Astronomy & Astrophysics manuscript no. output
June 13, 2026

The shear-to-cosmology paradigm

I. Hybrid field-level and simulation-based framework for weak-lensing surveys

Jiacheng Ding[1], Chen Su[1], Ji Yao[1,⋆], Le Zhang[2,3,⋆⋆], and Huanyuan Shan[1,⋆⋆⋆]

[1] Shanghai Astronomical Observatory, Chinese Academy of Sciences, Shanghai 200030, P. R. China
[2] School of Physics and Astronomy, Sun Yat-Sen University, Zhuhai 519082, P. R. China
[3] CSST Science Center for the Guangdong–Hong Kong–Macau Greater Bay Area, SYSU, Zhuhai 519082, P. R. China

Received May 07, 2026

ABSTRACT

Context. Precise cosmological inference from next-generation weak-lensing surveys requires non-Gaussian information beyond two-point statistics, while conventional field-level analyses based on convergence reconstruction can introduce additional systematics.
Aims. This work aims to develop a scalable and robust framework that directly maps shear fields to cosmological parameters, exploiting non-Gaussian information while avoiding convergence reconstruction to improve constraints from stage-IV weak-lensing surveys.
Methods. We propose a hybrid machine-learning (ML) framework that combines field-level inference (FLI) with simulation-based inference (SBI). The FLI network performs field-level compression by extracting informative, non-Gaussian features directly from shear field, which are then passed to SBI to model complex posteriors. To mitigate noise from intrinsic galaxy shapes, we further developed a blind, training-free, PCA-based shear denoising method.
Results. Using CSST-like mock catalogs, we show that shear-based inference achieves an approximately twofold improvement in FoM over convergence-based approaches. Furthermore, within the shear-based framework, the combination of PCA denoising and ML compression yields a 36.4% FoM improvement relative to standard two-point statistics.
Conclusions. This work establishes a scalable framework for shear-based cosmological inference that unlocks field-level non-Gaussian information and significantly improves constraints for stage-IV weak-lensing surveys.

Key words. weak-lensing cosmology – field-level inference – simulation-based inference – shear denoising

1. Introduction

Weak lensing (WL) is a powerful probe in modern cosmology. According to general relativity, light rays are deflected when passing through inhomogeneous matter distributions, leading to subtle distortions in the observed shapes of background galaxies. By statistically analyzing the shape distortions of millions of galaxies, one can directly map the large-scale distribution of total matter, including both dark matter and baryons, across the cosmic web. This makes WL an essential tool for constraining cosmological parameters, testing theories of gravity, and investigating the nature of dark energy (Kilbinger 2015; Mandelbaum 2018; Meneghetti 2022). In addition, WL-based dark-matter (DM) reconstruction techniques (Massey et al. 2007; Van Waerbeke et al. 2013; Shan et al. 2014; Oguri et al. 2018) and DM halo identification methods (Hoekstra et al. 2004; Tang & Fan 2005; Shan et al. 2012) have played a key role in tracing the growth and spatial distribution of large-scale structures (LSSs) in the Universe.

Over the past decade, stage-III WL surveys such as DES (Dark Energy Survey Collaboration et al. 2016), KiDS, and HSC (Aihara et al. 2018) have mapped thousands of square degrees of the sky with dedicated pipelines and achieved 1% control of systematics, establishing a solid foundation for precision cosmology. Building on these advances, ongoing and upcoming stage-IV surveys, including the China Space Station Survey Telescope (CSST; Yao et al. 2024; CSST Collaboration et al. 2026; Gong et al. 2025), *Euclid* (Euclid Collaboration et al. 2025), LSST (Ivezić et al. 2019), and the *Roman* Space Telescope (Spergel et al. 2015) will expand sky coverage to tens of thousands of square degrees and deliver substantially improved measurement precision, enabling unprecedented tests of dark energy, gravity, and structure formation.

Traditionally, cosmological inference has relied on two-point statistics within a Bayesian framework (Planck Collaboration et al. 2020; Leclercq & Heavens 2021; Padilla et al. 2021; Taylor & Markovič 2022). However, several challenges become increasingly severe as data quality improves. Achieving stage-IV accuracy in modeling WL observables is a particularly demanding task due to nonlinear gravitational evolution, baryonic effects, and complex observational systematics, including PSF modeling, survey geometry, photometric redshift uncertainties, and intrinsic alignments (Hu 1999; Schneider et al. 2002; Bernstein 2009; Schrabback et al. 2010; Benjamin et al. 2013; Yao et al. 2023).

In addition, late-time cosmic fields are highly non-Gaussian, such that two-point statistics capture only a limited fraction of the available information. While higher order statistics, such as peak counts, Minkowski functionals, and wavelet moments (Dietrich & Hartlap 2010; Shan et al.

⋆ E-mail: ji.yao@shao.ac.cn
⋆⋆ E-mail: zhangle7@mail.sysu.edu.cn
⋆⋆⋆ E-mail: hyshan@shao.ac.cn

Article published by EDP Sciences, to be cited as https://doi.org/10.1051/0004-6361/202658859

2018; Kratochvil et al. 2012; Petri et al. 2013; Cheng et al. 2020, 2025) are more sensitive to nonlinear structures, only a limited subset admit theoretical models that remain sufficiently accurate for cosmological inference in the presence of nonlinear evolution, baryonic uncertainties, and realistic observational systematics. This limitation significantly restricts their applicability in precision cosmology.

These challenges have motivated a shift toward field-level inference (FLI; Leclercq 2025), which operates directly on high-dimensional cosmic fields rather than compressed summary statistics. By working with the full fields, FLI naturally preserves non-Gaussian information and avoids reliance on approximate analytic descriptions of nonlinear structure. Instead, it leverages explicit forward modeling to simulate cosmological fields and observational effects, including survey masking, noise, and instrumental systematics, thereby providing a flexible and self-consistent framework for handling complex data models (Leclercq & Heavens 2021; Stopyra et al. 2024). However, these traditional FLI approaches, while statistically optimal in principle, face several practical limitations. Most notably, they operate in an extremely high-dimensional space, as the latent field typically contains millions of degrees of freedom, making posterior exploration computationally expensive and technically challenging. This high dimensionality places strong demands on sampling algorithms and often leads to scalability issues. In addition, traditional FLI relies on accurate and differentiable forward models that must capture nonlinear gravitational evolution and observational systematics; any mismatch or modeling inaccuracy can introduce biases in the inferred parameters.

Machine-learning (ML) techniques are increasingly applied in cosmological data processing and analysis (Dieleman et al. 2015; Petrillo et al. 2017; Schawinski et al. 2017; Zhang et al. 2024; Liu et al. 2025; Wu et al. 2025; Zhong et al. 2025). In cosmological FLI, ML can learn the complex nonlinear mapping between cosmological parameters and high-dimensional fields directly from simulations, greatly reducing the computational cost of traditional FLI (Fluri et al. 2019; Hortúa et al. 2020; Min et al. 2024). Furthermore, building on the ML-driven framework, simulation-based inference (SBI) offers a likelihood-free approach that learns the full posterior distribution of cosmological parameters from simulated observables (Cranmer et al. 2020; Lemos et al. 2023; Su et al. 2025; Zeghal et al. 2025). By leveraging the forward model, SBI naturally incorporates nonlinear evolution and observational systematics, overcoming the limitations of Gaussian-likelihood assumptions and the inefficiency of MCMC sampling in high-dimensional spaces. Consequently, ML-driven SBI has emerged as a scalable and accurate methodology for extracting cosmological information from Stage-IV WL surveys.

In WL cosmology, previous studies have primarily extracted cosmological information from the convergence field (Petri et al. 2013; Shan et al. 2018; Fluri et al. 2022; Lu et al. 2023) by analyzing its morphology, topology, and peak statistics (Liu & Haiman 2016). However, the convergence field is not directly observable and must be reconstructed from the measured shear, a process sensitive to the reconstruction algorithm and often degraded by complex survey masks. Additionally, shear measurements involve multiple calibration and correction steps that can introduce systematic uncertainties propagating into the convergence field. To mitigate these effects, this work applies FLI directly to shear fields, while also comparing results with convergence-based inference.

