

A roadmap to cosmological parameter analysis with third-order shear statistics

IV. Analytic cross-covariance between second- and third-order aperture masses

Niek Wielders^{1,2,*}, Laila Linke^{1,3}, Pierre A. Burger^{1,4,5}, Sven Heydenreich^{1,6},
Lucas Porth¹, and Peter Schneider¹

¹ University of Bonn, Argelander-Institut für Astronomie, Auf dem Hügel 71, 53121 Bonn, Germany

² School of Mathematics, Statistics and Physics, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK

³ Universität Innsbruck, Institut für Astro- und Teilchenphysik, Technikerstr. 25/8, 6020 Innsbruck, Austria

⁴ Waterloo Centre for Astrophysics, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada

⁵ Department of Physics and Astronomy, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada

⁶ Department of Astronomy and Astrophysics, University of California, Santa Cruz, 1156 High Street, Santa Cruz, CA 95064, USA

Received 11 March 2025 / Accepted 25 August 2025

ABSTRACT

Context. Weak gravitational lensing is a powerful probe of cosmology, with second-order shear statistics commonly used to constrain parameters such as the matter density Ω_m and the clustering amplitude S_8 . However, degeneracies between parameters persist and can be broken by including higher-order statistics, such as the third-order aperture mass. To jointly analyse second- and third-order statistics, an accurate model of their cross-covariance is essential.

Aims. This work derives and validates a non-tomographic analytical model for the cross-covariance between second- and third-order aperture mass statistics. Analytical models are computationally efficient and enable cosmological parameter inference across a range of models, in contrast to numerical covariances derived from simulations or resampling methods, which are either costly or biased.

Methods. We derived the cross-covariance from real-space estimators of the aperture mass. Substituting the Halofit power spectrum, BiHalofit bispectrum, and a halo-model-based tetraspectrum, the model was validated against numerical covariances from the N -body Scinet LIghtCone Simulations (SLICS) using both shear catalogues and convergence maps. We performed a Markov chain Monte Carlo parameter analysis using both analytical and numerical covariances for several filter scale combinations.

Results. The cross-covariance separates into three terms governed by the power spectrum, bispectrum, and tetraspectrum, with the latter dominating. While the analytical model qualitatively reproduces simulation results, differences arise due to modelling approximations and numerical evaluation issues. The analytical contours are systematically tighter, with a combined figure of merit that is 72% that of the numerical case, increasing to 80% when small-scale information is excluded. These differences largely stem from an underprediction of the second-order covariance.

Conclusions. This work completes the analytical covariance framework for second- and third-order aperture mass statistics, enabling joint parameter inference without the need for large simulation suites. While further refinement is needed to improve quantitative accuracy, the model represents a key step towards simulation-independent cosmic shear analyses.

Key words. gravitational lensing: weak – methods: analytical – methods: statistical – cosmological parameters – large-scale structure of Universe

1. Introduction

The Lambda cold dark matter (ΛCDM) model is the prevailing framework for understanding the evolution of the Universe, from the post-inflationary epoch to the development of the large-scale structure (LSS) observed today. It is characterised by six key parameters that must be empirically determined. Observational efforts to constrain these parameters fall into two broad categories: those that examine the early Universe through measurements of the cosmic microwave background (CMB), such as those provided by Planck (Planck Collaboration VI 2020), and those that probe the late-time Universe, notably via the LSS.

A longstanding area of interest in recent years has been the apparent tension between CMB-based and LSS-based determinations of the clustering amplitude parameter $S_8 = \sigma_8 \sqrt{\Omega_m/0.3}$, where σ_8 quantifies the normalisation of the matter power spectrum and Ω_m denotes the total matter density. Several LSS surveys have reported lower values of S_8 compared to those inferred from the CMB, with tensions reaching 2–3 σ significance (Di Valentino et al. 2021; Joudaki et al. 2020; Asgari et al. 2021). However, recent results from the KiDS-Legacy analysis (Wright et al. 2025) have shown that this tension can be resolved through a more robust

* Corresponding author: niek.wielders@gmail.com

statistical treatment of the data, suggesting that the discrepancy does not require new physics, but rather a refined methodological framework.

Nevertheless, the potential of weak gravitational lensing as a tool for probing cosmology remains undiminished. Through the measurement of background galaxy distortions induced by the intervening matter distribution, weak lensing offers a direct and unbiased probe of the total matter content, encompassing both baryonic and dark components, without assumptions about galaxy bias. This makes it uniquely suited for constraining key cosmological parameters such as Ω_m and S_8 (σ_8).

In the past decade, weak lensing analyses have primarily employed second-order statistics, such as the two-point shear correlation function, which have substantially improved parameter constraints (Hilbrandt et al. 2017; DES Collaboration 2022; Heymans et al. 2021; Dalal et al. 2023). However, second-order measures alone are limited in their ability to fully characterise the matter distribution, particularly at late times. The Gaussian approximation valid in the early Universe breaks down due to non-linear structure formation, which induces non-Gaussian features such as high-density peaks corresponding to galaxy clusters. These complexities are not captured by second-order statistics alone.

Moreover, degeneracies, especially between Ω_m and σ_8 , remain a significant challenge when relying solely on two-point statistics (Takada & Jain 2004; Kilbinger & Schneider 2005; Euclid Collaboration: Ajani et al. 2023). To break these degeneracies and access the richer information encoded in the LSS, higher-order statistics and non-Gaussian probes have become an increasingly active area of development. These methods hold the potential not only to tighten parameter constraints, but also to offer a more complete statistical description of the cosmic matter distribution.

In the following, we focus on third-order statistics, which provide a measure of the skewness of the matter distribution. Since second- and third-order statistics have different dependences on Ω_m and S_8 (Takada & Jain 2004; Kilbinger & Schneider 2005; Kayo et al. 2013; Pyne & Joachimi 2021), combining them in a joint parameter analysis helps to break degeneracies, as demonstrated by Heydenreich et al. (2023). Various higher-order statistics have been introduced to address the limitations of second-order statistics in weak gravitational lensing. Examples include the shear three-point correlation function (Schneider & Lombardi 2003; Schneider et al. 2005), peak-count statistics (Martinet et al. 2018; Harnois-Déraps et al. 2021), persistent homology (Heydenreich et al. 2021), and density split statistics (Gruen et al. 2018; Burger et al. 2022). For our analysis, we adopted second- and third-order aperture mass statistics, $\langle \mathcal{M}_{\text{ap}}^2 \rangle$ and $\langle \mathcal{M}_{\text{ap}}^3 \rangle$, due to several advantages. Firstly, they are scalars, which effectively compress data and simplify handling (Heydenreich et al. 2023). Secondly, they are unaffected by the mass-sheet degeneracy, a transformation that adds a constant $1 - \lambda$ to the dimensionless surface mass density, while linearly scaling it by λ , leaving lensing observables unchanged (Falco et al. 1985; Schneider & Seitz 1995). Thirdly, they naturally decompose E - and B -modes, where B -modes should not be present in gravitational lensing to leading order, providing a valuable check for systematic errors. Lastly, they can be analytically modelled in a comparatively simple manner.

This paper is part of a series that presents a comprehensive approach to cosmological parameter estimation using third-order shear statistics. The series begins with an in-depth analysis of the analytical model for $\langle \mathcal{M}_{\text{ap}}^3 \rangle$ (Heydenreich et al. 2023). The second paper derives the analytical auto-covariance for $\langle \mathcal{M}_{\text{ap}}^3 \rangle$ (Linke et al. 2023, L+23 hereafter). The third paper Porth et al. (2024) introduces an efficient method for computing aperture mass statistics by measuring the shear three-point correlation function (3PCF). Finally, the last paper (Burger et al. 2024) presents the first numerical cosmological parameter analysis of the fourth data release of the Kilo-Degree Survey (KiDS-1000), incorporating second- and third-order shear statistics, intrinsic alignments, and the effects of baryonic feedback. The current paper establishes the foundation for an analytical joint analysis of $\langle \mathcal{M}_{\text{ap}}^2 \rangle$ and $\langle \mathcal{M}_{\text{ap}}^3 \rangle$ by deriving the final missing component: the analytical cross-covariance. Once derived, we validated the cross-covariance through a joint parameter analysis on simulated data. To achieve this, we used the analytical third-order auto-covariance from L+23 and the analytical second-order auto-covariance computed in Linke et al. (2024). By combining these covariances with the cross-covariance, we were able to construct posterior distributions using Markov chain Monte Carlo (MCMC) sampling.

Using an analytical covariance instead of a numerical one provides several advantages. A numerical covariance can be obtained either from simulations or directly from observational data. In the case of simulations, the number of realisations required must significantly exceed the number of components in the data vector, making the approach computationally expensive. Running these simulations and computing the covariance is particularly demanding, as data vectors are often large. In contrast, an analytical model significantly reduces computational cost while maintaining accuracy. Alternatively, the covariance matrix can be estimated using jackknife resampling or bootstrapping. These methods divide the data field into multiple patches that are small enough to ensure there are more patches than data vector components, and compute the sample covariance over these patches. However, this approach has limitations. It neglects correlations on scales larger than the patch size and assumes that the patches are mutually independent, which is not strictly true since adjacent patches share boundaries. As a result, this method introduces a bias in the covariance estimate, whereas an analytical covariance remains unaffected. Moreover, an analytical approach enables a straightforward evaluation of the covariance under different cosmological models. This is crucial for parameter inference and forecasting, as observational data by definition only samples a single cosmology, and generating large suites of simulations for multiple cosmologies would be prohibitively expensive.

The paper is organised as follows. Section 2 provides a concise review of aperture mass statistics. Section 3 introduces real-space estimators for aperture mass statistics and derives the analytical cross-covariance from these estimators. Additionally, we present a numerical method for validating the analytical results. Section 4 compares the analytical and numerical cross-covariances, where the analytical model is constructed by adapting the publicly available third-order modelling code `threepoint`¹. Section 5 outlines the joint cosmological parameter estimation and its validation against numerical contours. Finally, Sect. 6 discusses our findings and presents concluding remarks.

¹ <https://github.com/sheydenreich/threepoint>

2. Theoretical background

This section briefly overviews the key quantities of weak gravitational lensing and aperture masses. A more detailed discussion can be found in [Schneider \(1996\)](#), [Bartelmann & Schneider \(2001\)](#), [Hoekstra & Jain \(2008\)](#) or [Bartelmann \(2010\)](#). This paper assumes a flat universe, such that the comoving angular-diameter distance equals the comoving radial distance $f_K(\chi) = \chi$. Furthermore, we use the flat-sky approximation. The matter density contrast is $\delta(\chi, \boldsymbol{\theta}, \chi) = \rho(\chi, \boldsymbol{\theta}, \chi) / \bar{\rho}(\chi) - 1$, where $\rho(\chi, \boldsymbol{\theta}, \chi)$ is the matter density at angular position $\boldsymbol{\theta}$ and comoving distance χ on the backward light cone, and $\bar{\rho}(\chi)$ is the average density at χ . Defining the Fourier transform of variables with a tilde, the polyspectra of the density contrast $\mathcal{P}_n^{(3d)}$ are defined by

$$\langle \tilde{\delta}(\boldsymbol{\ell}_1/\chi, \chi) \dots \tilde{\delta}(\boldsymbol{\ell}_n/\chi, \chi) \rangle_c = (2\pi)^3 \delta_D(\boldsymbol{\ell}_1/\chi + \dots + \boldsymbol{\ell}_n/\chi) \mathcal{P}_n^{(3d)}(\boldsymbol{\ell}_1/\chi, \dots, \boldsymbol{\ell}_n/\chi; \chi), \quad (1)$$

where $\langle \dots \rangle_c$ are cumulants or connected correlation functions, and δ_D is a Dirac delta function.

2.1. The convergence

The convergence, or the dimensionless surface mass density, can be calculated from δ by a weighted line-of-sight integration

$$\kappa(\boldsymbol{\theta}) = \frac{3H_0^2 \Omega_m}{2c^2} \int_0^\infty d\chi q(\chi) \chi \frac{\delta(\chi, \boldsymbol{\theta}, \chi)}{a(\chi)}, \quad \text{with the weight function } q(\chi) = \int_\chi^\infty d\chi' p(\chi') \frac{\chi' - \chi}{\chi'}, \quad (2)$$

the Hubble constant H_0 , the probability distribution of source galaxies in comoving distance $p(\chi) d\chi$, the speed of light c , and the scale factor $a(\chi)$, normalised to unity today. The n th-order polyspectra \mathcal{P}_n of the convergence are defined by

$$\langle \kappa(\boldsymbol{\theta}_1) \dots \kappa(\boldsymbol{\theta}_n) \rangle_c = \int \frac{d^2 \ell_1}{(2\pi)^2} \dots \frac{d^2 \ell_n}{(2\pi)^2} \mathcal{P}_n(\boldsymbol{\ell}_1, \dots, \boldsymbol{\ell}_n) (2\pi)^2 \delta_D(\boldsymbol{\ell}_1 + \dots + \boldsymbol{\ell}_n) e^{-i(\boldsymbol{\ell}_1 \cdot \boldsymbol{\theta}_1 + \dots + \boldsymbol{\ell}_n \cdot \boldsymbol{\theta}_n)}. \quad (3)$$

Here and in the following, integrals without explicit boundaries are integrals over \mathbb{R}^2 . Particularly, the power spectrum P , the bispectrum B , and the tetraspectrum P_5 are the convergence polyspectra important for our present discussion and are defined as

$$P(\ell) = \mathcal{P}_2(\boldsymbol{\ell}, -\boldsymbol{\ell}), \quad B(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2) = \mathcal{P}_3(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, -\boldsymbol{\ell}_1 - \boldsymbol{\ell}_2), \quad \text{and} \quad P_5(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, \boldsymbol{\ell}_3, \boldsymbol{\ell}_4) = \mathcal{P}_5(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, \boldsymbol{\ell}_3, \boldsymbol{\ell}_4, -\boldsymbol{\ell}_1 - \boldsymbol{\ell}_2 - \boldsymbol{\ell}_3 - \boldsymbol{\ell}_4), \quad (4)$$

where we made use of the homogeneity and isotropy of the Universe. Under the Limber approximation, they are related to the polyspectra of δ ($\mathcal{P}_n^{(3d)}$) by a weighted integration along the line of sight ([Limber 1954](#); [Kaiser & Jaffe 1997](#); [Kayo et al. 2013](#)):

$$\mathcal{P}_n(\boldsymbol{\ell}_1, \dots, \boldsymbol{\ell}_n) = \left(\frac{3H_0^2 \Omega_m}{2c^2} \right)^n \int_0^\infty d\chi \frac{q^n(\chi)}{\chi^{n-2} a^n(\chi)} \mathcal{P}_n^{(3d)}(\boldsymbol{\ell}_1/\chi, \dots, \boldsymbol{\ell}_n/\chi; \chi). \quad (5)$$

The convergence defined in Eq. (2) and its n th-order connected correlation function (Eq. (3)) will be used to calculate the aperture mass in the following subsection, from which our statistics are derived. If shape noise from the intrinsic ellipticity of galaxies, ϵ_i , is present, it must be accounted for by adding $\sigma_{\epsilon_i}^2/(2N)$ to the power spectrum $P(\ell)$, where σ_{ϵ_i} is the two-component intrinsic ellipticity dispersion and N is the galaxy number density. A proof of this correction can be found in Appendix B of [L+23](#). Since shape noise is Gaussian, it does not affect the bispectrum or tetraspectrum.

