frac{2\ell'+1}{4\pi}} \sqrt{\frac{2\ell''+1}{4\pi}} C_{\ell''}^{(\nu)} \\ &\quad \times \left\{ (D_{\ell'}^{EE} + D_{\ell'}^{BB}) {}_2Y_{\ell 2}(\theta) {}_2Y_{\ell'-2}(\theta) {}_0Y_{\ell'' 0}(\theta) \right. \\ &\quad \left. \pm (D_{\ell'}^{EE} - D_{\ell'}^{BB}) {}_2Y_{\ell-2}(\theta) {}_2Y_{\ell' 2}(\theta) {}_0Y_{\ell'' 0}(\theta) \right\}, \quad (\text{B3}) \end{aligned}$$

where $C_{\ell''}^{(\nu)}$ is the angular power spectrum of the scale cut $\nu(\theta)$. We used the identity ${}_s Y_{\ell m}^* = (-1)^{s+m} {}_s Y_{\ell -m}$ and the fact that the inverse spherical transform of $\xi_\pm(\theta) \xi_W^{(\gamma)}(\theta)$ in the integrand is the pseudo- D_ℓ as shown in Appendix E of Singh (2021). Integrating the product of

three spin-weighted spherical harmonics, we get

$$\begin{aligned}
 F_{\ell}^{EE/BB} &= \frac{1}{2} \sqrt{\frac{4\pi}{2\ell+1}} \sum_{\ell'\ell''} \sqrt{\frac{2\ell'+1}{4\pi}} \sqrt{\frac{2\ell''+1}{4\pi}} C_{\ell''}^{(v)} \\
 &\times \left\{ D_{\ell'}^{EE/BB} \sqrt{\frac{(2\ell+1)(2\ell'+1)(2\ell''+1)}{4\pi}} \right. \\
 &\left[\begin{pmatrix} \ell & \ell' & \ell'' \\ 2 & -2 & 0 \end{pmatrix} \begin{pmatrix} \ell & \ell' & \ell'' \\ 2 & -2 & 0 \end{pmatrix} + \begin{pmatrix} \ell & \ell' & \ell'' \\ 2 & -2 & 0 \end{pmatrix} \begin{pmatrix} \ell & \ell' & \ell'' \\ -2 & 2 & 0 \end{pmatrix} \right] \\
 &+ D_{\ell'}^{BB/EE} \sqrt{\frac{(2\ell+1)(2\ell'+1)(2\ell''+1)}{4\pi}} \\
 &\left[\begin{pmatrix} \ell & \ell' & \ell'' \\ 2 & -2 & 0 \end{pmatrix} \begin{pmatrix} \ell & \ell' & \ell'' \\ 2 & -2 & 0 \end{pmatrix} - \begin{pmatrix} \ell & \ell' & \ell'' \\ 2 & -2 & 0 \end{pmatrix} \begin{pmatrix} \ell & \ell' & \ell'' \\ -2 & 2 & 0 \end{pmatrix} \right] \left. \right\}, \tag{B4}
 \end{aligned}$$

where we used

$$\begin{aligned}
 &\int d\Omega {}_{s_1}Y_{\ell m_1}(\theta) {}_{s_2}Y_{\ell' m_2}(\theta) {}_{s_3}Y_{\ell'' m_3}(\theta) \\
 &= \sqrt{\frac{(2\ell+1)(2\ell'+1)(2\ell''+1)}{4\pi}} \\
 &\times \begin{pmatrix} \ell & \ell' & \ell'' \\ -s_1 & -s_2 & -s_3 \end{pmatrix} \begin{pmatrix} \ell & \ell' & \ell'' \\ m_1 & m_2 & m_3 \end{pmatrix}. \tag{B5}
 \end{aligned}$$

Using the phase property of the Wigner-3j symbols

$$\begin{pmatrix} \ell & \ell' & \ell'' \\ -s_1 & -s_2 & -s_3 \end{pmatrix} = (-1)^{(\ell+\ell'+\ell'')} \begin{pmatrix} \ell & \ell' & \ell'' \\ s_1 & s_2 & s_3 \end{pmatrix} \tag{B6}$$

we derive the cosmic shear pseudo- D_{ℓ} as

$$\begin{aligned}
 F_{\ell}^{EE/BB} &= \sum_{\ell'} \frac{2\ell'+1}{4\pi} \sum_{\ell''} (2\ell''+1) C_{\ell''}^{(v)} \left(\frac{1+(-1)^{\ell+\ell'+\ell''}}{2} \right) \\
 &\begin{pmatrix} \ell & \ell' & \ell'' \\ 2 & -2 & 0 \end{pmatrix}^2 \left(D_{\ell'}^{EE/BB} + D_{\ell'}^{BB/EE} \right) \\
 &= \sum_{\ell'} M_{\ell\ell'}^+ D_{\ell'}^{EE/BB} + M_{\ell\ell'}^- D_{\ell'}^{BB/EE}, \tag{B7}
 \end{aligned}$$

where the spin-2 coupling matrices are defined in equation (17) with the power spectrum of survey window $C_{\ell''}^{(W'')}$ replaced with the power spectrum of the scale cut $C_{\ell''}^{(v)}$. If the full range of scales in ℓ and θ are used, the last equation is equivalent to showing that the power spectrum and two-point correlation function estimators are identical.

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