In shear measurements, the dominant noise arises from the intrinsic shapes of galaxies. While intrinsic alignments (IA) induce correlations, these are largely confined to galaxies within the same redshift slice. As a result, the shear signal varies smoothly with redshift, whereas shape noise exhibits stochastic, non-smooth fluctuations. To exploit this property, we perform principal component analysis (PCA) on noisy shear maps across multiple redshift slices: the smooth, coherent components capture the underlying shear signal, while the rapidly fluctuating components correspond to random shape noise. This denoising procedure requires no explicit modeling, analogous to the blind-analysis methods commonly used in 21 cm intensity mapping (Alonso et al. 2015).

In this work, we developed a shear-to-cosmology framework, enabling more direct, information-efficient, and statistically optimal use of the data, particularly in the presence of complex survey geometries. Section 2 presents the methodological framework, including the FLI network and SBI strategy. Section 3 describes the mock pipeline for generating shear fields. Section 4 details shear denoising and convergence reconstruction. Section 5 presents the inference results, comparing shear- and convergence-based estimates and demonstrating the improvements from PCA denoising. Section 6 summarizes the main findings.

2. Methodology

The hybrid inference framework, illustrated in Fig. 1, consists of two main components: (1) an FLI network (inside the dashed box), trained to map WL shear fields to cosmological parameters to generate informative feature representations; (2) an SBI module (outside the dashed box) that estimates posteriors from summary statistics, using either conventional shear two-point correlation functions (2PCFs) or ML-derived features.

2.1. Field-level inference network

Field-level inference refers to statistical methods that extract physical fields and cosmological parameters directly from observed data, with the likelihood defined over the entire discretized field and without compressing the map-level information. A neural network can be considered a realization of FLI; it can encode latent representations while still enabling a near-direct mapping from the physical fields to cosmological parameters.

2.1.1. Input data

As shown in Fig. 1, the input sample is a six-dimensional shear field, including the four sky blocks across 16 photo-z slices. Its shape is $(B, N_\gamma, N_z, N_{\rm blk}, H, W) = (B, 2, 16, 4, 64, 64)$, where B is the batch size, N_γ the shear components, N_z the photo-z slices, $N_{\rm blk}$ the input sky blocks within each sky group, and (H, W) the spatial map size. Along the redshift direction, the 16 photo-z slices are produced by the denoising procedure (see Sect. 4.1), starting from the four photo-z bins of the CSST-like lensing sample (see Sect. 3.3) and covering the $0 < z < 4$ range. From the sky-region perspective, each input sky block ($6.4 \times 6.4\,\mathrm{deg}^2$) is constructed using two data-augmentation operations applied to the original $12.8 \times 12.8\,\mathrm{deg}^2$ sky blocks with a resolution of 0.1 deg/pixel (see Fig. 5). The first operation is random cropping, which increases the spatial diversity of the training samples by a factor of 64^2. The second is random masking, where multiple circular masks with varying radii, allowed to overlap, are applied to mimic realistic survey masks (as il-

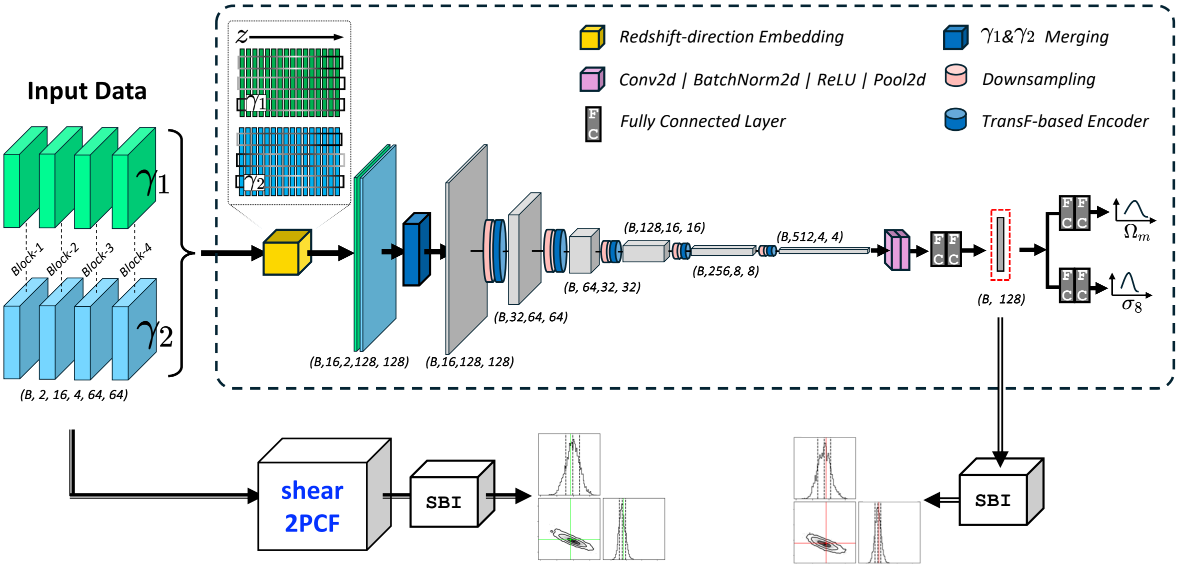


Fig. 1: Schematic overview of the proposed framework, consisting of: 1. a field-level inference (FLI) network (inside the dashed box) and 2. a simulation-based inference (SBI) module (outside the dashed box). The field-level shear input is a six-dimensional tensor with a shape of $(B, N_\gamma, N_z, N_{\rm blk}, H, W) = (B, 2, 16, 4, 64, 64)$, where B is the batch size, N_γ the shear components, N_z the number of photo-z slices, $N_{\rm blk}$ the number of blocks in each sky group (see Fig. 5), and (H, W) the spatial map size. The FLI network comprises four components: (a) a redshift-direction embedding layer, (b) a γ_1 & γ_2 merging layer, (c) transformer (`TransF`)-based encoders, and (d) fully connected inference layers to predict (Ω_m, σ_8). The gray boxes show the change of data shapes in the network pipeline. In parallel, SBI is applied either to the ML-derived feature representation (dashed red box) or to conventional two-point statistics (2PCF). The same FLI network structure was applied to a convergence map (κ) by duplicating it to form a two-component input (κ, κ).

lustrated in Fig. 2). For convergence-based inference, the corresponding convergence maps are subsequently reconstructed from the prepared shear fields using the `KS` algorithm (Sect. 4.2).

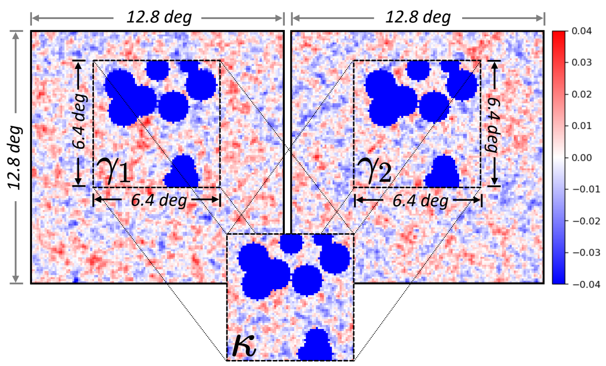


Fig. 2: Preparation of an input sky block for training the FLI network. At a resolution of 0.1 deg/pixel, based on the original 12.8×12.8 deg^2 blocks, we generated 6.4×6.4 deg^2 inputs via two augmentations: (1) random cropping (boosting spatial diversity by 64^2) and (2) random masking with overlapping circular masks to mimic survey masks. The corresponding convergence field can then be reconstructed from the prepared shear using the `KS` algorithm.

2.1.2. Network structure

The proposed FLI network (dashed box in Fig. 1) is composed of the following four key components that collaboratively extract cosmological information directly from the WL shear fields:

a) Redshift-direction embedding layer: The 3D convolutional layers with kernels elongated along the redshift axis capture the shear evolution across photo-z slices.

b) γ_1 & γ_2 merging layer: A transformer block utilizing the multi-head attention mechanism is employed to integrate the two shear components.

c) Transformer-based encoders: Unlike conventional convolutional encoders, the transformer (`TransF`)-based layer captures both local and global dependencies, naturally integrating multi-scale information across redshift slices and sky positions, thereby producing a more informative latent space for cosmological inference.

d) Inference layers: For each cosmological parameter, the inference layer consists of two fully connected layers predicting the posterior mean and log-variance. The layers are intentionally kept shallow to facilitate the subsequent SBI in decoding the latent vectors.