2.2. Definition of the aperture mass

The aperture mass \mathcal{M}_{ap} can be calculated by either a weighted two-dimensional integration over the convergence or the tangential shear γ_t . It was first introduced in [Schneider \(1996\)](#), and can be defined by either

$$\mathcal{M}_{\text{ap}}(\boldsymbol{\phi}, \theta) := \int d^2 \vartheta U_\theta(|\boldsymbol{\phi} - \boldsymbol{\vartheta}|) \kappa(\boldsymbol{\vartheta}) \quad \text{or} \quad \mathcal{M}_{\text{ap}}(\boldsymbol{\phi}, \theta) := \int d^2 \vartheta Q_\theta(|\boldsymbol{\phi} - \boldsymbol{\vartheta}|) \gamma_t(\boldsymbol{\vartheta}), \quad (6)$$

where $\boldsymbol{\phi}$ is the aperture centre coordinate, θ the filter radius, and U_θ and Q_θ are radially symmetric filter functions. U_θ is compensated, obeying the relation

$$\int_0^\infty d\vartheta \vartheta U_\theta(\vartheta) = 0, \quad \text{and is connected to } Q_\theta \text{ through } Q_\theta(\vartheta) = \frac{2}{\vartheta^2} \int_0^\vartheta d\vartheta' \vartheta' U_\theta(\vartheta') - U_\theta(\vartheta). \quad (7)$$

For our purpose, we use the exponential filter function proposed in [Crittenden et al. \(2002\)](#),

$$U_\theta(\vartheta) = \frac{1}{2\pi\theta^2} \left(1 - \frac{\vartheta^2}{2\theta^2} \right) e^{-\vartheta^2/(2\theta^2)}, \quad \text{such that} \quad Q_\theta(\vartheta) = \frac{\vartheta^2}{4\pi\theta^4} e^{-\vartheta^2/(2\theta^2)}. \quad (8)$$

The Fourier transform of U_θ is

$$\tilde{U}_\theta(\ell) = \frac{\ell^2 \theta^2}{2} e^{-\ell^2 \theta^2/2} =: \tilde{u}(\ell\theta). \quad (9)$$

The second- and third-order aperture mass statistics are essential for our present discussion and are defined, respectively, by

$$\langle \mathcal{M}_{\text{ap}}^2 \rangle(\theta_1, \theta_2) := \langle \mathcal{M}_{\text{ap}}(\boldsymbol{\phi}, \theta_1) \mathcal{M}_{\text{ap}}(\boldsymbol{\phi}, \theta_2) \rangle \quad \text{and} \quad \langle \mathcal{M}_{\text{ap}}^3 \rangle(\theta_1, \theta_2, \theta_3) := \langle \mathcal{M}_{\text{ap}}(\boldsymbol{\phi}, \theta_1) \mathcal{M}_{\text{ap}}(\boldsymbol{\phi}, \theta_2) \mathcal{M}_{\text{ap}}(\boldsymbol{\phi}, \theta_3) \rangle, \quad (10)$$

where $\langle \dots \rangle$ denotes the spatial average over $\boldsymbol{\phi}$. We note that since $\langle \mathcal{M}_{\text{ap}} \rangle = 0$, as the expectation value of the density contrast δ is by definition zero, the second- and third-order cumulants of the aperture mass are equal to the spatial averages: $\langle \mathcal{M}_{\text{ap}}^3 \rangle_{\text{c}} = \langle \mathcal{M}_{\text{ap}}^3 \rangle$ and $\langle \mathcal{M}_{\text{ap}}^2 \rangle_{\text{c}} = \langle \mathcal{M}_{\text{ap}}^2 \rangle$. Inserting the definition of the aperture mass calculated from the convergence (Eq. (6)) into the Eq. (10), using that the spatial averages are equal to the cumulants and substituting the cumulants of the convergence with the polyspectra as defined in Eq. (3), one finds

$$\begin{aligned} \langle \mathcal{M}_{\text{ap}}^2 \rangle(\theta_1, \theta_2) &= \int d^2\vartheta_1 U_{\theta_1}(|\boldsymbol{\phi} - \boldsymbol{\vartheta}_1|) \int d^2\vartheta_2 U_{\theta_2}(|\boldsymbol{\phi} - \boldsymbol{\vartheta}_2|) \int \frac{d^2\ell_1}{(2\pi)^2} \int \frac{d^2\ell_2}{(2\pi)^2} \mathcal{P}_2(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2) (2\pi)^2 \delta_{\text{D}}(\boldsymbol{\ell}_1 + \boldsymbol{\ell}_2) e^{-i(\boldsymbol{\ell}_1 \cdot \boldsymbol{\vartheta}_1 + \boldsymbol{\ell}_2 \cdot \boldsymbol{\vartheta}_2)} \\ &= \int \frac{d^2\ell_1}{(2\pi)^2} \int \frac{d^2\ell_2}{(2\pi)^2} \tilde{u}(\ell_1 \theta_1) \tilde{u}(\ell_2 \theta_2) \mathcal{P}_2(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2) (2\pi)^2 \delta_{\text{D}}(\boldsymbol{\ell}_1 + \boldsymbol{\ell}_2) e^{-i(\boldsymbol{\ell}_1 + \boldsymbol{\ell}_2) \cdot \boldsymbol{\phi}} \\ &= \int \frac{d^2\ell}{(2\pi)^2} \tilde{u}(\ell \theta_1) \tilde{u}(\ell \theta_2) P(\ell), \end{aligned} \quad (11)$$

for the second-order aperture mass statistics, where going from the first to the second line, we performed the Fourier transformation on the U_{θ} filters. In the last step, we integrated the Dirac delta function. It is important to note that taking equal aperture radii for the second-order aperture mass statistics (i.e., $\theta_1 = \theta_2$) captures all the information in this order. With the chosen \tilde{u} filter (Eq. (9)), the dispersion at two distinct aperture radii can be expressed as a dispersion at an average aperture radius (Schneider et al. 2005)

$$\langle \mathcal{M}_{\text{ap}}(\theta_1) \mathcal{M}_{\text{ap}}(\theta_2) \rangle = \int \frac{d^2\ell}{(2\pi)^2} P(\ell) \frac{\ell^4 \theta_1^2 \theta_2^2}{4} e^{-\ell^2(\theta_1^2 + \theta_2^2)/2} = \frac{4\theta_1^2 \theta_2^2}{(\theta_1^2 + \theta_2^2)^2} \left\langle \mathcal{M}_{\text{ap}}^2 \left(\sqrt{\frac{\theta_1^2 + \theta_2^2}{2}} \right) \right\rangle. \quad (12)$$

Therefore, using equal aperture radii in the dispersion preserves all relevant information, and it is sufficient to evaluate $\langle \mathcal{M}_{\text{ap}}^2 \rangle(\theta) := \langle \mathcal{M}_{\text{ap}}^2 \rangle(\theta, \theta)$ with a single aperture radius as input. Equivalently, for the third-order

$$\begin{aligned} \langle \mathcal{M}_{\text{ap}}^3 \rangle(\theta_1, \theta_2, \theta_3) &= \int \frac{d^2\ell_1}{(2\pi)^2} \int \frac{d^2\ell_2}{(2\pi)^2} \int \frac{d^2\ell_3}{(2\pi)^2} \tilde{u}(\ell_1 \theta_1) \tilde{u}(\ell_2 \theta_2) \tilde{u}(\ell_3 \theta_3) \mathcal{P}_3(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, \boldsymbol{\ell}_3) (2\pi)^2 \delta_{\text{D}}(\boldsymbol{\ell}_1 + \boldsymbol{\ell}_2 + \boldsymbol{\ell}_3) e^{-i(\boldsymbol{\ell}_1 + \boldsymbol{\ell}_2 + \boldsymbol{\ell}_3) \cdot \boldsymbol{\phi}} \\ &= \int \frac{d^2\ell_1}{(2\pi)^2} \int \frac{d^2\ell_2}{(2\pi)^2} \tilde{u}(\ell_1 \theta_1) \tilde{u}(\ell_2 \theta_2) \tilde{u}(|\boldsymbol{\ell}_1 + \boldsymbol{\ell}_2| \theta_3) B(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2). \end{aligned} \quad (13)$$

The simplification to equal aperture radii used above is not applicable to the third-order statistics. This can be observed from Eq. (9), which sharply peaks around $\ell\theta = \sqrt{2}$. For $\langle \mathcal{M}_{\text{ap}}^3 \rangle$, if one were to take $\theta_1 = \theta_2 = \theta_3$ in Eq. (13), only equilateral configurations of the bispectrum with $\ell_1 \sim \ell_2 \sim |\boldsymbol{\ell}_1 + \boldsymbol{\ell}_2|$ would be probed (Schneider et al. 2005). However, the non-equilateral aperture radii configurations depend on different bispectrum configurations and provide additional information in this order. Therefore, distinct aperture radii should be allowed for third-order statistics.

3. Derivation of the analytical cross-covariance

In a real survey or simulation, convolving with the weight function U_{θ} over an infinite area is not feasible. Therefore, following L+23, we define a real-space estimator of the aperture mass as

$$\hat{\mathcal{M}}_{\text{ap}}(\boldsymbol{\phi}, \theta) := \int_{A'} d^2\vartheta U_{\theta}(|\boldsymbol{\phi} - \boldsymbol{\vartheta}|) \kappa(\boldsymbol{\vartheta}), \quad (14)$$

where A' represents the finite survey area. By spatially averaging products of this real-space estimator over the aperture centre $\boldsymbol{\phi}$, we estimate the second- and third-order aperture mass statistics. However, since we are working with a finite survey area, this spatial average is itself an approximation. Moreover, the averaging region must be smaller than the full survey area, as the aperture mass involves integrating over a region surrounding the aperture centre. Placing aperture centres too close to the survey boundary would introduce biases due to boundary effects. To mitigate these effects, we estimate the second- and third-order aperture mass statistics by averaging over an area A , which is entirely contained within the survey region A' as illustrated in Fig. 1. The area A is chosen such that 99.9% of the effective support of Q_{θ} is within A' when the aperture centre $\boldsymbol{\phi}$ is placed on the boundary of A and the largest aperture radius θ_{max} is used. This condition is satisfied when the boundary of A is positioned $4\theta_{\text{max}}$ inside the boundary of A' (Heydenreich et al. 2023). The field boundary of A is determined using Q_{θ} rather than U_{θ} because U_{θ} is a compensated filter function, meaning its support is not well-defined in the same way. Given this set-up, the integrals over the finite survey area A' can be approximated as integrals over the entire real space \mathbb{R}^2 , since the convolution with U_{θ} contributes negligibly outside of A' . The real-space statistics of the aperture mass are thus given by

$$\hat{\mathcal{M}}_{\text{ap}}^2(\theta) = \frac{1}{A} \int_A d^2\phi \left[\prod_{i=1}^2 \int d^2\vartheta_i U_{\theta}(|\boldsymbol{\phi} - \boldsymbol{\vartheta}_i|) \kappa(\boldsymbol{\vartheta}_i) \right] \quad \text{and} \quad \hat{\mathcal{M}}_{\text{ap}}^3(\theta_1, \theta_2, \theta_3) = \frac{1}{A} \int_A d^2\phi \left[\prod_{i=1}^3 \int d^2\vartheta_i U_{\theta_i}(|\boldsymbol{\phi} - \boldsymbol{\vartheta}_i|) \kappa(\boldsymbol{\vartheta}_i) \right]. \quad (15)$$

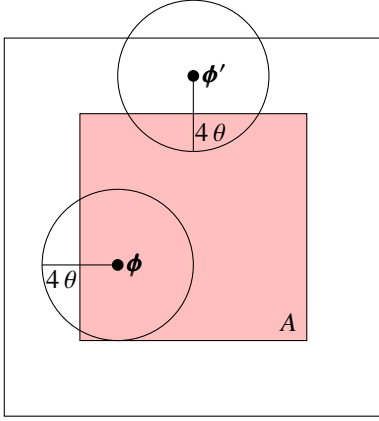


Fig. 1. Aperture centre placement in a data field (from L+23). Given a data field of area A' (not necessarily a square field), apertures are only placed within a smaller field A , which lies well within A' . The boundary of A is chosen such that 99.9% of the effective support of Q_θ lies within A (which corresponds to $\sim 4\theta$) if the aperture centre lies on this boundary. To illustrate this, two aperture centres are depicted: ϕ is within A and ϕ' is outside of A . It can be seen that the circle around ϕ' , illustrating the effective support of Q_θ , extends beyond A' , thus introducing bias into the calculated aperture mass at this point. To exclude this effect, we only average the aperture centre over the red region A .

Table 1. Aperture scale radii permutations for $T_{PB,2}$ and $T_{PB,3}$.

$T_{PB,2}$ permutations	$T_{PB,3}$ permutations
$\theta_1 \theta_4 \theta_5 \theta_2 \theta_3$	$\theta_1 \theta_2 \theta_4 \theta_3 \theta_5$
$\theta_2 \theta_4 \theta_5 \theta_1 \theta_3$	$\theta_1 \theta_3 \theta_4 \theta_2 \theta_5$
$\theta_3 \theta_4 \theta_5 \theta_1 \theta_2$	$\theta_2 \theta_3 \theta_4 \theta_1 \theta_5$

Notes. The first permutation of each term corresponds to the explicit terms in Eq. (18). Interchanging the aperture radii left (right) of the vertical bars is an invariant property.