Together, these layers enable a direct and flexible mapping from observed shear fields to cosmological parameters.

To assess the impact of convergence reconstruction on cosmological inference, we apply the shear-based FLI network to convergence data. Since the network expects two-component inputs, we duplicate the convergence map to form a compatible two-component input (κ, κ). This setup enables a direct comparison between shear- and convergence-based inputs while keeping the network architecture and training procedure identical, isolating the effect of input type on inference performance and providing insight into the information loss in reconstructed convergence maps relative to the original shear maps.

2.1.3. Loss function

Loss function plays a central role in guiding the learning of compression statistics in neural networks and fundamentally determines the type and structure of the information extracted by the model. Alsing & Wandelt (2018) discusses theoretical approaches to optimal data compression, which, in principle, are capable of preserving the full Fisher information content of the data. However, in practical cosmological data analysis, such optimal compression schemes often require more computational resources, thereby increasing the training time of the network. Lanzieri et al. (2025) has shown that different loss function formulations implicitly constrain the types of features learned by the network. For instance, a regression-based objective using mean squared error (MSE) encourages the network to learn the conditional mean of the parameters, but does not guarantee that the learned representation is a sufficient statistic for non-Gaussian posteriors. By further introducing a regularization term, such as the Kullback–Leibler (KL) divergence, the latent representation can be constrained to approximate a prescribed prior distribution, thereby improving the structure and generalization of the latent space and mitigating overfitting to some extent.

Motivated by this, the proposed FLI network was designed to directly model the posteriors of the cosmological parameters by predicting the means, μ_i; log-variances, $\log\sigma_i^2$; and parameter samples, θ_i obtained via the reparameterization trick. We then adopted a joint loss function that combines a mean squared error (MSE) term with a KL divergence term, defined as

$$L_{\rm tot} = \sum_{i=1}^{N_\theta} \left(\underbrace{\left(\theta_i - \hat{\theta}_i\right)^2}_{L_{\rm MSE}} + \underbrace{\frac{\beta}{2}\left(\mu_i^2 + \sigma_i^2 - \log\sigma_i^2 - 1\right)}_{L_{\rm KL}} \right), \quad (1)$$

where $\hat{\theta}_i$ denotes the ground-truth cosmological parameters, and N_θ is the dimensionality of the target parameter space.

During the training process, the validation set is strictly excluded from all parameter updates. It is used solely to monitor the loss and automatically adjust the learning rate when the validation loss plateaus or worsens. Consequently, the evolution of the validation loss provides a direct indication of the model's performance and training progress. Notably, to evaluate both the shear-based and convergence-based datasets, we employed the same network architecture as mentioned in Sect. 2.1.

2.2. Simulation-based inference

Simulation-based inference (SBI) is a Bayesian framework for parameter estimation when the likelihood is intractable but forward simulations are available. It trains neural density estimators (NDE) on simulated parameter–observation pairs to approximate the posterior. The workflow typically first compresses the forward-modeling maps or summary statistics into a low-dimensional feature space using a neural network and then trains a conditional density model, often a normalizing flow, to relate these features to the target parameters. Once trained, the model provides an amortized posterior estimator applicable to any (mock) observation for efficient inference.

We used the `sbi`[1] package (Tejero-Cantero et al. 2020; Boelts et al. 2025) and adopted neural posterior estimation (NPE; Papamakarios & Murray 2016), specifically the NPE-C (SNPE-C) variant. The goal of NPE is to directly learn an approximation of the posterior distribution $p(\theta \mid x)$ from a set of simulated parameter–data pairs $(\theta,\ x)$. This is achieved by training a conditional NDE $q_\phi(\theta \mid x)$ to model the distribution of parameters conditioned on observations. The posterior is modeled with a conditional normalizing flow implemented as a neural spline flow (NSF), which provides a flexible parameterization of complex posterior distributions. The training relies solely on samples $(\theta,\ x) \sim p(\theta, x)$, which were obtained by first drawing parameters from the prior and then generating corresponding simulations through the forward model.

The network parameters, ϕ, were optimized by maximizing the conditional likelihood of the simulated parameters given the data; this is equivalent to minimizing the expected negative log-likelihood over the joint distribution. This objective can be interpreted as minimizing the expected Kullback–Leibler divergence, $D_{\rm KL}$ between the true posterior $p(\theta \mid x)$, and its neural approximation $q_\phi(\theta \mid x)$, ensuring that the learned model provides an accurate approximation of the posterior for all simulated pairs. The $D_{\rm KL}$ is non-negative and vanishes if, and only if, the two distributions are identical, which leads to the loss function

$$\mathcal{L}(\phi) = -\mathbb{E}_{p(\theta,x)}\left[\log q_\phi(\theta|x)\right]. \quad (2)$$

In this paper, from the perspective of SBI, the FLI module (Fig. 1) can be interpreted as a data-compression process that encodes the full WL shear field into low-dimensional but fully informative summary statistics. Another popular summary statistic for the shear field is the shear–shear correlations. In the following sections, we compare the constraining power of these two types of summary statistics and demonstrate the significant improvements achieved through ML-based compression.

2.2.1. Summary statistics: ML-derived feature representations

As illustrated in Fig. 1, the neural network compresses each shear map into a 128-dimensional feature representation, denoted as $\boldsymbol{x}_{\tt ML}$. The extracted features depend on both the network architecture and the training strategy. Guided by the design of the loss function, the neural compressor is trained to produce informative summary statistics for cosmological parameter inference. Although this objective does not formally guarantee sufficient statistics, it has been shown in practice to retain a substantial fraction of the cosmological information relevant for parameter estimation. For the convergence-based network, we adopted the same architecture as in the shear case but trained it independently, enabling the network to transform each convergence map into a representation of the same length (128).

2.2.2. Summary statistics: Shear 2PCFs

In the traditional statistics of cross-z 2PCFs, the shear data originally consist of 16 photo-z slices. To reduce the computational cost, adjacent slices are averaged to produce eight z-slices, resulting in 36 distinct pairs. For each pair (a, b), the correlation functions are defined as

$$\xi_+^{(a,b)}(\vartheta) = \left\langle \gamma_{\rm t}^a\, \gamma_{\rm t}^b \right\rangle(\vartheta) + \left\langle \gamma_\times^a\, \gamma_\times^b \right\rangle(\vartheta), \quad (3)$$

$$\xi_-^{(a,b)}(\vartheta) = \left\langle \gamma_{\rm t}^a\, \gamma_{\rm t}^b \right\rangle(\vartheta) - \left\langle \gamma_\times^a\, \gamma_\times^b \right\rangle(\vartheta), \quad (4)$$

where $\gamma_{\rm t}$ and $\gamma_\times$ denote the tangential and cross-components of the shear, respectively. Based on the `TreeCorr`[2] (Jarvis et al.

[1] `https://sbi-dev.github.io/sbi`

[2] `https://rmjarvis.github.io/TreeCorr`

2004; Jarvis 2015) library, the measurements were performed over angular scales in the range $\vartheta_{\rm min} = 6$ to $\vartheta_{\rm max} = 120$, sampled at $N_\vartheta = 5$ logarithmically spaced points. To combine the information from all z-slice pairs, we first stacked the ξ_+ and ξ_- measurements across different pairs,

$$\hat{\xi}_+ = \texttt{stack}(\xi_+^{p1}, \xi_+^{p2}, \cdots, \xi_+^{p36}), \quad (5)$$

$$\hat{\xi}_- = \texttt{stack}(\xi_-^{p1}, \xi_-^{p2}, \cdots, \xi_-^{p36}), \quad (6)$$

and subsequently concatenated the two to construct the summary statistics, $\hat{\xi}_{+-} = \texttt{stack}(\hat{\xi}_+, \hat{\xi}_-)$ with a length of 360. Although the traditional statistics ($\hat{\xi}_{+-}$) involve larger data volumes than ML-derived features, subsequent SBI results show that the latter capture richer cosmological information and yield tighter parameter constraints.

3. Mock pipeline

3.1. N-body simulation

To generate weak-lensing maps for training and validation, we used the publicly available CosmoGridV1[3] simulation suite (Kacprzak et al. 2023; Fluri et al. 2022), which provides a diverse set of cosmologies suitable for stage-IV lensing survey conditions. The mock catalogs were constructed to match the typical redshift distribution and galaxy number density expected in forthcoming wide-field lensing surveys (e.g., CSST, *Euclid*, and *Roman*).

CosmoGridV1 consists of 2,500 wCDM cosmologies spanning $\Omega_m, \sigma_8, w_0, n_s, \Omega_b, h_0$, assuming three degenerate massive neutrinos with a fixed total mass of $\sum m_\nu = 0.06\,$eV. For each cosmology, seven realizations with different initial conditions were produced. The full suite includes cosmologies drawn from both wide- and narrow-grid priors. In this work, we only used the "`run0`" realization of 1,260 cosmologies sampled from the narrow-grid prior spanning the parameter ranges $\Omega_m \in [0.15, 0.45]$, $\sigma_8 \in [0.5, 1.3]$, $w_0 \in [-1.25, -0.75]$, $n_s \in [0.93, 1.00]$, $\Omega_b \in [0.04, 0.05]$, and $h_0 \in [0.65, 0.75]$. Based on estimates from two-point statistics ($\partial C_\ell / \partial\theta(\ell)$, (Yao et al. 2024), the four parameters of (n_s, h_0, w_0, Ω_b) highly depend on angular sensitivity across more than two orders of angular scales, which is limited by our "6 arcmin" small-scale cut due to the simulation limit. Therefore, as a proof of methodology, we focused on (Ω_m, σ_8) to represent the cosmological constraining power in this work.

Based on the DM simulation snapshots, CosmoGridV1 also provides projected full-sky DM light-cone shells covering the redshift range $0 < z < 4$, stored in HEALPix format with resolutions of $N_{\rm side} = 512$ or 1024. Owing to computational limitations, we used the $N_{\rm side} = 512$ shells as the basis for the ray-tracing procedure to generate the weak-lensing signals in this work.

3.2. Ray tracing

Utilizing the ~68 redshift shells provided in the CosmoGridV1 data covering the redshift interval of $z \in [0, 4]$, we employed the full-sky ray-tracing algorithm `DORIAN`[4] (Ferlito et al. 2024) to generate WL shear signals at the following 23 redshifts: $z \in$ {0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90, 1.00, 1.10, 1.20, 1.30, 1.40, 1.50, 1.60, 1.70, 1.80, 2.00, 2.20, 2.40, 2.80, 3.20, 3.60}.

Notably, the simulation time steps are not fully synchronized across different cosmologies, leading to discrepancies in the redshift sampling of the DM light-cone shells, especially at high redshifts. Such inconsistencies can introduce systematic errors and spurious features into the machine-learning dataset. To address this, we adopted a coarser redshift sampling for the shear fields in the high-redshift regime, which both mitigates potential systematics and better matches the expected galaxy distributions in WL surveys, while substantially reducing computational costs.

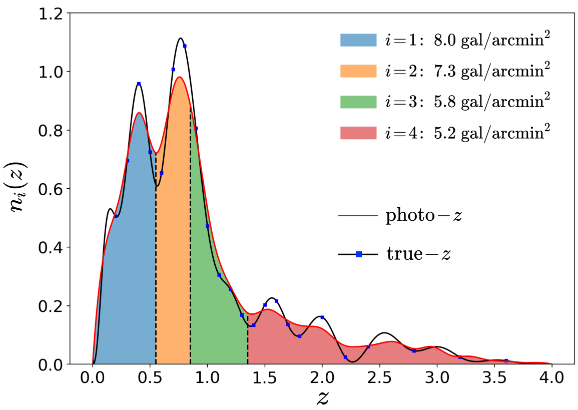


Fig. 3: CSST-like photo-z distribution. The red curve shows the photo-z distribution with photo-z uncertainties of $\sigma(z) = \sigma_z(1+z)$, $\sigma_z = 0.05$. Using this z-dependent kernel, the true-z distribution (marked curve) was obtained by deconvolution. A mock true-z galaxy catalog was generated assuming a galaxy number density of 26 gal/arcmin^2 and intrinsic shape noise of $\sigma_e = 0.288$. Incorporating photo-z uncertainties, the mock photo-z catalog was then constructed. To enhance the S/N of the WL signal, the photo-z catalog was divided into 4 z-bins, each shown in a different color with its corresponding galaxy number density.

3.3. Survey setup

To validate the hybrid framework under realistic conditions, we employed mock photo-z galaxy catalogs with specifications comparable to high-precision surveys such as CSST. These mocks include galaxy shape noise and irregular survey masking to emulate the observational conditions expected in large-scale weak-lensing programs.

Weak-lensing shear measurements rely on accurately determined galaxy ellipticities. In theory, the observed ellipticity (ε) of a galaxy can be approximated as a linear combination of its intrinsic ellipticity (ε^s) and the lensing shear (γ), given by

$$\varepsilon = \frac{\varepsilon^s + \gamma}{1 + \varepsilon^s \gamma^*}, \quad (7)$$

where $\gamma = \gamma_1 + \gamma_2\,{\rm i}$, $\varepsilon^s = \varepsilon_1^s + \varepsilon_2^s\,{\rm i}$, and " i " denotes the imaginary unit. It is assumed that the two components of ε^s follow a Gaussian distribution with standard deviation of $\sigma_e = 0.288$, which is consistent with the value adopted in the KiDS survey (Hildebrandt et al. 2020; Blake et al. 2020). Ideally, the ensemble average of galaxy ellipticities can provide an unbiased shear estimator, with $\langle\varepsilon^s\rangle = 0$ and $\gamma = \langle\varepsilon\rangle$. In practice, however, the accuracy is limited by the finite surface density of usable source galaxies: insufficient samples leave residual

[3] http://www.cosmogrid.ai
[4] https://gitlab.mpcdf.mpg.de/fferlito/dorian

contributions from intrinsic shapes, introducing galaxy shape noise into the WL signal. The third-generation surveys typically achieve galaxy number densities of approximately 10–20 gal/arcmin^2, whereas upcoming or ongoing fourth-generation surveys are expected to reach densities as high as about 30 gal/arcmin^2. China Space Station Survey Telescope (CSST) is a stage-IV space-based observatory for wide-field photometric imaging and slitless spectroscopy, targeting a sky coverage of 17,500 deg^2 (Cao et al. 2018; Gong et al. 2019; Zhan 2021). Its upcoming WL survey is expected to achieve a galaxy number density of $\sim$26 gal/arcmin^2 (Gong et al. 2019; Liu et al. 2024).

Cropping shear maps from full-sky HEALPix maps at different redshifts yields a series of pure multi-redshift shear maps. To generate the mock WL catalog, galaxies are placed at the centers of the corresponding mesh pixels, ignoring their finer spatial distribution within each pixel. Based on the photometric-redshift (photo-z) distribution of CSST lensing samples (Yao et al. 2024), a CSST-like catalog can then be generated through the steps listed below.

(1) We adopted the photo-z distribution of CSST lensing galaxies (red curve in Fig. 3), assuming a photo-z uncertainty parameterized as $\sigma(z) = \sigma_z(1 + z)$, with $\sigma_z = 0.05$ (Cao et al. 2018). The corresponding true-z distribution is then recovered through a deconvolution procedure.

(2) Using the recovered true-z distribution (black curve in Fig. 3) and an assumed galaxy number density of 26 gal/arcmin^2, we generated a mock sample of galaxies with redshifts drawn accordingly.

(3) Finally, by assigning shear signals from simulated shear fields at multiple redshifts (blue squares in Fig. 3) and incorporating intrinsic shape noise with a dispersion of $\sigma_e = 0.288$, we constructed the CSST-like photo-z galaxy catalog.