The cross-covariance of the real-space estimators is then defined as

$$C_{\hat{\mathcal{M}}_{\text{ap}}^3, \hat{\mathcal{M}}_{\text{ap}}^2}(\theta_1, \theta_2, \theta_3; \theta_4) = \langle \hat{\mathcal{M}}_{\text{ap}}^3 \hat{\mathcal{M}}_{\text{ap}}^2 \rangle(\theta_1, \theta_2, \theta_3; \theta_4) - \langle \hat{\mathcal{M}}_{\text{ap}}^3 \rangle(\theta_1, \theta_2, \theta_3) \langle \hat{\mathcal{M}}_{\text{ap}}^2 \rangle(\theta_4). \quad (16)$$

Here and in the following, when a comma ‘,’ is used, the function is invariant under permutations of its arguments, whereas when a semi-colon ‘;’ is used, they are not invariant. The first term in Eq. (16) is given as

$$\begin{aligned} \langle \hat{\mathcal{M}}_{\text{ap}}^3 \hat{\mathcal{M}}_{\text{ap}}^2 \rangle(\theta_1, \theta_2, \theta_3; \theta_4) &= \frac{1}{A^2} \int_A d^2 \phi_1 \int_A d^2 \phi_2 \left[\prod_{i=1}^3 \int d^2 \vartheta_i U_{\theta_i}(|\phi_1 - \vartheta_i|) \right] \left[\prod_{j=4}^5 \int d^2 \vartheta_j U_{\theta_j}(|\phi_2 - \vartheta_j|) \right] \\ &\times \langle \kappa(\vartheta_1) \kappa(\vartheta_2) \kappa(\vartheta_3) \kappa(\vartheta_4) \kappa(\vartheta_5) \rangle. \end{aligned} \quad (17)$$

Here, the fifth-order correlation function can be separated into cumulants,

$$\begin{aligned} \langle \hat{\mathcal{M}}_{\text{ap}}^3 \hat{\mathcal{M}}_{\text{ap}}^2 \rangle(\theta_1, \theta_2, \theta_3; \theta_4) &= \frac{1}{A^2} \int_A d^2 \phi_1 \int_A d^2 \phi_2 \left[\prod_{i=1}^3 \int d^2 \vartheta_i U_{\theta_i}(|\phi_1 - \vartheta_i|) \right] \left[\prod_{j=4}^5 \int d^2 \vartheta_j U_{\theta_j}(|\phi_2 - \vartheta_j|) \right] \\ &\times \left\{ \langle \kappa(\vartheta_1) \kappa(\vartheta_2) \kappa(\vartheta_3) \rangle_c \langle \kappa(\vartheta_4) \kappa(\vartheta_5) \rangle_c \right. \\ &+ \langle \kappa(\vartheta_1) \kappa(\vartheta_4) \kappa(\vartheta_5) \rangle_c \langle \kappa(\vartheta_2) \kappa(\vartheta_3) \rangle_c + 2 \text{ Perm. of } (\theta_1, \theta_2, \theta_3) \\ &+ 2 \left[\langle \kappa(\vartheta_1) \kappa(\vartheta_2) \kappa(\vartheta_4) \rangle_c \langle \kappa(\vartheta_3) \kappa(\vartheta_5) \rangle_c + 2 \text{ Perm. of } (\theta_1, \theta_2, \theta_3) \right] \\ &\left. + \langle \kappa(\vartheta_1) \kappa(\vartheta_2) \kappa(\vartheta_3) \kappa(\vartheta_4) \kappa(\vartheta_5) \rangle_c \right\} \\ &=: T_{PB,1}(\theta_1, \theta_2, \theta_3; \theta_4) + T_{PB,2}(\theta_1, \theta_2, \theta_3; \theta_4) + T_{PB,3}(\theta_1, \theta_2, \theta_3; \theta_4) + T_{P5}(\theta_1, \theta_2, \theta_3; \theta_4), \end{aligned} \quad (18)$$

where we defined the T -expressions such that $T_{PB,1}$ corresponds to the term in the second line, $T_{PB,2}$ to the third, $T_{PB,3}$ to the fourth, and T_{P5} to the fifth. The permutations of $T_{PB,2}$ and $T_{PB,3}$ can be found in Table 1. The third term ($T_{PB,3}$) includes a factor of two because swapping $\kappa(\vartheta_4)$ with $\kappa(\vartheta_5)$ results in equivalent terms, as both are connected to θ_4 . Remember $\langle \kappa \rangle = 0$ because the expectation value of the density contrast δ is zero; thus, the second- and third-order cumulants of the convergence are equal to the respective moments: $\langle \kappa^3 \rangle_c = \langle \kappa^3 \rangle$ and $\langle \kappa^2 \rangle_c = \langle \kappa^2 \rangle$ (but $\langle \kappa^5 \rangle_c \neq \langle \kappa^5 \rangle$). As a result, $T_{PB,1}$ cancels the second term on the r.h.s. of Eq. (16), observing it can be written as the expectation value of $\hat{\mathcal{M}}_{\text{ap}}^2$ times the expectation value of $\hat{\mathcal{M}}_{\text{ap}}^3$, such that

$$C_{\hat{\mathcal{M}}_{\text{ap}}^3, \hat{\mathcal{M}}_{\text{ap}}^2}(\theta_1, \theta_2, \theta_3; \theta_4) = T_{PB,2}(\theta_1, \theta_2, \theta_3; \theta_4) + T_{PB,3}(\theta_1, \theta_2, \theta_3; \theta_4) + T_{P5}(\theta_1, \theta_2, \theta_3; \theta_4). \quad (19)$$

The second term $T_{\text{PB},2}$ can be rewritten by substituting the definition for the convergence polyspectra (Eq. (3)) into Eq. (18), cancelling $\boldsymbol{\ell}_3$ and $\boldsymbol{\ell}_5$ by integrating over the Dirac delta functions and performing the Fourier transformations of the U_θ functions

$$\begin{aligned}
 T_{\text{PB},2}(\theta_1, \theta_2, \theta_3; \theta_4) &= \frac{1}{A^2} \int_A d^2\phi_1 \int_A d^2\phi_2 \left[\prod_{i=1}^3 \int d^2\vartheta_i U_{\theta_i}(|\boldsymbol{\phi}_1 - \boldsymbol{\vartheta}_i|) \right] \left[\prod_{j=4}^5 \int d^2\vartheta_j U_{\theta_j}(|\boldsymbol{\phi}_2 - \boldsymbol{\vartheta}_j|) \right] \prod_{k=1}^5 \int \frac{d^2\ell_k}{(2\pi)^6} \\
 &\quad \times \delta_{\text{D}}(\boldsymbol{\ell}_2 + \boldsymbol{\ell}_3) \delta_{\text{D}}(\boldsymbol{\ell}_1 + \boldsymbol{\ell}_4 + \boldsymbol{\ell}_5) \mathcal{P}_3(\boldsymbol{\ell}_1, \boldsymbol{\ell}_4, \boldsymbol{\ell}_5) \mathcal{P}_2(\boldsymbol{\ell}_2, \boldsymbol{\ell}_3) e^{-i(\boldsymbol{\ell}_1 \cdot \boldsymbol{\vartheta}_1 + \boldsymbol{\ell}_2 \cdot \boldsymbol{\vartheta}_2 + \boldsymbol{\ell}_3 \cdot \boldsymbol{\vartheta}_3 + \boldsymbol{\ell}_4 \cdot \boldsymbol{\vartheta}_4 + \boldsymbol{\ell}_5 \cdot \boldsymbol{\vartheta}_5)} \\
 &\quad + 2 \text{ Perm. of } (\theta_1, \theta_2, \theta_3) \\
 &= \frac{1}{A^2} \int_A d^2\phi_1 \int_A d^2\phi_2 \left[\prod_{i=1}^3 \int d^2\vartheta_i U_{\theta_i}(|\boldsymbol{\phi}_1 - \boldsymbol{\vartheta}_i|) \right] \left[\prod_{j=4}^5 \int d^2\vartheta_j U_{\theta_j}(|\boldsymbol{\phi}_2 - \boldsymbol{\vartheta}_j|) \right] \int \frac{d^2\ell_1}{(2\pi)^2} \int \frac{d^2\ell_2}{(2\pi)^2} \\
 &\quad \times \int \frac{d^2\ell_4}{(2\pi)^2} B(\boldsymbol{\ell}_1, \boldsymbol{\ell}_4) P(\ell_2) e^{-i(\boldsymbol{\ell}_1 \cdot (\boldsymbol{\vartheta}_1 - \boldsymbol{\vartheta}_5) + \boldsymbol{\ell}_2 \cdot (\boldsymbol{\vartheta}_2 - \boldsymbol{\vartheta}_3) + \boldsymbol{\ell}_4 \cdot (\boldsymbol{\vartheta}_4 - \boldsymbol{\vartheta}_5))} + 2 \text{ Perm. of } (\theta_1, \theta_2, \theta_3) \\
 &= \frac{1}{A^2} \int_A d^2\phi_1 \int_A d^2\phi_2 \int \frac{d^2\ell_1}{(2\pi)^2} \int \frac{d^2\ell_2}{(2\pi)^2} \int \frac{d^2\ell_4}{(2\pi)^2} B(\boldsymbol{\ell}_1, \boldsymbol{\ell}_4) P(\ell_2) \\
 &\quad \times \tilde{u}(\ell_1\theta_1) \tilde{u}(\ell_2\theta_2) \tilde{u}(\ell_2\theta_3) \tilde{u}(\ell_4\theta_4) \tilde{u}(|\boldsymbol{\ell}_1 + \boldsymbol{\ell}_4|\theta_4) e^{-i\boldsymbol{\ell}_1 \cdot (\boldsymbol{\phi}_1 - \boldsymbol{\phi}_2)} + 2 \text{ Perm. of } (\theta_1, \theta_2, \theta_3) .
 \end{aligned} \tag{20}$$

Collecting all terms containing information on the survey area and geometry in the geometry factor $G_A(\boldsymbol{\ell})$ defined as (L+23)

$$G_A(\boldsymbol{\ell}) = \frac{1}{A^2} \int_A d^2\phi_1 \int_A d^2\phi_2 e^{-i\boldsymbol{\ell} \cdot (\boldsymbol{\phi}_1 - \boldsymbol{\phi}_2)} , \tag{21}$$

we find the final result

$$\begin{aligned}
 T_{\text{PB},2}(\theta_1, \theta_2, \theta_3; \theta_4) &= \int \frac{d^2\ell_1}{(2\pi)^2} \int \frac{d^2\ell_2}{(2\pi)^2} \int \frac{d^2\ell_4}{(2\pi)^2} B(\boldsymbol{\ell}_1, \boldsymbol{\ell}_4) P(\ell_2) \tilde{u}(\ell_1\theta_1) \tilde{u}(\ell_2\theta_2) \tilde{u}(\ell_2\theta_3) \tilde{u}(\ell_4\theta_4) \tilde{u}(|\boldsymbol{\ell}_1 + \boldsymbol{\ell}_4|\theta_4) \\
 &\quad \times G_A(\boldsymbol{\ell}_1) + 2 \text{ Perm. of } (\theta_1, \theta_2, \theta_3) .
 \end{aligned} \tag{22}$$

By equivalent calculations, one finds

$$\begin{aligned}
 T_{\text{PB},3}(\theta_1, \theta_2, \theta_3; \theta_4) &= 2 \int \frac{d^2\ell_1}{(2\pi)^2} \int \frac{d^2\ell_2}{(2\pi)^2} \int \frac{d^2\ell_3}{(2\pi)^2} B(\boldsymbol{\ell}_2, \boldsymbol{\ell}_3) P(\ell_1) \tilde{u}(\ell_1\theta_1) \tilde{u}(\ell_2\theta_2) \tilde{u}(\ell_3\theta_3) \tilde{u}(|\boldsymbol{\ell}_2 + \boldsymbol{\ell}_3|\theta_4) \tilde{u}(\ell_1\theta_4) \\
 &\quad \times G_A(\boldsymbol{\ell}_1 + \boldsymbol{\ell}_2 + \boldsymbol{\ell}_3) + 2 \text{ Perm. of } (\theta_1, \theta_2, \theta_3) ,
 \end{aligned} \tag{23}$$

and

$$\begin{aligned}
 T_{\text{P5}}(\theta_1, \theta_2, \theta_3; \theta_4) &= \int \frac{d^2\ell_1}{(2\pi)^2} \int \frac{d^2\ell_2}{(2\pi)^2} \int \frac{d^2\ell_3}{(2\pi)^2} \int \frac{d^2\ell_4}{(2\pi)^2} P_5(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, \boldsymbol{\ell}_3, \boldsymbol{\ell}_4) \tilde{u}(\ell_1\theta_1) \tilde{u}(\ell_2\theta_2) \tilde{u}(\ell_3\theta_3) \tilde{u}(\ell_4\theta_4) \\
 &\quad \times \tilde{u}(|\boldsymbol{\ell}_1 + \boldsymbol{\ell}_2 + \boldsymbol{\ell}_3 + \boldsymbol{\ell}_4|\theta_4) G_A(\boldsymbol{\ell}_1 + \boldsymbol{\ell}_2 + \boldsymbol{\ell}_3) .
 \end{aligned} \tag{24}$$

Equations ((22)–(24)) can subsequently be substituted into Eq. (19) to obtain the analytical covariance of the real-space estimator. Next, we investigate how these terms scale with the survey field size.

3.1. Large survey area approximation

This subsection illustrates how the cross-covariance terms behave as the field size increases, by taking the limit $A \rightarrow \infty$ in the geometry factor (Eq. (21)). In this limit, one finds (L+23)

$$\lim_{A \rightarrow \infty} G_A(\boldsymbol{\ell}) = \frac{(2\pi)^2}{A} \delta_{\text{D}}(\boldsymbol{\ell}) . \tag{25}$$

Substituting and eliminating $\boldsymbol{\ell}_1$ by integrating over the Dirac delta function, the large-field approximation of $T_{\text{PB},2}$ becomes

$$\begin{aligned}
 T_{\text{PB},2}^\infty(\theta_1, \theta_2, \theta_3; \theta_4) &= \frac{1}{A} \int \frac{d^2\ell_2}{(2\pi)^2} \int \frac{d^2\ell_4}{(2\pi)^2} B(\mathbf{0}, \boldsymbol{\ell}_4) P(\ell_2) \tilde{u}(0) \tilde{u}(\ell_2\theta_2) \tilde{u}(\ell_2\theta_3) \tilde{u}(\ell_4\theta_4) \tilde{u}(\ell_4\theta_4) \\
 &\quad + 2 \text{ Permutations of } (\theta_1, \theta_2, \theta_3) = 0 ,
 \end{aligned} \tag{26}$$

because $\tilde{u}(0) = 0$, as can be seen from Eq. (9). Thus, $T_{\text{PB},2}$ arises solely from the limited size of the survey field and will therefore be referred to as a finite-field term.

For $T_{\text{PB},3}$ and T_{P5} , as given in Eqs. (23) and (24) respectively, the foregoing limiting procedure leads to

$$\begin{aligned}
 T_{\text{PB},3}^\infty(\theta_1, \theta_2, \theta_3; \theta_4) &= \frac{2}{A} \int \frac{d^2\ell_1}{(2\pi)^2} \int \frac{d^2\ell_2}{(2\pi)^2} B(\boldsymbol{\ell}_2, -\boldsymbol{\ell}_1 - \boldsymbol{\ell}_2) P(\ell_1) \tilde{u}(\ell_1\theta_1) \tilde{u}(\ell_2\theta_2) \tilde{u}(|\boldsymbol{\ell}_1 + \boldsymbol{\ell}_2|\theta_3) \tilde{u}(\ell_1\theta_4) \tilde{u}(\ell_1\theta_4) \\
 &\quad + 2 \text{ Permutations of } (\theta_1, \theta_2, \theta_3) ,
 \end{aligned} \tag{27}$$

and

$$T_{P_5}^\infty(\theta_1, \theta_2, \theta_3; \theta_4) = \frac{1}{A} \int \frac{d^2 \ell_1}{(2\pi)^2} \int \frac{d^2 \ell_2}{(2\pi)^2} \int \frac{d^2 \ell_4}{(2\pi)^2} P_5(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, -\boldsymbol{\ell}_1 - \boldsymbol{\ell}_2, \boldsymbol{\ell}_4) \tilde{u}(\ell_1 \theta_1) \tilde{u}(\ell_2 \theta_2) \tilde{u}(|\boldsymbol{\ell}_1 + \boldsymbol{\ell}_2| \theta_3) \tilde{u}(\ell_4 \theta_4) \tilde{u}(\ell_4 \theta_4), \quad (28)$$

such that the cross-covariance in the large-field approximation simplifies to

$$C_{\mathcal{M}_{\text{ap}}^{3,2}}^\infty(\theta_1, \theta_2, \theta_3; \theta_4) = T_{\text{PB},3}^\infty(\theta_1, \theta_2, \theta_3; \theta_4) + T_{P_5}^\infty(\theta_1, \theta_2, \theta_3; \theta_4). \quad (29)$$

While $T_{\text{PB},3}$ and T_{P_5} scale approximately as $1/A$, $T_{\text{PB},2}$ exhibits a steeper decline with increasing field size.

Since the geometry term (Eq. (21)) is a rapidly oscillating function, its integration presents numerical challenges. Because our computational resources are largely sufficient for this task (albeit with some limitations; see Sect. 6.1), we retain the full expressions. However, in scenarios with limited computing power, or when computational efficiency is prioritised, large-field approximations may serve as useful alternatives.