As a preliminary investigation within the shear-to-cosmology framework, this work focuses solely on extracting cosmological information in the presence of galaxy shape noise and photo-z uncertainties, deferring the inclusion of PSF effects, shear measurement biases, mean redshift biases, intrinsic alignments (IAs), baryonic feedback, reduced shear, source clustering, magnification, and other systematics to future studies using larger simulation sets.

4. Data preprocessing

The goal of employing FLI is to maximally extract the non-Gaussian information encoded in the data. However, conventional smoothing operations inevitably suppress such features, thereby motivating the development of denoising techniques tailored for FLI. These methods aim to reduce noise while preserving the essential non-Gaussian structures of the field.

From an information-theoretic perspective, denoising does not introduce additional cosmological information into the data. Instead, it facilitates more efficient information extraction by improving the pixel-level signal-to-noise (S/N) ratio. In practice, ML models operate under a trade-off between effective information extraction and the control of overfitting, where limited data and finite model capacity can hinder the full utilization of the available information. By enhancing the S/N ratio, denoised data enable models to more reliably capture the underlying cosmological signal, thereby improving the fidelity and constraining power of field-level inference.

Several ML-based denoising approaches, such as GAN-based models (Shirasaki et al. 2021) and diffusion models (Aoyama et al. 2026), have been applied to WL mass maps. In contrast, leveraging the distinct statistical properties of WL signals and shape noise, we aim to develop a blind, training-free denoising technique that does not rely on any learned model. This approach constitutes a key methodological component of this work.

4.1. Shear denoising

Weak-lensing signals are strongly redshift-dependent, whereas shape noise is largely uncorrelated with redshift. This motivated the use of a PCA algorithm along the redshift direction to separate signal from noise. However, the global nature of PCA causes high S/N modes from high-redshift bins to leak into lower-redshift bins when wide redshift bins are used, diluting the true redshift evolution. To mitigate this, we performed a PCA independently within narrower redshift intervals, limiting cross-bin signal mixing. This requires subdividing galaxies within each photo-z bin, enabling the denoising procedure to better preserve the redshift-dependent structure of the shear signal. Hence, the entire mock galaxy catalog over the entire photo-z range is divided into four photo-z bins, as shown in Fig. 3. Considering both the photo-z distribution of CSST galaxies and the characteristic decline of WL strength at lower redshifts, the number of galaxies in the low photo-z bins is appropriately increased to boost the WL signal in those bins.

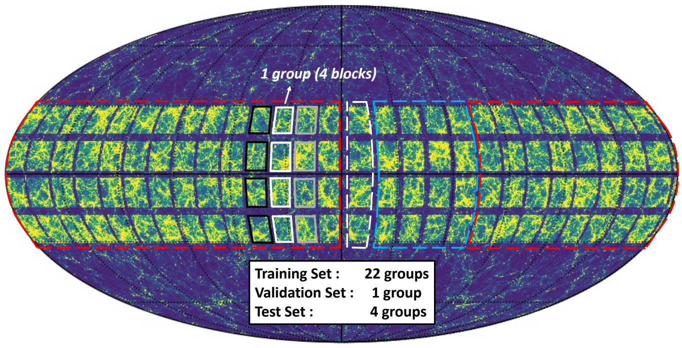
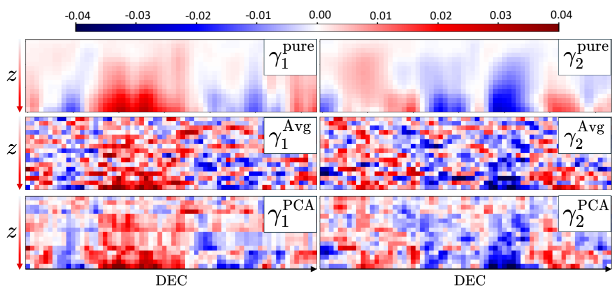


Fig. 5: Illustration of sky-region partitioning into training, validation, and test sets. For each cosmology, the dataset contains 108 sky blocks, grouped in fours within the same declination region. Groups enclosed by red, blue, and white dashed boxes indicate the training, test, and validation sets, respectively, with no overlapping sky regions. Each block is a 128×128 mesh spanning 12.8×12.8 deg^2 and is randomly cropped to 6.4×6.4 deg^2 before input to the neural network. Each group of cropped blocks forms a sample, covering 163.84 deg^2.



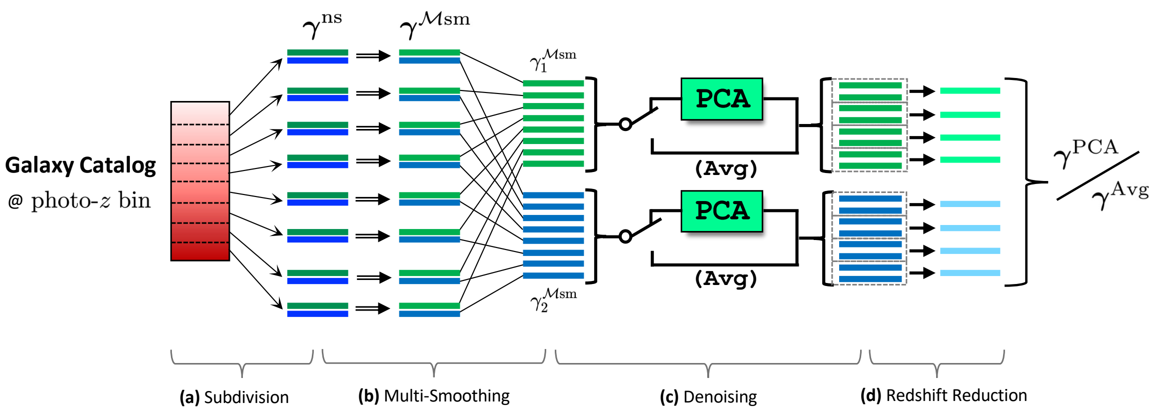


Fig. 4: Pipeline of shear measurement in each photo-z bin including four steps to enhance the shear signal quality. When PCA processing is applied, the denoised shear field is denoted as $\boldsymbol{\gamma}^{\rm PCA}$; otherwise, the resulting field is denoted as $\boldsymbol{\gamma}^{\rm Avg}$.

Fig. 6: Visualization of shear denoising. The bottom panels show shear fields from the pipeline in Fig. 4 including PCA denoising ($\boldsymbol{\gamma}^{\rm PCA}$), while the middle panels show shear fields from the pipeline without PCA ($\boldsymbol{\gamma}^{\rm Avg}$). The top panels display the corresponding pure shear signals after the same multi-smoothing operation ($\boldsymbol{\gamma}^{\rm pure}$). In each panel, the horizontal axis corresponds to declination and the vertical axis to redshift. As shown in the figure, PCA denoising clearly preserves more reliable redshift structures, which provide the key information for our inference framework.

For each sky block in Fig. 5, the shear data can be obtained through the following steps (illustrated in Fig. 4), incorporating analytical operations to suppress noise and enhance the shear signal quality.

(a) Subdivision: To ensure statistical stability, each subset must contain a sufficient number of galaxies (typically 20–40 per pixel), given the finite sample size within each photo-z bin. At the same time, the number of subsets determines the number of effective PCA modes; using too few subsets reduces the decomposition capacity and limits the ability to separate signal from noise. We therefore adopted eight subsets as a pragmatic compromise between these competing requirements. Accordingly, within each photo-z bin and sky block, we divided galaxies into eight subsets by their photo-z values and then computed noisy shear maps ($\boldsymbol{\gamma}^{\rm ns}$) by averaging the ellipticities per subset, expressed as $\boldsymbol{\gamma}^{\rm ns} = \langle \boldsymbol{\epsilon} \rangle$. Notably, we adopt the simplified precondition here that all galaxies reside entirely within pixel boundaries, neglecting the complexities associated with those intersecting pixel edges.

(b) Multi-smoothing: To improve PCA decomposition by enhancing the S/N ratio via high-frequency noise suppression, we smooth the noisy shear map in each subset using a set of Gaussian kernels,

$$\begin{aligned}\boldsymbol{\gamma}^{\rm sm}(\boldsymbol{\theta}) &= [\boldsymbol{\gamma}^{\rm ns} \otimes \mathcal{G}]\,(\boldsymbol{\theta}) \\ &\rightleftharpoons\ \ \tilde{\boldsymbol{\gamma}}^{\rm sm}(\boldsymbol{\ell}) = \tilde{\mathcal{G}}(\boldsymbol{\ell}) \cdot \tilde{\boldsymbol{\gamma}}^{\rm ns}(\boldsymbol{\ell}),\end{aligned} \tag{8}$$

where the Gaussian kernel in Fourier space is given by

$$\tilde{\mathcal{G}}(\boldsymbol{\ell}) := \exp(-\frac{|\boldsymbol{\ell}|^2}{2\tilde{\sigma}_{\rm sm}^2}). \tag{9}$$

We adopted a set of smoothing scales in Fourier space, $\tilde{\sigma}_{\rm sm}$ = [10, 20, 40] pixel^{-1}, which correspond to Gaussian kernel widths in real space of $\sigma_{\rm sm}$ = [1.00, 0.50, 0.25] deg, respectively. Intuitively, feeding multiple independently smoothed results into a neural network preserves multi-scale, non-Gaussian features. Although increased smoothing improves the S/N ratio, it reduces local complexity, making overfitting more likely. Averaging multiple smoothed results may offer a balance between information retention and overfitting prevention, expressed as

$$\boldsymbol{\gamma}^{\mathcal{M}\rm sm}(\boldsymbol{\theta}) = \frac{1}{N_{\rm sm}} \sum_{n=1}^{N_{\rm sm}} \boldsymbol{\gamma}^{\rm sm}(\boldsymbol{\theta}). \tag{10}$$

(c) Denoising: Conventional denoising algorithms (such as smoothing and Wiener-filter) normally down-weight the high-ℓ modes at map level, while ignoring the fact that shear signals are correlated along the redshift direction and shape noises are not. Based on the resulting $\boldsymbol{\gamma}^{\mathcal{M}\rm sm}$ from eight subsets, one can further suppress the noise level via the denoising algorithm PCA:

$$\begin{bmatrix} \boldsymbol{\gamma}^{\rm PCA}_{(1)} \\ \boldsymbol{\gamma}^{\rm PCA}_{(2)} \\ \cdots \\ \boldsymbol{\gamma}^{\rm PCA}_{(8)} \end{bmatrix} = \mathtt{PCA}\left(\begin{bmatrix} \boldsymbol{\gamma}^{\mathcal{M}\rm sm}_{(1)} \\ \boldsymbol{\gamma}^{\mathcal{M}\rm sm}_{(2)} \\ \cdots \\ \boldsymbol{\gamma}^{\mathcal{M}\rm sm}_{(8)} \end{bmatrix}\right). \tag{11}$$

Based on the PCA decomposition, the input shear data were separated into eight orthogonal modes. We only retained the first two smooth principal components for shear reconstruction, since the true shear field varies smoothly with redshift, whereas the shape noise manifests as random fluctuations.

(d) Redshift reduction: Combining the outputs from the four photo-z bins yields a four-dimensional shear dataset ($N_\gamma = 2$, $N_z = 32$, H, W), where N_γ and N_z denote the two shear components and the redshift dimension after denoising, respectively. To reduce its data size for ML or statistical analyses, every two consecutive redshift layers are averaged, resulting in ($N_\gamma = 2$, $N_z = 16$, H, W). The processed data are denoted as $\boldsymbol{\gamma}^{\rm PCA}$ when PCA is applied and $\boldsymbol{\gamma}^{\rm Avg}$ otherwise.

At present, a systematic sensitivity analysis of these design choices (e.g., the number of subsets and the smoothing scale) is still lacking. Nevertheless, empirical results suggest that the method remains robust to moderate variations, provided the parameters stay within physically and statistically motivated ranges. In practice, the number of subsets is governed by a trade-off between ensuring sufficient galaxy counts per subset and retaining enough degrees of freedom for effective PCA decomposition, while the smoothing scale reflects a balance between noise suppression and the preservation of small-scale structures. As long as these considerations are satisfied, we do not expect qualitative changes in the overall performance.

It is worth noting that we used a certain strategy: the mask effect is applied to the prepared shear maps ($\boldsymbol{\gamma}^{\rm PCA}$ or $\boldsymbol{\gamma}^{\rm Avg}$). Although this treatment does not fully correspond to the actual mask, it can significantly accelerate the training of the neural network.

4.2. Convergence mapping

Based on the prepared shear field ($\boldsymbol{\gamma}^{\rm PCA}$ or $\boldsymbol{\gamma}^{\rm Avg}$), generated with random masks during training, we applied the KS algorithm (Kaiser & Squires 1993) to reconstruct the convergence field (κ), retaining only its E-mode component:

$$\kappa^{\rm prep}(\boldsymbol{\theta}) \leftarrow \texttt{KS(}\,\boldsymbol{\gamma}^{\rm prep}(\boldsymbol{\theta})\,\texttt{):}\quad \tilde{\kappa}^{\rm prep}(\boldsymbol{\ell}) = D^{*}(\boldsymbol{\ell})\,\tilde{\boldsymbol{\gamma}}^{\rm prep}(\boldsymbol{\ell}), \qquad (12)$$
$$D(\boldsymbol{\ell}) = \frac{\ell_1^2 - \ell_2^2 + 2\,\mathrm{i}\,\ell_1\ell_2}{|\boldsymbol{\ell}|^2},$$

where $D(\boldsymbol{\ell})$ is the projection operator, " i " denotes the imaginary unit, and the option "prep" can be specified as "PCA" or "Avg".

5. Results

5.1. Loss evolution

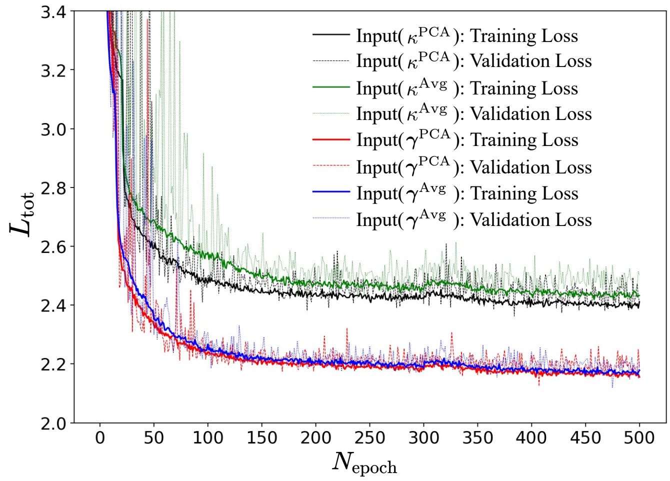


Fig. 7: Evolution of loss (`batch size=32`) for different input data: the shear maps $\boldsymbol{\gamma}^{\rm PCA}$ (with denoising) and $\boldsymbol{\gamma}^{\rm Avg}$ (without denoising), the corresponding convergence maps $\kappa^{\rm PCA}$ and $\kappa^{\rm Avg}$ via KS algorithm. In neural-network training, the validation set was excluded from parameter optimization and serves solely to trigger automatic learning rate reduction based on variations in its loss. Consequently, the evolution of validation loss reflects the model's performance improvement.

As illustrated in Fig. 6, PCA denoising effectively enhances the S/N of the shear data and further reduces the training loss. To evaluate its impact, we compared the FLI results with and without PCA denoising. Based on the loss function described in Sect. 2.1.3, Fig. 7 presents the training loss evolution with a batch size of 32 for four types of input data. The first two are convergence fields—the averaged convergence ($\kappa^{\rm Avg}$) and the PCA-denoised convergence ($\kappa^{\rm PCA}$)—both reconstructed from their corresponding shear fields using the KS algorithm. The other two are shear fields: the averaged shear ($\gamma^{\rm Avg}$) and the PCA-denoised shear ($\gamma^{\rm PCA}$). As shown in Fig. 7, the shear-based approaches achieve substantially lower losses than the convergence-based ones, since convergence reconstruction introduces additional systematic errors, particularly in the presence of sky masking.