A concluding remark on the derived cross-covariance and its large-field approximation is that to compute their values, we estimate P_5 in T_{P_5} using a halo decomposition (Appendix A). T_{P_5} is then approximated by keeping the two dominant terms, the one-halo term and one of the two-halo terms:

$$T_{P_5}(\theta_1, \theta_2, \theta_3; \theta_4) \approx T_{P_5,1\text{h}}(\theta_1, \theta_2, \theta_3; \theta_4) + T_{P_5,2\text{h}}(\theta_1, \theta_2, \theta_3; \theta_4). \quad (30)$$

In the large-field approximation $T_{P_5,2\text{h}}^\infty = 0$, which means this is a finite field term, and thus we find $T_{P_5}^\infty \approx T_{P_5,1\text{h}}^\infty$.

3.2. Aperture mass correlation functions

This section demonstrates that the covariance terms (Eqs. (22)–(24)) can alternatively be constructed with aperture mass correlation functions. These correlation functions can be numerically estimated from N -body simulations and therefore enable us to verify the individual components of our analytical cross-covariance.

The aperture mass correlation functions are defined as

$$\begin{aligned} \mathcal{M}_{\text{ap}}^{n,m}(\theta_1, \dots, \theta_m; \theta_{m+1}, \dots, \theta_n; \boldsymbol{\eta}) &:= \left\langle \mathcal{M}_{\text{ap}}(\boldsymbol{\phi}, \theta_1) \dots \mathcal{M}_{\text{ap}}(\boldsymbol{\phi}, \theta_m) \mathcal{M}_{\text{ap}}(\boldsymbol{\phi} + \boldsymbol{\eta}, \theta_{m+1}) \dots \mathcal{M}_{\text{ap}}(\boldsymbol{\phi} + \boldsymbol{\eta}, \theta_n) \right\rangle \\ &= \left[\prod_{i=1}^m \int d^2 \vartheta_i U_{\theta_i}(\vartheta_i) \dots U_{\theta_m}(\vartheta_m) \right] \left[\prod_{j=m+1}^n \int d^2 \vartheta_j U_{\theta_{m+1}}(|\boldsymbol{\vartheta}_{m+1} + \boldsymbol{\eta}|) \dots U_{\theta_n}(|\boldsymbol{\vartheta}_n + \boldsymbol{\eta}|) \right] \left\langle \kappa(\boldsymbol{\vartheta}_1) \dots \kappa(\boldsymbol{\vartheta}_n) \right\rangle, \end{aligned} \quad (31)$$

where, without loss of generality, we set $\boldsymbol{\phi} = \mathbf{0}$ after the equality sign by shifting the origin. Substituting the correlation functions into Eq. (18), $T_{\text{PB},2}$ is rewritten as

$$T_{\text{PB},2}(\theta_1, \theta_2, \theta_3; \theta_4) = \frac{1}{A^2} \mathcal{M}_{\text{ap}}^{2,0}(\theta_2, \theta_3; \mathbf{0}) \int_A d^2 \phi_1 \int_A d^2 \phi_2 \mathcal{M}_{\text{ap}}^{3,1}(\theta_1; \theta_4, \theta_4; \boldsymbol{\phi}_1 - \boldsymbol{\phi}_2) + 2 \text{ Permutations of } (\theta_1, \theta_2, \theta_3). \quad (32)$$

By replacing the $\boldsymbol{\phi}_2$ integral by an $\boldsymbol{\eta} = \boldsymbol{\phi}_1 - \boldsymbol{\phi}_2$ integral and integrating over $d[2]\phi_1$, we obtain $A E_A(\boldsymbol{\eta})$, where $E_A(\boldsymbol{\eta})$ represents the probability that for a point $\boldsymbol{\phi}_1$ within A , the point $\boldsymbol{\phi}_1 + \boldsymbol{\eta}$ also lies within A . For a square survey area, $E_A(\boldsymbol{\eta})$ is given in Appendix A.2 of Heydenreich et al. (2020). By expanding these algebraic steps, we find

$$T_{\text{PB},2}(\theta_1, \theta_2, \theta_3; \theta_4) = \frac{1}{A} \mathcal{M}_{\text{ap}}^{2,0}(\theta_2, \theta_3; \mathbf{0}) \int d^2 \eta E_A(\boldsymbol{\eta}) \mathcal{M}_{\text{ap}}^{3,1}(\theta_1; \theta_4, \theta_4; \boldsymbol{\eta}) + 2 \text{ Permutations of } (\theta_1, \theta_2, \theta_3). \quad (33)$$

After generating the aperture mass correlation function maps from simulations, we azimuthally average them over the polar angle of $\boldsymbol{\eta}$ due to the statistical isotropy of the Universe. Defining the aperture mass correlations as a function of distance η by $\bar{\mathcal{M}}_{\text{ap}}^{n,m}$ (Eq. (31)) cannot be used in this context, as it requires a vector-valued input $\boldsymbol{\eta}$ and integrating $E_A(\boldsymbol{\eta})$ over the polar angle of $\boldsymbol{\eta}$, one finds

$$T_{\text{PB},2}(\theta_1, \theta_2, \theta_3; \theta_4) = \frac{2\pi}{A} \bar{\mathcal{M}}_{\text{ap}}^{2,0}(\theta_2, \theta_3; 0) \int_0^\infty d\eta E_A(\eta) \bar{\mathcal{M}}_{\text{ap}}^{3,1}(\theta_1; \theta_4, \theta_4; \eta) + 2 \text{ Permutations of } (\theta_1, \theta_2, \theta_3), \quad (34)$$

where for a survey of square area, $E_A(\eta)$ is given in Appendix A.2 of Heydenreich et al. (2020)² as

$$E_A(\eta) = \begin{cases} \frac{1}{A\pi} [A\pi - (4\sqrt{A} - \eta)\eta] & \text{for } \eta \leq \sqrt{A}, \\ \frac{2}{\pi} \left[2\sqrt{\frac{\eta^2}{A} - 1} - 1 - \frac{\eta^2}{2A} - \arccos\left(\frac{\sqrt{A}}{\eta}\right) + \arcsin\left(\frac{\sqrt{A}}{\eta}\right) \right] & \text{for } \sqrt{A} \leq \eta \leq \sqrt{2A}, \\ 0 & \text{for } \sqrt{2A} \leq \eta. \end{cases} \quad (35)$$

² For the case $\sqrt{A} \leq \eta \leq \sqrt{2A}$ Eq. (A.11) in Heydenreich et al. (2020) has a factor 2 typo.

Following the same procedure for $T_{\text{PB},1}$ and $T_{\text{PB},3}$ leads to

$$\begin{aligned} T_{\text{PB},1}(\theta_1, \theta_2, \theta_3; \theta_4) &= \mathcal{M}_{\text{ap}}^{3,0}(\theta_1, \theta_2, \theta_3; \mathbf{0}) \mathcal{M}_{\text{ap}}^{2,0}(\theta_4, \theta_4; \mathbf{0}) \\ &= \bar{\mathcal{M}}_{\text{ap}}^{3,0}(\theta_1, \theta_2, \theta_3; 0) \bar{\mathcal{M}}_{\text{ap}}^{2,0}(\theta_4, \theta_4; 0), \end{aligned} \quad (36)$$

$$\begin{aligned} T_{\text{PB},3}(\theta_1, \theta_2, \theta_3; \theta_4) &= \frac{2}{A} \int d^2\eta E_A(\boldsymbol{\eta}) \mathcal{M}_{\text{ap}}^{3,2}(\theta_2, \theta_3; \theta_4; \boldsymbol{\eta}) \mathcal{M}_{\text{ap}}^{2,1}(\theta_1; \theta_4; \boldsymbol{\eta}) + 2 \text{ Permutations of } (\theta_1, \theta_2, \theta_3) \\ &= \frac{4\pi}{A} \int_0^\infty d\eta E_A(\eta) \bar{\mathcal{M}}_{\text{ap}}^{3,2}(\theta_2, \theta_3; \theta_4; \eta) \bar{\mathcal{M}}_{\text{ap}}^{2,1}(\theta_1; \theta_4; \eta) + 2 \text{ Permutations of } (\theta_1, \theta_2, \theta_3). \end{aligned} \quad (37)$$

We note that because Eq. (18) is written in cumulants, T_{P5} cannot be simply connected to the fifth-order correlation function. Therefore, to calculate T_{P5} in terms of correlation functions, we subtract $T_{\text{PB},1}$, $T_{\text{PB},2}$, and $T_{\text{PB},3}$ as defined above, such that

$$\begin{aligned} T_{\text{P5}}(\theta_1, \theta_2, \theta_3; \theta_4) &= \frac{1}{A} \int d^2\eta E_A(\boldsymbol{\eta}) \mathcal{M}_{\text{ap}}^{5,3}(\theta_1, \theta_2, \theta_3; \theta_4, \theta_4; \boldsymbol{\eta}) \\ &\quad - T_{\text{PB},1}(\theta_1, \theta_2, \theta_3; \theta_4) - T_{\text{PB},2}(\theta_1, \theta_2, \theta_3; \theta_4) - T_{\text{PB},3}(\theta_1, \theta_2, \theta_3; \theta_4) \\ &= \frac{2\pi}{A} \int_0^\infty d\eta E_A(\eta) \bar{\mathcal{M}}_{\text{ap}}^{5,3}(\theta_1, \theta_2, \theta_3; \theta_4, \theta_4; \eta) \\ &\quad - T_{\text{PB},1}(\theta_1, \theta_2, \theta_3; \theta_4) - T_{\text{PB},2}(\theta_1, \theta_2, \theta_3; \theta_4) - T_{\text{PB},3}(\theta_1, \theta_2, \theta_3; \theta_4). \end{aligned} \quad (38)$$

4. Analytical model validation

With a model for the analytical cross-covariance of the second- and third-order aperture mass statistics in hand, it is important to test its consistency with N -body simulations.

4.1. Covariance estimation methods from simulations

A numerical cross-covariance can be obtained from simulations in two ways. The first and most straightforward method involves taking the unbiased sample covariance from multiple simulation realisations. For this approach, a significantly larger number of realisations (lines of sight) is required than the number of components in the data vector, and these realisations must be independent of one another. As discussed in Sect. 3, we need to exclude the boundaries of the fields by cutting off four times the largest aperture radius, θ_{max} . After spatially averaging the second- and third-order aperture mass over its centre coordinate $\boldsymbol{\phi}$ ($\hat{\mathcal{M}}_{\text{ap}}^2$ and $\hat{\mathcal{M}}_{\text{ap}}^3$, respectively) for each realisation i , the sample covariance is computed using

$$C_{\hat{\mathcal{M}}_{\text{ap}}^{3,2}}^{\text{sim}}(\theta_1, \theta_2, \theta_3; \theta_4) = \frac{1}{N-1} \sum_{i=1}^N [\hat{\mathcal{M}}_{\text{ap}}^3(\theta_1, \theta_2, \theta_3)]_i [\hat{\mathcal{M}}_{\text{ap}}^2(\theta_4)]_i - \frac{1}{N(N-1)} \sum_{i=1}^N [\hat{\mathcal{M}}_{\text{ap}}^3(\theta_1, \theta_2, \theta_3)]_i \sum_{j=1}^N [\hat{\mathcal{M}}_{\text{ap}}^2(\theta_4)]_j, \quad (39)$$

where the sums are taken over the realisations, and N represents the number of realisations.

The second method involves using aperture mass correlation function maps, as outlined in Sect. 3.2. It is important to note that the same boundary exclusion must be applied for this method. The advantage of this approach is that it allows us to obtain the T -terms separately.

The sum of $T_{\text{PB},2}$, $T_{\text{PB},3}$, and T_{P5} , computed from the aperture mass correlation function maps (Eqs. (34), (37), and (38)), reproduces the full sample covariance. We use the individually measured T -terms solely for comparison with their analytical counterparts. In analyses that rely on the full numerical covariance, such as the cosmological parameter estimations in Sect. 5, we adopt the sample covariance as our reference, as this allows us to estimate errors via bootstrapping.

4.2. Validation simulation – SLICS

We use the Scinet-LightCone Simulations (SLICS hereafter; Harnois-Déraps & van Waerbeke 2015; Harnois-Déraps et al. 2018) to validate the analytical model. These are dark matter only N -body simulations from which we use both shear catalogues and convergence maps. The SLICS are a suite of ray-tracing simulations which use 1536^3 particles placed in a $505 h^{-1}$ Mpc box. A flat Λ CDM cosmology is assumed, with normalised Hubble constant $h = 0.69$, the normalisation of the power spectrum $\sigma_8 = 0.83$, the equation of state of dark matter $w = -1$, matter density parameter $\Omega_{\text{m}} = 0.29$, baryon density parameter $\Omega_{\text{b}} = 0.047$, and primordial power spectrum spectral index $n_{\text{s}} = 0.969$. For the set of convergence maps, we use 715 independent lines of sight, all having an area of $10 \times 10 \text{ deg}^2$, placed on a 7745×7745 pixel grid. All maps have a redshift of $z = 0.897$, assume the flat-sky approximation, and do not contain shape noise. The aperture mass maps are calculated using the first equation in Eq. (6). The shear catalogue set contains 927 independent lines of sight, and has a galaxy number density

$$n(z) \propto z^2 e^{-(z/z_0)^\beta}, \quad (40)$$

where $z_0 = 0.637$, $\beta = 1.5$, and the galaxy density is normalised to 30 arcmin^{-2} , mimicking a density that is expected in the Euclid survey (Laureijs et al. 2011; Euclid Collaboration: Mellier et al. 2025). The galaxies have a maximum redshift of $z_{\text{max}} = 3$. Each

realisation has again an area of $10 \times 10 \text{ deg}^2$. Every galaxy contains a shear, upon which an intrinsic ellipticity is added as shape noise. This is done by adding a random ellipticity to the shear from a Gaussian with a two-component ellipticity dispersion of $\sigma_\epsilon^2 = (0.37)^2$. The aperture mass is estimated by a weighted summation over the ellipticities of the galaxies

$$\hat{\mathcal{M}}_{\text{ap}}(\boldsymbol{\phi}, \theta) = \frac{1}{n_{\text{gal}}(\boldsymbol{\phi})} \sum_i^N Q_\theta(|\boldsymbol{\phi} - \boldsymbol{\vartheta}_i|) \epsilon_{t,i}, \quad \text{with} \quad n_{\text{gal}}(\boldsymbol{\phi}) = \sum_i^N Q_\theta(|\boldsymbol{\phi} - \boldsymbol{\vartheta}_i|), \quad (41)$$

where N is the total number of galaxies in the catalogue and $\epsilon_{t,i}$ is the sum of the tangential shear plus the tangential intrinsic ellipticity for the galaxy i . This is a discrete approximation of the second definition of the aperture mass in Eq. (6). We test all filter combinations of $\theta \in \{4', 8', 16'\}$. The aperture centres of the convergence maps and shear catalogues are placed in a region of $7.87 \times 7.87 \text{ deg}^2$ after boundary removal.

4.3. Results

To compute the analytical cross-covariance using Eqs. (22)–(24), we require the convergence polyspectra. These are obtained by substituting the underlying matter polyspectra: the matter power spectrum is taken from the revised `Halofit` (Takahashi et al. 2012), and the bispectrum from `BiHalofit` (Takahashi et al. 2020). The tetraspectrum is approximated using the halo model, following the formalism of Cooray & Sheth (2002), and detailed in Appendix A. Each polyspectrum is then projected into convergence space via Limber integration (Eq. (5)) to obtain the polyspectra for the convergence.