5.2. Predictions by FLI

Under the assumption of Gaussian distribution, the trained FLI network can output the posteriors of the target parameters (Ω_m, σ_8) directly. As shown in Fig. 5, the prepared test set includes four discrete sky groups per cosmology, covering a total of 655.36 deg^2. In order to highlight the differences between the input data types: ($\boldsymbol{\gamma}^{\rm Avg}$, $\boldsymbol{\gamma}^{\rm PCA}$, $\kappa^{\rm Avg}$, $\kappa^{\rm PCA}$), we averaged the FLI predictions from the four sky groups for each cosmology in Figs. 8 and 9. As shown in the figure, qualitatively, shear-based inference outperforms convergence-based inference. Furthermore, applying PCA denoising reduces the systematic bias of the predictions relative to the true values.

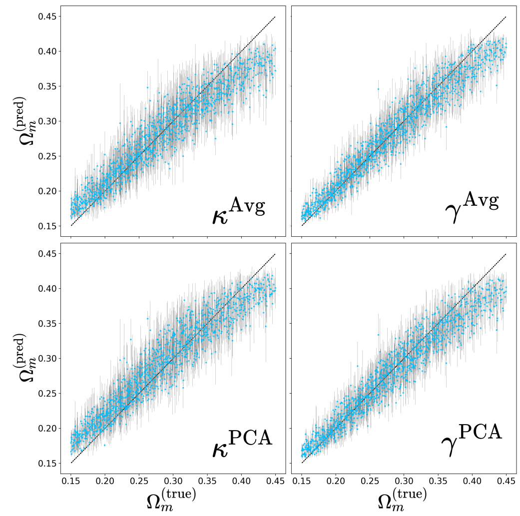


Fig. 8: Predictions of Ω_m directly from networks trained and tested on four types of input data: ($\boldsymbol{\gamma}^{\rm Avg}$, $\boldsymbol{\gamma}^{\rm PCA}$, $\kappa^{\rm Avg}$, $\kappa^{\rm PCA}$). For both shear- and convergence-based inputs, the networks share the same architecture (Fig. 1) but were trained independently. Each point, evaluated on the test set, denotes the average prediction over four sky groups (covering 655.36 deg^2) within the same cosmology, with error bars indicating the standard deviation. Dashed lines mark the true parameter values.

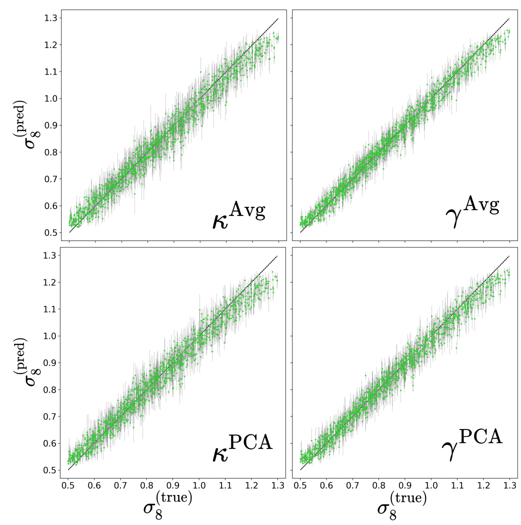


Fig. 9: Predictions of σ_8 directly from networks trained and tested on four types of input data: ($\gamma^{\rm Avg}$, $\gamma^{\rm PCA}$, $\kappa^{\rm Avg}$, $\kappa^{\rm PCA}$). Details are as in Fig. 8.

5.3. Validation of predictions

The probability integral transform (PIT) tool is a standard diagnostic method for evaluating the reliability of model-predicted probability distributions. When a model's cumulative distribution function (CDF) is well calibrated, the PIT values obtained by evaluating observed data against the CDF follow a uniform distribution. PIT thus offers a simple yet effective way to assess both the calibration and uncertainty of probabilistic predictions.

In this work, cosmological parameters are inferred using two approaches: (1) directly from the FLI network, and (2) via SBI applied to summary statistics. We performed PIT analyses on the posteriors obtained from both methods to evaluate their inference quality and provide an additional consistency check for our framework.

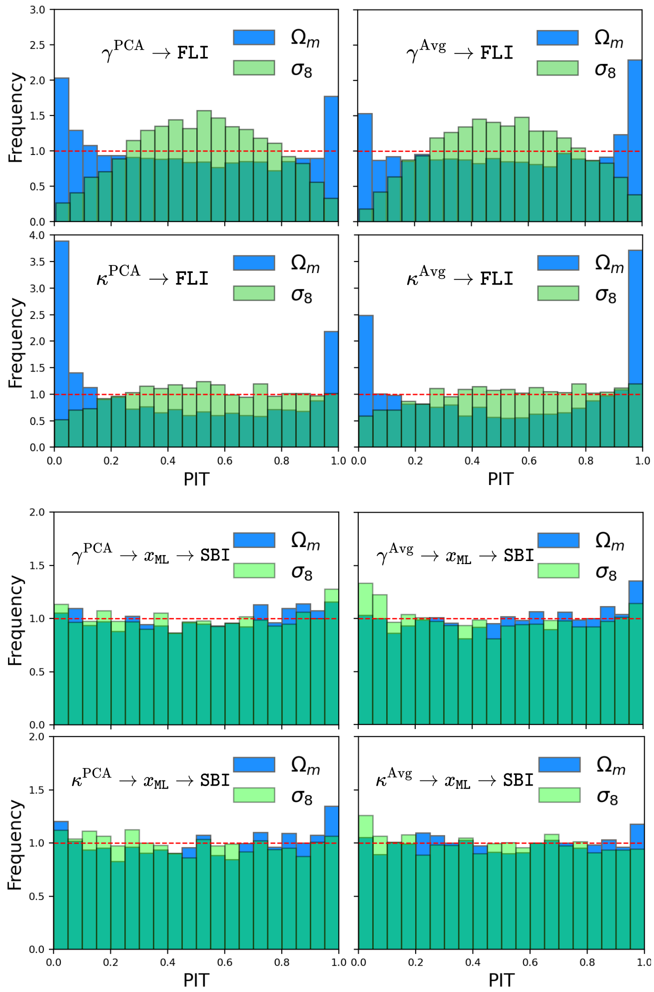


Fig. 10: Comparison of PIT validation results between field-level inference (FLI) and ML-derived simulation-based inference (SBI), performed using four types of input data: ($\gamma^{\rm PCA}$, $\gamma^{\rm Avg}$, $\kappa^{\rm PCA}$, $\kappa^{\rm Avg}$). Compared with FLI (top panels), the PIT distribution of SBI (bottom panels) is closer to a uniform distribution, indicating that the SBI posteriors are better calibrated and more consistent with the true parameters.

As shown in Fig. 10, the PIT distributions from SBI are noticeably closer to uniform than those from the FLI network, indicating better calibrated posteriors. This is expected because, although both methods use the same ML-derived feature representations, the inference module in the FLI network, comprising fully connected layers (Fig. 1), assumes Gaussian posteriors, which limits flexibility and accuracy. In contrast, SBI directly learns the mapping from features to parameter posteriors using simulated data, enabling it to capture complex, non-Gaussian, and even multi-modal distributions. By explicitly modeling the full posterior, SBI naturally accounts for parameter correlations and uncertainties, yielding more reliable and accurate inference than the Gaussian-constrained network outputs.

5.4. Cosmological outputs

SBI offers a flexible and statistically rigorous framework for cosmological parameter estimation from complex data. In this study, we employed two distinct forms of summary statistics: (1) ML-derived feature representation ($x_{\rm ML}$) and (2) the shear 2PCF ($\hat{\xi}_{+-}$). The constraining power on the target cosmological parameters is quantified by the figure of merit (FoM), normalized per square degree.

In principle, given a fixed neural compressor, stochastic variations across independent SBI realizations should be negligible if the neural density estimator (NDE) is fully converged. In practice, however, factors such as finite training data, limited model capacity, and random initialization of the NDE can introduce small variations in the inferred posteriors.