4.3.1. Validity of the large-field approximation

First, we compare the analytical terms $T_{\text{PB},3}$ and $T_{\text{P5},1\text{h}}$ with their large-field approximation as shown in the top panels of Fig. 2. $T_{\text{PB},3}$ agrees with its large-field approximation within 5% for $\sim 70\%$ of the filter combinations, both for the convergence maps without shape noise and the shear catalogues with shape noise. The larger deviations ($\sim 20\%$) are present mostly for small aperture radii combinations (combinations with $4'$). It seems larger aperture radii converge to the large-field approximation more rapidly than smaller ones. The large-field approximation generally overpredicts $T_{\text{PB},3}$. For $T_{\text{P5},1\text{h}}$, we see similar results; the one-halo term and the large-field approximation agree within less than 5% for $\sim 80\%$ of the filter combinations in both datasets. For the shear catalogues, the filter combination $(4', 8', 16'; 4')$ differs by about 17%. This larger discrepancy is probably due to the inaccuracy of the integration in $T_{\text{P5},1\text{h}}$. To calculate it, we have to perform a nine-dimensional integration (one dimension coming from the mass integral in Eq. (A.4) of the halo model and the other eight as seen in Eq. 24), two of which have to integrate over the oscillating function G_A . This is computationally intensive, so we were restricted in calculating it more precisely. Also, for $T_{\text{P5},1\text{h}}$, we see that small filter radii have the largest deviation from the large-field approximation. It should be noted that above, we compared the one-halo term with the large-field approximation, as the two-halo term $T_{\text{P5},2\text{h}}$ is zero in this approximation (see Appendix A). In the bottom left panel of Fig. 2, we show the relative contribution of $T_{\text{P5},2\text{h}}$ to the total $T_{\text{P5}} = T_{\text{P5},1\text{h}} + T_{\text{P5},2\text{h}}$. For both datasets, $T_{\text{P5},2\text{h}}$ makes up around 60% of T_{P5} for small filter radii combinations, and for larger radii combinations, we see a minimal contribution of 30%. Although $T_{\text{P5},2\text{h}}$ rapidly decreases with field size, at the field size of the SLICS simulations, the contribution is still significant. So, whereas approximating $T_{\text{P5},1\text{h}}$ by T_{P5}^∞ does not introduce a significant error, the large-field approximation is not valid for T_{P5} on SLICS-sized fields, and $T_{\text{P5},2\text{h}}$ still has to be added to find a meaningful result. The bottom right panel of Fig. 2 shows the fractional difference between the large-field approximation of the cross-covariance $C_{\hat{\mathcal{M}}_{\text{ap}}^{3,2}}^\infty = T_{\text{PB},3}^\infty + T_{\text{P5},1\text{h}}^\infty$ plus the two-halo term $T_{\text{P5},2\text{h}}$ and the cross-covariance $C_{\hat{\mathcal{M}}_{\text{ap}}^{3,2}}$. Almost all filter combinations show an agreement within 5%. The filter combination $(16', 16', 16'; 4')$ is the only one showing a larger discrepancy ($\sim 12\%$).

4.3.2. Validation with N -body simulations

In Figs. 3 and 4, we compare the individual terms of the cross-covariance, both analytical and numerical, for the convergence maps and the shear catalogues, respectively. For all aperture radii combinations, $T_{\text{P5},1\text{h}} + T_{\text{P5},2\text{h}}$ (which depends on the tetraspectrum) dominates. Furthermore, we observe that for almost all combinations, $T_{\text{PB},2}$ is sub-dominant.

T_{P5} in general slightly over-predicts its numerical counterpart. The analytical $T_{\text{PB},3}$ is marginally smaller than the numerical one for the convergence maps. Conversely, for the shear catalogues, we see that the analytical $T_{\text{PB},3}$ is generally larger than the numerical one. The cause of the minor discrepancies between the numerical and analytical terms can be various. First, we expect the power-, bi-, and tetraspectrum to differ slightly in our model code and the SLICS simulation (see Appendix A in Heydenreich et al. 2023 for an example of the discrepancy between the measured and modelled bispectrum of the `BiHalofit`). This can be due to the binning, finite resolution, smoothing of the simulation fields, and uncertainties in `Halofit` and `BiHalofit`. Also, we describe T_{P5} with the halo model, which is an approximate model. Moreover, we only include the most significant contribution of the two-halo term. Taking these discrepancies into account, it can be stated that the numerical and analytical terms agree reasonably well. Only for the SLICS with shape noise, we see that the modelled $T_{\text{PB},2}$ does not match the one measured from the simulation. This is probably due to the difficulties in measuring a finite-field term when noise is included. Taking the large-field approximation in Eq. (34) would mean neglecting $E_A(\boldsymbol{\eta})$, after which the integral is, by definition, zero. That means the only reason it is non-zero for a finite field is the inclusion of $E_A(\boldsymbol{\eta})$. This function is a slowly varying function that leads to an amplification of the shape noise in the measured correlation functions. We, therefore, do not expect an exact match between the analytical and measured $T_{\text{PB},2}$. Fortunately, as the finite-field term is sub-dominant, this has a negligible effect on the total cross-covariance. For the shear catalogues, we show the

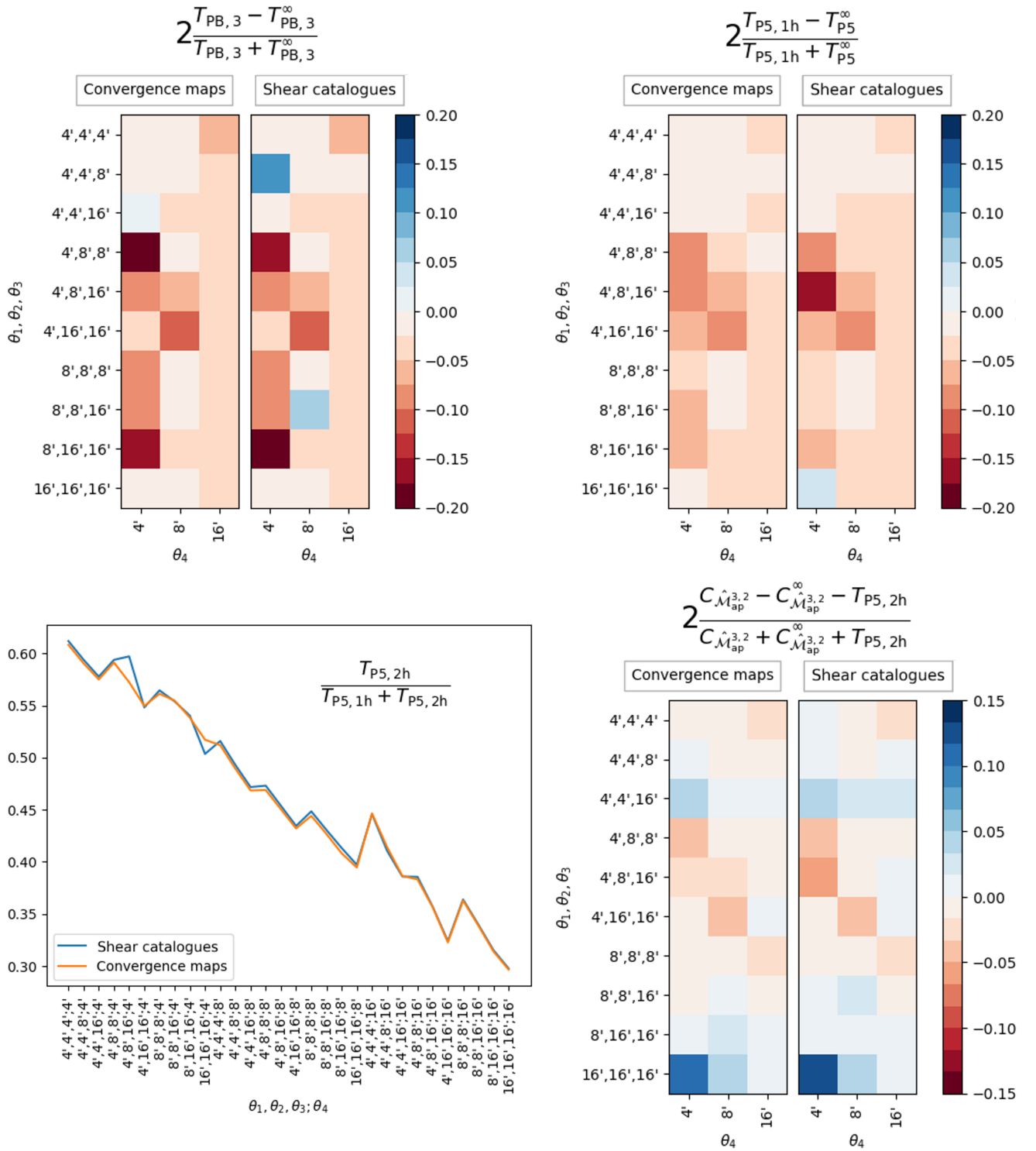


Fig. 2. *Top left:* fractional difference between $T_{PB,3}$ and $T_{PB,3}^\infty$ for the convergence maps and the shear catalogues with shape noise. *Top right:* fractional difference between $T_{P5,1h}$ and T_{P5}^∞ for the convergence maps and the shear catalogues. T_{P5}^∞ does not include the two-halo term $T_{P5,2h}$ as it is a finite-field term and therefore zero in the large-field approximation (Appendix A). *Bottom left:* relative contribution of the two-halo term $T_{P5,2h}$ to the total T_{P5} . The blue line for the shear catalogues and the orange line for the convergence maps of the SLICS. *Bottom right:* fractional difference between the analytical cross-covariance and its large-field approximation plus the two-halo term $T_{P5,2h}$ for both datasets.

total analytical cross-covariance and the sample covariance, including error bars, in Fig. 4. The model cross-covariance is just the sum of the separate terms. It is seen that the tetraspectrum term almost exclusively determines it.

Also shown in the heat maps in Figs. 3 and 4 is the difference between the analytical cross-covariance and the sample covariance, normalised by the bootstrapping error. We observe that all filter combinations agree within $3.5\sigma_{\text{boot}}$, with more than 50% lying within

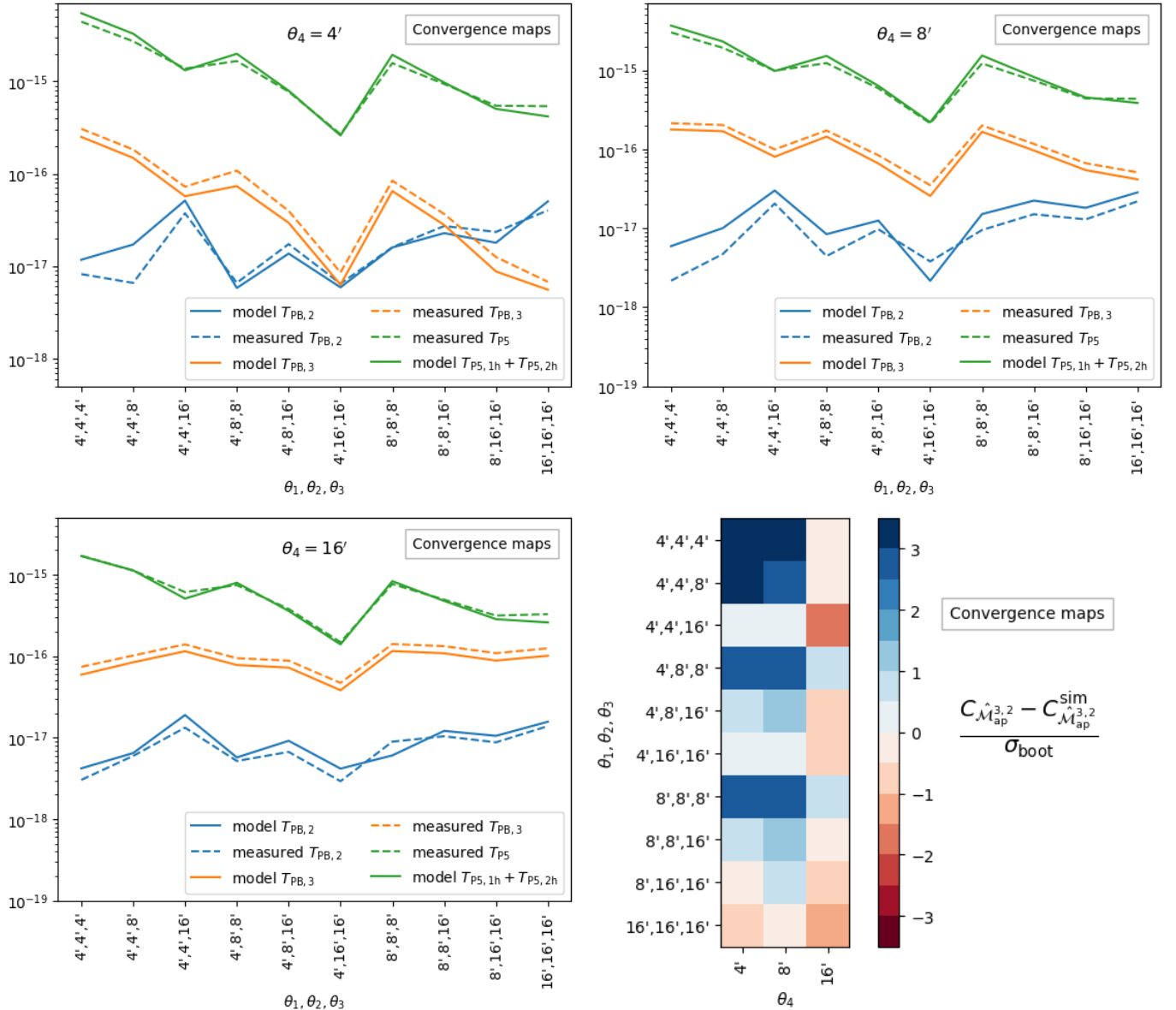


Fig. 3. Individual terms of the model cross-covariance, plotted together with the numerically measured ones of the SLICS without shape noise (convergence maps). The top left plot displays the terms when fixing $\theta_4 = 4'$, the top right when fixing $\theta_4 = 8'$, and the bottom left fixing $\theta_4 = 16'$. The bottom right heat map shows the difference between the modelled cross-covariance and the sample covariance normalised by the bootstrapping error.

the $1\sigma_{\text{boot}}$ range. The model slightly over-predicts the sample covariance for most filter combinations, with the largest differences occurring for the smaller filter combinations. This is likely due to the differences mentioned earlier.

5. Cosmological parameter estimation

With an analytical cross-covariance of the second- and third-order aperture mass statistics now available, we can combine it with the analytical covariances of \hat{M}_{ap}^2 and \hat{M}_{ap}^3 to form a joint covariance matrix. The left panel of Fig. 5 illustrates the fractional difference between the analytical and numerical joint covariance matrices for the SLICS with shape noise (shear catalogues). In this section, we compare the posterior distribution of S_8 and Ω_m obtained from the analytical joint covariance with those derived from the SLICS simulation with shape noise. We use a Markov Chain Monte Carlo (MCMC) sampling method for this comparison. It should be noted that the purpose of this paper is not to provide realistic estimates for either the clustering or the matter density parameter. Rather, this section aims to demonstrate that similar posterior distributions can be obtained from the analytical model when compared to the numerical posteriors. The posterior distribution of the analytical model is proportional to

$$P[\mathbf{m}(\Theta) | \mathbf{d}, \tilde{C}] \propto |\tilde{C}|^{-\frac{1}{2}} e^{-\chi^2/2}, \quad (42)$$

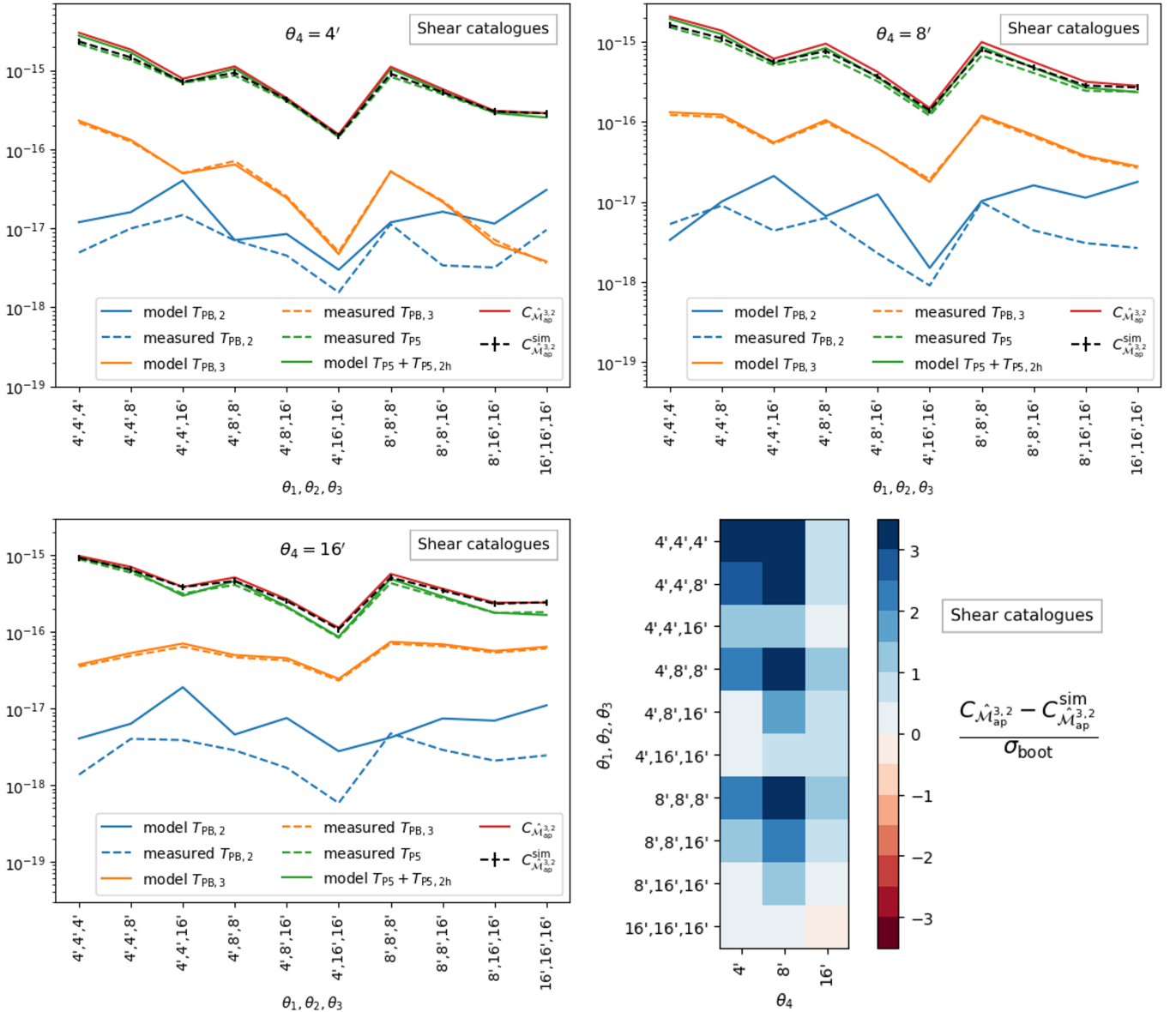


Fig. 4. Individual terms of the analytical cross-covariance, plotted together with the numerically measured ones of the SLICS with shape noise (shear catalogues). The total analytical cross-covariance and the measured sample covariance are also plotted. The top left plot displays the terms when fixing $\theta_4 = 4'$, the top right when fixing $\theta_4 = 8'$, and the bottom left plot when fixing $\theta_4 = 16'$. The bottom right heat map shows the difference between the modelled cross-covariance and the sample covariance normalised by the bootstrapping error.

where \mathbf{m} is the model vector containing the predicted values of $\langle \mathcal{M}_{\text{ap}} \rangle^2(\theta_i)$ and $\langle \mathcal{M}_{\text{ap}} \rangle^3(\theta_j, \theta_k, \theta_l)$ given the cosmological parameters Θ ; \mathbf{d} is the data vector, containing the measured second- and third-order aperture mass statistics; and \tilde{C} is the corresponding covariance matrix. For the numerical parameter estimation, we need to consider that we estimated the covariance from a finite sample. A correction for this is described in Percival et al. (2022). The numerical posterior distribution is proportional to

$$P[\mathbf{m}(\Theta) | \mathbf{d}, \tilde{C}] \propto |\tilde{C}|^{-\frac{1}{2}} \left(1 + \frac{\chi^2}{n_r - 1} \right)^{-m/2}, \quad (43)$$

where n_r is the number of realisations. m is a power-law index

$$m = n_\theta + 2 + \frac{n_r - 1 + B(n_d - n_\theta)}{1 + B(n_d - n_\theta)}, \quad \text{with } B = \frac{n_r - n_d - 2}{(n_r - n_d - 1)(n_r - n_d - 4)}, \quad (44)$$

n_θ being the number of parameters Θ and n_d the dimension of \mathbf{d} . The χ^2 -distribution in Eqs. (42) and (43) is calculated as

$$\chi^2 = [\mathbf{m}(\Theta) - \mathbf{d}]^T \tilde{C}^{-1} [\mathbf{m}(\Theta) - \mathbf{d}]. \quad (45)$$

We make our comparison by substituting \tilde{C} by C_{joint} and $C_{\text{joint}}^{\text{sim}}$, which are the analytical and numerical joint covariances, respectively, into Eqs. (42), (43), and (45). The numerical covariance $C_{\text{joint}}^{\text{sim}}$ is a positive definite matrix as it is calculated from the sample

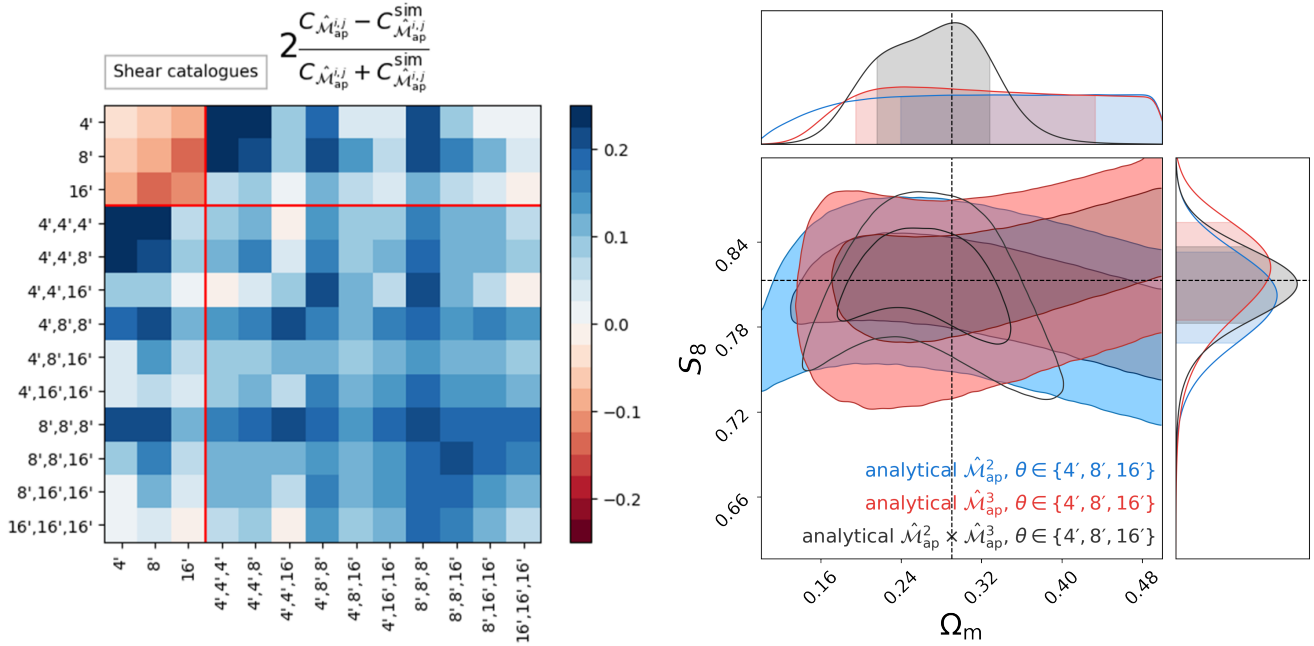


Fig. 5. *Left:* fractional difference between the analytical and numerical joint covariance matrix of the second- and third-order aperture mass statistics. The red lines divide the parts of the joint covariance matrix. The 3×3 matrix at the top left is the covariance matrix of \hat{M}_{ap}^2 calculated in Linke et al. (2024). The 10×10 matrix at the bottom right is the covariance matrix of \hat{M}_{ap}^3 from L+23. The non-square matrices at the top right and bottom left are the cross-covariance derived in this paper. *Right:* analytical S_8 – Ω_m contours from the second- and third-order aperture mass statistics, shown separately and in combination.

covariance. Our analytic covariance, as calculated to match the SLICS with shape noise, is not positive definite due to approximations and numerical errors, which poses a problem. Some of the subdominant eigenvalues of C_{joint} are found to be negative. This leads to a χ^2 -distribution with a non-zero probability of finding negative χ^2 values, leading to incorrect posterior distributions.

To mitigate this problem, we approximate our joint analytic covariance with a principal component analysis (PCA). This method corresponds to performing an eigendecomposition of the covariance matrix and cutting away the subdominant (negative) eigenvalues and the corresponding eigenvectors. Because our covariance matrix is ill-conditioned, meaning the eigenvalues are close to zero, numerical instability can arise in the eigendecomposition, causing issues in the accuracy of the eigenvalues and eigenvectors. These instabilities arise due to finite precision in computer calculations, causing small perturbations or rounding errors. A numerically more stable method is the singular value decomposition (SVD). It is a generalisation to the eigendecomposition and is theoretically equivalent for symmetric matrices. Performing an SVD factorises our covariance matrix

$$C_{\text{joint}} = U \Sigma V^T, \quad (46)$$

where Σ is a matrix with the singular values on the diagonal in descending order (which correspond to the absolute values of the eigenvalues for a symmetric matrix), and U and V^T are real matrices with orthonormal columns and rows, respectively. The diagonal matrix scales coordinates, and the matrices U and V^T represent rotations and reflections of the space. Up to different signs per column, U and V are equivalent to the matrix of eigenvectors of C_{joint} (but U and V are not the same; only the absolute values of the components are). We then cut away the diagonal terms corresponding to the negative eigenvalues and the corresponding columns and rows of U and V^T , respectively. In our case, the original matrices are square 13×13 matrices. After the principal component analysis, \tilde{U} is a 13×10 , $\tilde{\Sigma}$ a 10×10 , and \tilde{V}^T a 10×13 matrix (as we had three negative eigenvalues), where the tildes represent the matrices after the PCA, such that

$$\tilde{C}_{\text{joint}} = \tilde{U} \tilde{\Sigma} \tilde{V}^T. \quad (47)$$

The new matrix \tilde{C}_{joint} is a positive semi-definite matrix. As \tilde{C}_{joint} is a rank 10 matrix with dimensions 13×13 , it is singular. Hence, it is non-invertible. Taking the inverse of the separate terms on the right-hand side of Eq. (47) is thus not possible as \tilde{U} and \tilde{V} are not square matrices. Therefore, we take their pseudoinverse (denoted as a cross). As \tilde{U} has orthonormal columns and \tilde{V}^T orthonormal rows, their pseudoinverse matrices are equal to their transpose matrices ($\tilde{\Sigma}$ has an inverse, and thus the pseudoinverse is equal to the inverse)

$$(\tilde{U} \tilde{\Sigma} \tilde{V}^T)^+ = (\tilde{V}^T)^+ \tilde{\Sigma}^+ \tilde{U}^+ = \tilde{V} \tilde{\Sigma}^{-1} \tilde{U}^T. \quad (48)$$

Substituting this approximation into Eq. (45), we find

$$\chi^2 \approx [\mathbf{m}(\Theta) - \mathbf{d}]^T \tilde{V} \tilde{\Sigma}^{-1} \tilde{U}^T [\mathbf{m}(\Theta) - \mathbf{d}], \quad (49)$$

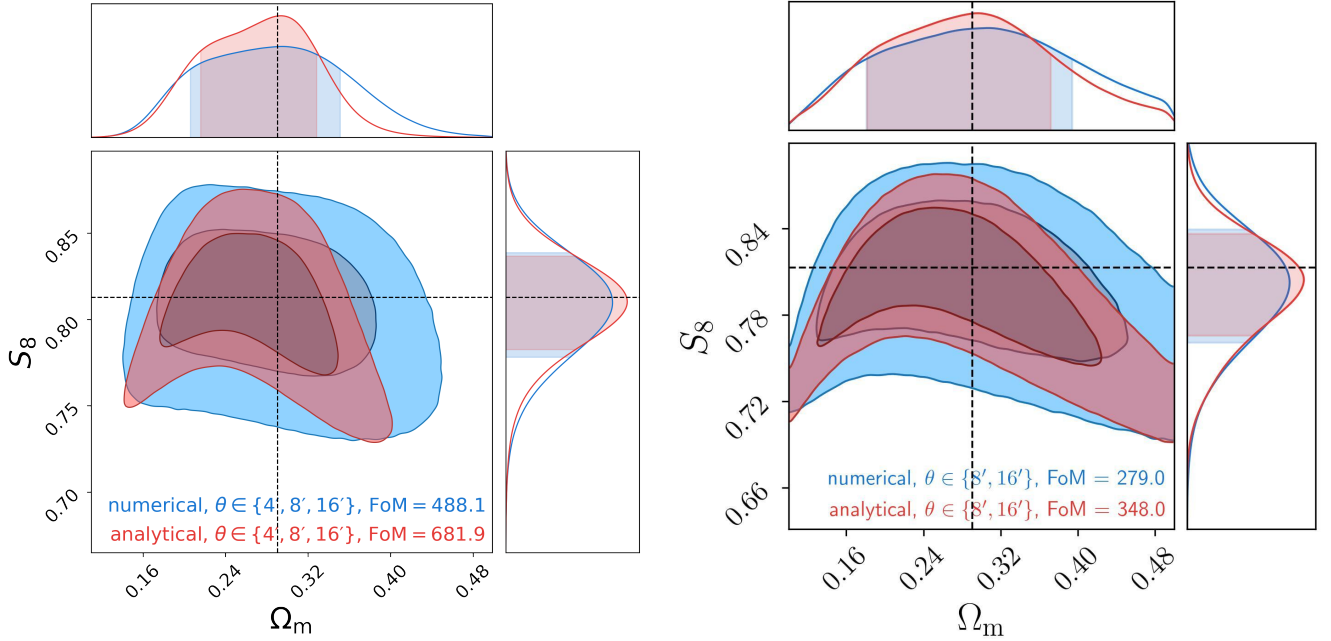


Fig. 6. Posterior contours of S_8 with Ω_m obtained via MCMC. The blue contour is obtained via the sample covariance of the SLICS shear catalogues. The red contour is calculated via the analytical covariance. *Left:* using all the filter radii combinations in $\theta \in \{4', 8', 16'\}$. *Right:* using only the combinations of $\theta \in \{8', 16'\}$.

which has a strictly positive χ^2 -distribution. With this set-up, we perform a Metropolis-Hastings MCMC, trained on 6500 model points sampled in a Latin hypercube spanning the cosmological parameter space Ω_m and S_8 . We use an emulator adopting the CosmoPower framework from [Spurio Mancini et al. \(2022\)](#) to efficiently compute the model predictions $\langle \hat{M}_{\text{ap}}^2(\theta_i) \rangle$ and $\langle \hat{M}_{\text{ap}}^3(\theta_j, \theta_k, \theta_l) \rangle$ for any given parameter vector Θ . The data vector \mathbf{d} is constructed from the emulator evaluated at the fiducial cosmological parameters corresponding to the SLICS simulations. All other cosmological parameters not varied in the analysis are fixed to their SLICS values.

The right panel of Fig. 5 shows the analytical posterior contours for \hat{M}_{ap}^2 , \hat{M}_{ap}^3 , and their joint posterior contour $\hat{M}_{\text{ap}}^2 \times \hat{M}_{\text{ap}}^3$. The degeneracy directions of the second- and third-order aperture mass statistics differ noticeably, resulting in significantly tighter constraints when the two are combined. A comparison between the analytical and numerical $S_8 - \Omega_m$ contours is shown in the left panel of Fig. 6. We calculate the figure of merit (FoM) with a parameter covariance \hat{C} obtained from the MCMC process as

$$\text{FoM} = \frac{1}{\sqrt{\det \hat{C}}} . \quad (50)$$

The fraction of the numerical over the analytical FoM is 0.72. The analytical posteriors are noticeably smaller than the numerical ones. In the left panel of Fig. 5, the largest discrepancies between the analytical and numerical cross-covariance occur at smaller filter radii scales (combinations with $4'$). Therefore, we also perform an MCMC with combinations of $8'$ and $16'$ only, as shown in the right panel of Fig. 6. Excluding $4'$, C_{joint} does not contain any negative eigenvalues, so we did not perform a PCA. Here, the fraction of the numerical over the analytical FoM is 0.80, which is an improvement compared to when all filter radii combinations are included.

6. Discussion

We derived an analytical model for the non-tomographic cross-covariance between the real-space estimators \hat{N}_{ap}^2 and \hat{N}_{ap}^3 , and validated it against simulated data from SLICS, including both shape-noise-free convergence maps and shear catalogues with noise. The cross-covariance from the simulations was numerically determined through either the sample covariance or aperture mass correlation functions. For our selected set of aperture radii combinations, we found that the maximum deviation between the analytical and numerical results was 3.5 standard deviations on the bootstrapping uncertainty, with over 70% of the cases falling well within the two-sigma range. This was true for both datasets. The most likely cause of the discrepancies is a difference between the polyspectra used in the model and those from the simulation.

6.1. The large-field approximation

We find that the large-field approximation has a negligible effect on the terms $T_{\text{P5,1h}}$ and $T_{\text{PB,3}}$. These can therefore be safely replaced by their large-field counterparts, T_{P5}^{∞} and $T_{\text{PB,3}}^{\infty}$, for a SLICS sized or larger field, if a reduction in computational cost is desired or if the integration over the survey window function G_A leads to numerical instability. The finite-field term $T_{\text{PB,2}}$ is consistently

subdominant across the range of filter combinations tested. Its omission has little impact on the overall cross-covariance and it can be neglected without significantly affecting the result. In contrast, the finite-field term $T_{P5,2h}$ is essential for achieving an accurate prediction. For combinations involving small aperture radii, it can contribute up to 60% of the total T_{P5} , which is the dominant component of the cross-covariance. Neglecting this term would therefore lead to a substantial underestimation of the covariance at small scales. Since the amplitude of finite-field corrections decreases more rapidly than linearly with increasing survey area, their relevance is expected to diminish in Stage IV surveys. In particular, for significantly larger survey fields than those of SLICS, the contribution from $T_{PB,2}$ is negligible. Nonetheless, we recommend retaining $T_{P5,2h}$ in any analysis that includes small-scale filters, even for larger-area surveys.

6.2. Analytical vs. numerical cross-covariance and origins of inaccuracy

The analytical cross-covariance model presented in this work agrees well with the numerical results from simulations across a broad range of filter combinations. For the configurations tested, 50% of the analytical predictions fall within the 1σ range of the bootstrap uncertainties, with the largest individual deviation reaching 3.5σ . The agreement is notably better for combinations involving larger aperture radii ($16'$), while the largest discrepancies arise at small filter scales. Testing the model at even larger aperture radii, such as $32'$, in larger-volume simulations, for example, the full-sky GL simulations of Takahashi et al. (2017), would provide valuable insight into the behaviour of the analytic cross-covariance at lower ℓ and help isolate contributions from different model components.

There are several possible reasons for the observed discrepancies at smaller filter radii. The most significant source of inaccuracy is most likely the halo model used to compute the tetraspectrum term T_{P5} , which dominates the cross-covariance at all aperture scales. Among its components, the two-halo contribution $T_{P5,2h}$ is particularly impactful at small radii. As no more accurate tetraspectrum model is currently available, this approximation remains necessary, though the inclusion of additional two-halo contributions may improve the model's fidelity, especially for the small filter scales, where $T_{P5,2h}$ is dominant. While the power spectrum and bispectrum models used to compute $T_{PB,2}$ and $T_{PB,3}$ also differ from those implicit in the simulations, these terms are subdominant, typically by one to two orders of magnitude, and thus contribute less to the overall discrepancy.

A secondary limitation arises from numerical integration accuracy. The calculation of T_{P5} involves high-dimensional integrals that are computationally expensive and sensitive to binning. Due to performance constraints, we employed relatively coarse binning in parts of the integration, which limits the achievable precision. Some of the residual discrepancies are likely due to this numerical limitation.

Several other approximations were used in the model but are unlikely to contribute significantly to the observed discrepancies. First, the flat sky approximation was used. This is a valid approximation, as the scales probed in this are below $\sim 2^\circ$, where the induced error is negligible (Kilbinger et al. 2017). Moreover, this approximation was also applied in the simulations, meaning it cannot account for any discrepancies. However, it is important to consider this approximation when extending to larger fields, where larger filter radii can be probed, as it is expected to introduce greater errors at those scales. Additionally, the Limber approximation was used. This approximation becomes unreliable on scales larger than 6° (Deshpande & Kitching 2020). However, non-linear structure growth is minimal at these scales, meaning third-order statistics provide little additional information, making this limitation largely inconsequential.

In conclusion, the primary source of inaccuracy in the analytical cross-covariance model lies in the treatment of the tetraspectrum. Numerical integration precision contributes to a lesser extent, whereas other standard approximations, such as the flat-sky and Limber approximations, have a negligible impact on the angular scales considered.

6.3. Analytic vs. numerical cosmological inference

By combining the analytical cross-covariance derived here with the analytical \hat{M}_{ap}^3 covariance of L+23 and the \hat{M}_{ap}^2 covariance of Linke et al. (2024), we constructed a joint analytical covariance matrix for use in cosmological parameter inference. We then compared its performance to the numerically derived covariance from SLICS in constraining the parameters Ω_m and S_8 . The resulting constraints are broadly consistent between the two approaches. The analytical model yields slightly tighter confidence contours (Fig. 6), indicating slightly optimistic constraining power. However, interpreting this difference requires care, as the overall agreement of the covariance components is not uniform. The fractional difference matrix in the left panel of Fig. 5 shows that the analytical \hat{M}_{ap}^2 auto-covariance underpredicts its numerical counterpart, while both the cross-covariance and the \hat{M}_{ap}^3 auto-covariance tend to overpredict. This asymmetry suggests that the origin of the discrepancy in the contours is not easily attributable to a single term, such as the cross-covariance. In addition to the discrepancies for the cross-covariance mentioned before, a plausible explanation for the underpredicting \hat{M}_{ap}^2 auto-covariance lies in the modelling of the trispectrum. The analytical \hat{M}_{ap}^2 covariance includes a trispectrum term computed using only the one-halo contribution. If the trispectrum extracted from simulations (e.g. SLICS) is larger due to two-halo or higher-order contributions, this would lead to an underestimation of the variance in the analytical \hat{M}_{ap}^2 covariance, thereby shifting the contours accordingly. In contrast, the cross-covariance and \hat{M}_{ap}^3 covariance may slightly overpredict their counterparts due to different approximations or modelling assumptions.

In practice, the joint analytical covariance matrix is prone to numerical instability. Small-scale inaccuracies and inconsistencies among components can cause the matrix to become ill-conditioned, complicating operations such as matrix inversion. To mitigate this, PCA and SVD were applied to regularise the matrix. While PCA improves numerical stability, it also removes information from the data vector, leading to a slight degradation of the constraining power. These numerical issues highlight the sensitivity of cosmological inference to the stability and internal consistency of the covariance matrix, particularly when combining multiple orders

of statistics. Future improvements to analytical models should aim not only for greater physical accuracy but also for improved numerical robustness.

6.4. Further comments

In realistic surveys, data often contain gaps due to masking, which prevents the direct computation of aperture mass statistics from convergence or shear maps using Eq. (6). Aperture centres located near masked regions lead to weighted integrals over incomplete areas, introducing biases in the aperture mass. To address this, the estimators $\langle \mathcal{M}_{\text{ap}}^2 \rangle$ and $\langle \mathcal{M}_{\text{ap}}^3 \rangle$ should instead be derived from their relations to the two- and three-point shear correlation functions (2PCF and 3PCF), as proposed by Schneider et al. (2005). However, even if masking is not explicitly accounted for, L+23 demonstrated that the analytical $\hat{\mathcal{M}}_{\text{ap}}^3$ covariance can still provide meaningful constraints by establishing upper and lower bounds. The upper bound is computed assuming a field size equal to the total unmasked survey area, while the lower bound excludes a boundary region equal to four times the largest aperture radius. For a SLICS-sized field, these bounds were found to be tight: the upper bound exceeded the lower bound by only $\sim 15\%$. In Stage IV surveys, the relative size of the excluded boundary will be smaller, implying even tighter bounds. Although we have not explicitly tested this for the analytical cross-covariance derived here, the underlying scaling arguments suggest that similar bounds should apply. This makes the analytical cross-covariance a practical tool even in the presence of realistic masking.

The analytical cross-covariance between $\hat{\mathcal{M}}_{\text{ap}}^2$ and $\hat{\mathcal{M}}_{\text{ap}}^3$ is closely related to the Fourier-space cross-covariance between the 3D power spectrum and bispectrum, as calculated by Sugiyama et al. (2020). While their formalism is three-dimensional, their approach closely parallels ours and can be translated to the 2D case via the Limber approximation. A direct comparison reveals that their expressions correspond to our large-field terms. Specifically, their Eq. (23) maps to our $T_{\text{PB},3}^{\infty}$ and their Eq. (26) corresponds to our T_{PS}^{∞} . This correspondence is evident from the structural similarity of their k -space with our ℓ -space configurations. Sugiyama et al. (2020) note that their expressions do not include super-sample covariance (SSC) terms, which explains the absence of certain contributions. In Fourier space, SSC arises from modes larger than the survey area and must be added explicitly. In contrast, real-space calculations naturally incorporate these large-scale modes within the measured survey window, eliminating the need for separate SSC corrections. A detailed discussion of this distinction, and the implicit inclusion of SSC in real-space statistics, is provided in Linke et al. (2024).

6.5. Future work

One clear avenue for extending this work is the incorporation of tomography. A tomographic approach provides significantly more information on the evolution of large-scale structure and is particularly effective at reducing degeneracies between cosmological parameters such as Ω_m and S_8 . It also enables meaningful constraints on a time-varying dark energy equation of state, $w(z)$. However, extending the current model to a tomographic framework is non-trivial. It leads to a substantial increase in the size of both the data vector and the covariance matrix, which increases computational cost and the risk of numerical instability. Analytical derivation of the tomographic covariance is, however, still more feasible than relying on simulations, given the substantial resources required to generate tomographic mock catalogues, but it introduces new challenges for validation. In particular, as simulation-based benchmarks are often unavailable at high dimensionality, ensuring the numerical stability and internal consistency of tomographic analytical models remains a key difficulty.

While this work focused on the cross-covariance between aperture mass variance and skewness, the combination $\hat{\mathcal{M}}_{\text{ap}}^2 \times \hat{\mathcal{M}}_{\text{ap}}^3$ is not the optimal choice for joint second- and third-order statistics in real data analyses. A primary limitation is that aperture mass statistics require knowledge of the two- and three-point correlation functions down to arbitrarily small angular separations and, formally, up to infinite separation. In practice, this is problematic as close galaxy pairs may be blended or lost to detection limits in real data and finite resolution and pixelisation prevent accurate small-scale correlation measurements in simulations. These limitations can lead to E -/ B -mode leakage and loss of accuracy. A more robust alternative is to use COSEBIs (Complete Orthogonal Sets of E -/ B -mode Integrals) for the second-order component (Schneider et al. 2010). COSEBIs are designed to operate over a finite angular range $[\theta_{\text{min}}, \theta_{\text{max}}]$, with filter functions that are both complete and orthogonal. This ensures the mathematically exact separation of the E - and B -modes. While some residual leakage may occur due to numerical effects or masking, COSEBIs are considerably less sensitive to the limitations described above. Furthermore, COSEBIs offer a natural data compression: typically, the first 6 modes capture the majority of the cosmological information, even in tomographic analyses (Wright et al. 2025). A promising next step is therefore the development of an analytical cross-covariance between COSEBIs and $\hat{\mathcal{M}}_{\text{ap}}^3$, which would combine robustness, compactness, and sensitivity to non-Gaussian features. Additionally, second-order cosmic shear studies have also used the shear correlation functions ξ_{\pm} , so a joint $\xi_{\pm} \times \mathcal{M}_{\text{ap}}^3$ covariance would also be a valuable extension.

6.6. Summary

In this work, we presented the first analytical real-space calculation of the cross-covariance between second- and third-order aperture mass statistics, completing the joint covariance matrix required for cosmological parameter estimation. For upcoming Stage IV surveys, analytical covariance matrices offer several advantages over simulation-based ones, including flexibility, scalability, and reduced noise. These benefits are most effectively realised when combined with COSEBIs for second-order statistics, and when extended to tomographic configurations. Nonetheless, limitations including the approximations inherent in the halo model, numerical instabilities in matrix operations, and difficulties in validation continue to hinder the broader adoption of analytical covariance matrices. Further developments are required to improve model accuracy, reduce numerical sensitivity, and enabling a tomographic set-up.

Acknowledgements. We would like to thank Joachim Harnois-Déraps for making public the SLICS mock data, which can be found at <http://slics.roe.ac.uk/>. We thank the anonymous referee for the useful comments. Funded by the TRA Matter (University of Bonn) as part of the Excellence Strategy of the federal and state governments. This work has been supported by the Deutsche Forschungsgemeinschaft through the project SCHN 342/15-1 and DFG SCHN 342/13. LL is supported by the Austrian Science Fund (FWF) [ESP 357-N]. PAB and SH acknowledge support from the German Academic Scholarship Foundation. LP acknowledges support from the DLR grant 50QE2002. SH is supported by the US Department of Energy, Office of High Energy Physics under Award Number DE-SC0019301.

References

- Asgari, M., Lin, C.-A., Joachimi, B., et al. 2021, *A&A*, **645**, A104
 Bartelmann, M. 2010, *Class. Quant. Grav.*, **27**, 233001
 Bartelmann, M., & Schneider, P. 2001, *Phys. Rep.*, **340**, 291
 Burger, P., Friedrich, O., Harnois-Déraps, J., & Schneider, P. 2022, *A&A*, **661**, A137
 Burger, P. A., Porth, L., Heydenreich, S., et al. 2024, *A&A*, **683**, A103
 Cooray, A., & Sheth, R. 2002, *Phys. Rep.*, **372**, 1
 Crittenden, R. G., Natarajan, P., Pen, U.-L., & Theuns, T. 2002, *ApJ*, **568**, 20
 Dalal, R., Li, X., Nicola, A., et al. 2023, *Phys. Rev. D*, **108**, 123519
 DES Collaboration 2022, *Phys. Rev. D*, **105**, 023520
 Deshpande, A. C., & Kitching, T. D. 2020, *Phys. Rev. D*, **101**, 103531
 Di Valentino, E., Anchordoqui, L. A., Özgür, A., et al. 2021, *Astropart. Phys.*, **131**, 102604
 Euclid Collaboration (Ajani, V., et al.) 2023, *A&A*, **675**, A120
 Euclid Collaboration (Mellier, Y., et al.) 2025, *A&A*, **697**, A1
 Falco, E. E., Gorenstein, M. V., & Shapiro, I. I. 1985, *ApJ*, **289**, L1
 Gruen, D., Friedrich, O., Krause, E., et al. 2018, *Phys. Rev. D*, **98**, 023507P
 Harnois-Déraps, J., & van Waerbeke, L. 2015, *MNRAS*, **450**, 2857
 Harnois-Déraps, J., Amon, A., Choi, A., et al. 2018, *MNRAS*, **481**, 1337
 Harnois-Déraps, J., Martinet, N., Castro, T., et al. 2021, *MNRAS*, **506**, 1623
 Heydenreich, S., Schneider, P., Hildebrandt, H., et al. 2020, *A&A*, **634**, A104
 Heydenreich, S., Brück, B., & Harnois-Déraps, J. 2021, *A&A*, **648**, A74
 Heydenreich, S., Linke, L., Burger, P., & Schneider, P. 2023, *A&A*, **672**, A44
 Heymans, C., Tröster, T., Asgari, M., et al. 2021, *A&A*, **646**, A140
 Hildebrandt, H., Viola, M., Heymans, C., et al. 2017, *MNRAS*, **465**, 1454
 Hoekstra, H., & Jain, B. 2008, *Ann. Rev. Nucl. Part. Sci.*, **58**, 99
 Joudaki, S., Hildebrandt, H., Traykova, D., et al. 2020, *A&A*, **638**, L1
 Kaiser, N., & Jaffe, A. 1997, *ApJ*, **484**, 545
 Kayo, I., Takada, M., & Jain, B. 2013, *MNRAS*, **429**, 344
 Kilbinger, M., & Schneider, P. 2005, *A&A*, **442**, 69
 Kilbinger, M., Heymans, C., Asgari, M., et al. 2017, *MNRAS*, **472**, 2126
 Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, ArXiv e-prints [arXiv:1110.3193]
 Limber, D. N. 1954, *ApJ*, **119**, 655
 Linke, L., Heydenreich, S., Burger, P. A., & Schneider, P. 2023, *A&A*, **672**, A185
 Linke, L., Burger, P. A., Heydenreich, S., Porth, L., & Schneider, P. 2024, *A&A*, **681**, A33
 Martinet, N., Schneider, P., Hildebrandt, H., et al. 2018, *MNRAS*, **474**, 712
 Percival, W. J., Friedrich, O., Sellentin, E., & Heavens, A. 2022, *MNRAS*, **510**, 3207
 Planck Collaboration VI. 2020, *A&A*, **641**, A6
 Porth, L., Heydenreich, S., Burger, P., Linke, L., & Schneider, P. 2024, *A&A*, **689**, A227
 Pyne, S., & Joachimi, B. 2021, *MNRAS*, **503**, 2300
 Schneider, P. 1996, *MNRAS*, **283**, 837
 Schneider, P., & Lombardi, M. 2003, *A&A*, **397**, 809
 Schneider, P., & Seitz, C. 1995, *A&A*, **294**, 411
 Schneider, P., Kilbinger, M., & Lombardi, M. 2005, *A&A*, **431**, 9
 Schneider, P., Eifler, T., & Krause, E. 2010, *A&A*, **520**, A116
 Sheth, R. K., & Tormen, G. 1999, *MNRAS*, **308**, 119
 Spurio Mancini, A., Piras, D., Alsing, J., Joachimi, B., & Hobson, M. P. 2022, *MNRAS*, **511**, 1771
 Sugiyama, N. S., Saito, S., Beutler, F., & Seo, H.-J. 2020, *MNRAS*, **497**, 1684
 Takada, M., & Jain, B. 2004, *MNRAS*, **348**, 897
 Takahashi, R., Sato, M., Nishimichi, T., Taruya, A., & Oguri, M. 2012, *ApJ*, **761**, 152
 Takahashi, R., Hamana, T., Shirasaki, M., et al. 2017, *ApJ*, **850**, 24
 Takahashi, R., Nishimichi, T., Namikawa, T., et al. 2020, *ApJ*, **895**, 113
 Wright, A. H., Stözlner, B., Asgari, M., et al. 2025, *A&A*, in press, <https://doi.org/10.1051/0004-6361/202554908>

Appendix A: The tetraspectrum halo model

In this appendix, we approximate the tetraspectrum using a halo model as in [Cooray & Sheth \(2002\)](#). The purpose of this estimation is to model T_{P5} (Eq. 24), which can be slightly modified and expressed as

$$T_{P5}(\theta_1, \theta_2, \theta_3; \theta_4) = \int \frac{d^2 \ell_1}{(2\pi)^2} \int \frac{d^2 \ell_2}{(2\pi)^2} \int \frac{d^2 s}{(2\pi)^2} \int \frac{d^2 \ell_4}{(2\pi)^2} P_5(\ell_1, \ell_2, \mathbf{s} - \ell_1 - \ell_2, \ell_4) \tilde{u}(\ell_1 \theta_1) \tilde{u}(\ell_2 \theta_2) \\ \times \tilde{u}(|\mathbf{s} - \ell_1 - \ell_2| \theta_3) \tilde{u}(\ell_4 \theta_4) \tilde{u}(|\mathbf{s} + \ell_4| \theta_4) G_A(\mathbf{s}), \quad (\text{A.1})$$

where $\mathbf{s} = \ell_1 + \ell_2 + \ell_3$ was used, such that the geometry factor only depends on a single input value. The tetraspectrum of the convergence is given by the Limber integration (Eq. 5),

$$P_5(\ell_1, \ell_2, \mathbf{s} - \ell_1 - \ell_2, \ell_4) = \left(\frac{3H_0^2 \Omega_m}{2c^2} \right)^5 \int_0^\infty d\chi \frac{q^5(\chi)}{\chi^3 a^5(\chi)} \mathcal{P}_5^{(3d)}(\ell_1/\chi, \ell_2/\chi, (\mathbf{s} - \ell_1 - \ell_2)/\chi, \ell_4/\chi, (-\mathbf{s} - \ell_4)/\chi; \chi), \quad (\text{A.2})$$

where the tetraspectrum of the density contrast can be decomposed into terms that include up to five halos. Specifically, the one-halo term corresponds to all five points in the correlation function lying within a single halo, while the two-halo term involves the points being distributed between two halos. As a result, the tetraspectrum of the density contrast can be expressed as

$$\mathcal{P}_5^{(3d)}(\ell_1/\chi, \ell_2/\chi, (\mathbf{s} - \ell_1 - \ell_2)/\chi, \ell_4/\chi, (-\mathbf{s} - \ell_4)/\chi; \chi) = I_5^0(\ell_1/\chi, \ell_2/\chi, (\mathbf{s} - \ell_1 - \ell_2)/\chi, \ell_4/\chi, (-\mathbf{s} - \ell_4)/\chi; \chi) \\ + P_L^{(3d)}(\ell_1/\chi) I_1^1(\ell_1/\chi; \chi) I_4^1(\ell_2/\chi, (\mathbf{s} - \ell_1 - \ell_2)/\chi, \ell_4/\chi, (-\mathbf{s} - \ell_4)/\chi; \chi) + 4 \text{ Perm.} \\ + P_L^{(3d)}(s/\chi) I_3^1(\ell_1/\chi, \ell_2/\chi, (\mathbf{s} - \ell_1 - \ell_2)/\chi; \chi) I_2^1(\ell_4/\chi, (\mathbf{s} - \ell_4)/\chi; \chi) + 9 \text{ Perm.} \\ + 3 \text{ halo terms} + 4 \text{ halo terms} + 5 \text{ halo terms}, \quad (\text{A.3})$$

where $P_L^{(3d)}$ is the linear density contrast power spectrum, and the I_n^l terms are defined as

$$I_n^l(\mathbf{k}_1, \dots, \mathbf{k}_n; \chi) = \int dm n[m, z(\chi)] \left(\frac{m}{\bar{\rho}} \right)^n b^l[m, z(\chi)] \prod_{i=1}^n \tilde{u}_{\text{NFW}}(\mathbf{k}_i, m), \quad (\text{A.4})$$

where m is the halo mass, $n(m, z)$ is the halo mass function from [Sheth & Tormen \(1999\)](#), $\bar{\rho}$ is the comoving mean density, $b(m, z)$ the halo bias ([Sheth & Tormen 1999](#)), and \tilde{u}_{NFW} is the Fourier transformed and normalised Navarro–Frenk–White-halo profile, which is defined as $u_{\text{NFW}}(r, m) = \rho(r, m)/m$, where ρ is the halo density profile. We note that when I_n^l terms are multiplied (meaning multiple halos), the masses, and thus also the integrations, are distinct per halo.

As the distance between halos is much greater than within one halo, the one-halo term will dominate on small scales (large ℓ/χ). However, in the one-halo term, we only have correlations up to the scale of the largest halo. Therefore, we expect the one-halo term to be subdominant on scales larger than halos. It is shown in [Pyne & Joachimi \(2021\)](#), that on large scales, in the assumption that $s \ll \ell_1, \ell_2, \ell_4$, the two-halo term depending on $P_L^{(3d)}(s/\chi)$ dominates. Therefore, we can approximate Eq. (A.3) as

$$\mathcal{P}_5^{(3d)}(\ell_1/\chi, \ell_2/\chi, (\mathbf{s} - \ell_1 - \ell_2)/\chi, \ell_4/\chi, (-\mathbf{s} - \ell_4)/\chi; \chi) \approx I_5^0(\ell_1/\chi, \ell_2/\chi, (\mathbf{s} - \ell_1 - \ell_2)/\chi, \ell_4/\chi, (-\mathbf{s} - \ell_4)/\chi; \chi) \\ + P_L^{(3d)}(s/\chi) I_3^1(\ell_1/\chi, \ell_2/\chi, (\mathbf{s} - \ell_1 - \ell_2)/\chi; \chi) I_2^1(\ell_4/\chi, (\mathbf{s} - \ell_4)/\chi; \chi) \\ \approx I_5^0(\ell_1/\chi, \ell_2/\chi, (-\ell_1 - \ell_2)/\chi, \ell_4/\chi, -\ell_4/\chi; \chi) + P_L^{(3d)}(s/\chi) I_3^1(\ell_1/\chi, \ell_2/\chi, (-\ell_1 - \ell_2)/\chi; \chi) I_2^1(\ell_4/\chi, -\ell_4/\chi; \chi), \quad (\text{A.5})$$

where in the second approximation, we neglected \mathbf{s} except for in the linear power spectrum, as $P_L^{(3d)}(0) = 0$. Now, filling in this approximation into Eq. (A.1), and using isotropy, yields

$$T_{P5}(\theta_1, \theta_2, \theta_3; \theta_4) \approx \left(\frac{3H_0^2 \Omega_m}{2c^2} \right)^5 \int \frac{d^2 \ell_1}{(2\pi)^2} \int \frac{d^2 \ell_2}{(2\pi)^2} \int \frac{d^2 s}{(2\pi)^2} \int \frac{d^2 \ell_4}{(2\pi)^2} \int_0^\infty d\chi \frac{q^5(\chi)}{\chi^3 a^5(\chi)} \\ \times \tilde{u}(\ell_1 \theta_1) \tilde{u}(\ell_2 \theta_2) \tilde{u}(|\ell_1 + \ell_2| \theta_3) \tilde{u}(\ell_4 \theta_4) \tilde{u}(|\mathbf{s} + \ell_4| \theta_4) G_A(\mathbf{s}) \\ \times \left[I_5^0(\ell_1/\chi, \ell_2/\chi, |\ell_1 + \ell_2|/\chi, \ell_4/\chi, \ell_4/\chi; \chi) + P_L^{(3d)}(s/\chi) I_3^1(\ell_1/\chi, \ell_2/\chi, |\ell_1 + \ell_2|/\chi; \chi) I_2^1(\ell_4/\chi, \ell_4/\chi; \chi) \right] \\ = T_{P5,1h}(\theta_1, \theta_2, \theta_3; \theta_4) + T_{P5,2h}(\theta_1, \theta_2, \theta_3; \theta_4), \quad (\text{A.6})$$

where

$$T_{P5,2h}(\theta_1, \theta_2, \theta_3; \theta_4) = \left(\frac{3H_0^2 \Omega_m}{2c^2} \right)^5 \int_0^\infty d\chi \frac{q^5(\chi)}{\chi^3 a^5(\chi)} \left[\int \frac{d^2 s}{(2\pi)^2} G_A(\mathbf{s}) P_L^{(3d)}(s/\chi) \right] \int \frac{d^2 \ell_1}{(2\pi)^2} \int \frac{d^2 \ell_2}{(2\pi)^2} \int \frac{d^2 \ell_4}{(2\pi)^2} \\ \times \tilde{u}(\ell_1 \theta_1) \tilde{u}(\ell_2 \theta_2) \tilde{u}(|\ell_1 + \ell_2| \theta_3) \tilde{u}(\ell_4 \theta_4) \tilde{u}(|\mathbf{s} + \ell_4| \theta_4) I_3^1(\ell_1/\chi, \ell_2/\chi, |\ell_1 + \ell_2|/\chi; \chi) I_2^1(\ell_4/\chi, \ell_4/\chi; \chi) \quad (\text{A.7})$$

will be zero in the large-field limit, where the integration over the Dirac delta function in Eq. (25) gives us a factor $P_L^{(3d)}(0) = 0$. Thus, we find a finite-field term similar to $T_{PB,2}$.