Each SBI run yields a distinct approximate posterior conditioned on its learned parameters. This reflects the stochastic nature of an NDE, including random initialization and stochastic optimization during training, which introduces variability across realizations. To assess the impact of these effects, we performed ten independent SBI realizations using a fixed trained FLI network as the neural compressor. We find that the inferred posteriors are broadly consistent across runs, indicating good stability of the SBI procedure. Based on this consistency, we report results from a representative NDE model that achieves strong constraining power, in order to illustrate the achievable performance of the method. The full set of inferred posteriors from the ten SBI runs is presented in Appendix A.

To demonstrate the SBI performance in each case, we adopted the fiducial cosmology characterized by ($\Omega_m^{\rm true}$, $\sigma_8^{\rm true}$) = (0.3062, 0.7615). Figure 11 presents the posterior distributions inferred from three SBI schemes, all applied to PCA-denoised weak lensing fields. The left panel shows the results based on 2PCF summaries derived from the PCA-denoised shear field ($\gamma^{\rm PCA}$), whereas the middle and right panels correspond to ML-derived features extracted from the PCA-denoised convergence ($\kappa^{\rm PCA}$) and shear ($\gamma^{\rm PCA}$), respectively. A comparison between the left and right panels reveals that ML-derived features enhance the parameter constraining power by 29.3% relative to the 2PCF-based summaries. In general, ML-powered compressors are expected to provide more informative summary statistics. However, the comparison between the left and middle panels shows that the shear-based 2PCF achieves stronger constraining power than the convergence-based ML-derived features. This is mainly due to the limited accuracy of convergence-field reconstruction using the Kaiser–Squires inversion. For our adopted setup, consisting of a 6.4° × 6.4° field of view with a 6 arcmin resolution, the reconstructed convergence maps deviate noticeably from the true field, indicating non-negligible reconstruction errors arising from information loss and systematics in the shear-to-convergence inversion, including finite-field effects, boundary artifacts, and pixelization. In contrast, shear is

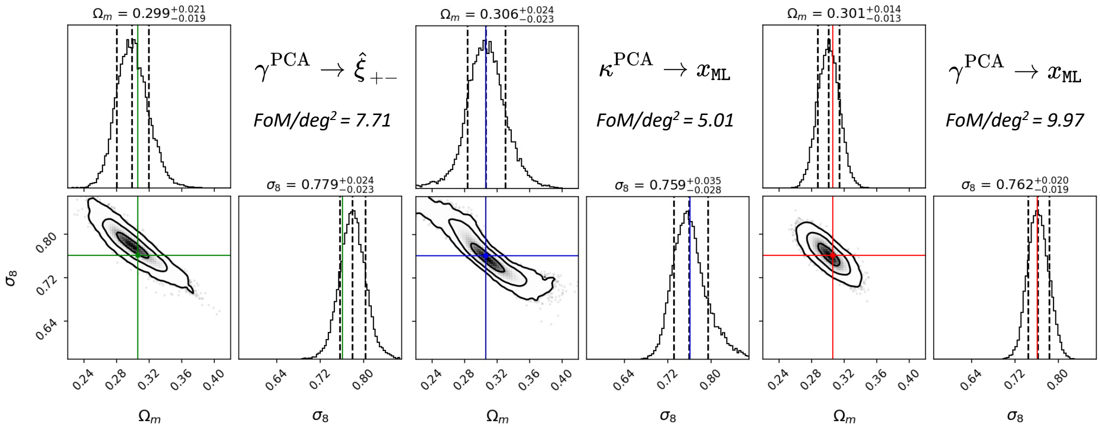


Fig. 11: Posteriors inferred via SBI using three types of summary statistics over the sky area of 655.4 deg^2, with true cosmological parameters $\Omega_{\rm m}^{\rm true} = 0.3062$ and $\sigma_8^{\rm true} = 0.7615$. The left panel shows the posterior using the 2PCF ($\hat{\xi}_{+-}$) of the PCA-denoised shear field ($\gamma^{\rm PCA}$). The middle panel uses ML-derived features from the convergence field ($\kappa^{\rm PCA}$), reconstructed from $\gamma^{\rm PCA}$ via the KS method. The right panel uses ML-derived features directly extracted from $\gamma^{\rm PCA}$. Based on the values of FoM/deg^2, ML-derived features improve the constraining power by 29.3% over the 2PCFs, and the shear-based method outperforms the kappa-based method by 99.0%.

the directly observed quantity and preserves the original lensing information more faithfully. Therefore, performing inference directly in the shear domain can better retain cosmological information and avoid reconstruction-induced biases. Consistently, the shear-based ML representation outperforms its convergence-based counterpart by 99.0%, further highlighting the advantage of direct shear-field analysis for cosmological parameter inference.

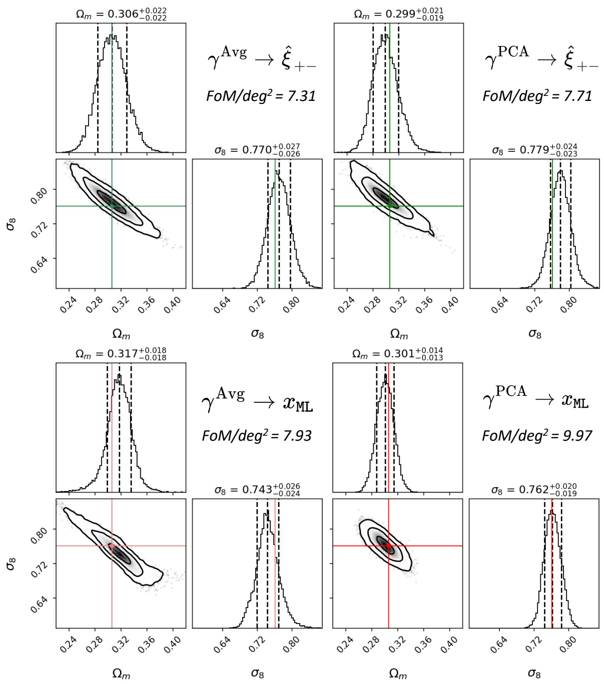


Fig. 12: Comparison of SBI posteriors obtained from $\gamma^{\rm Avg}$ and $\gamma^{\rm PCA}$ over 655.4 deg^2, highlighting the impact of PCA denoising on two types of summary statistics: the 2PCFs ($\hat{\xi}_{+-}$; top panels) and ML-derived features ($x_{\tt ML}$; bottom panels). Based on FoM/deg^2, PCA denoising improves the constraining power by 25.7% when using ML-derived features.

Table 1: Cosmological constraints from SBI using a CSST-like WL survey covering 655.4 deg^2, with true parameters $(\Omega_{\rm m}^{\rm true}, \sigma_8^{\rm true}) = (0.3062, 0.7615)$.

Data	Denoising		Sum.	Ω_m	σ_8	FoM
	$\mathcal{M}_{\rm sm}$	PCA	Stat.			/deg^2
$\gamma^{\rm Avg}$	✓	–	2PCF	$0.306^{+0.022}_{-0.022}$	$0.770^{+0.027}_{-0.026}$	7.31
			$x_{\tt ML}$	$0.317^{+0.018}_{-0.018}$	$0.743^{+0.026}_{-0.024}$	7.93
$\gamma^{\rm PCA}$	✓	✓	2PCF	$0.302^{+0.021}_{-0.020}$	$0.775^{+0.028}_{-0.028}$	7.71
			$x_{\tt ML}$	$0.301^{+0.014}_{-0.013}$	$0.762^{+0.020}_{-0.019}$	**9.97**
$\kappa^{\rm Avg}$	✓	–	$x_{\tt ML}$	$0.332^{+0.027}_{-0.025}$	$0.717^{+0.028}_{-0.028}$	4.99
$\kappa^{\rm PCA}$	✓	✓	$x_{\tt ML}$	$0.306^{+0.024}_{-0.023}$	$0.759^{+0.035}_{-0.028}$	5.01

Notes. The first column lists the input data types, which correspond to the different denoising operations shown in the second column. The third column indicates the type of summary statistics. 2PCF is the two-point correlation function, while $x_{\tt ML}$ are the ML-derived features. The final column (FoM/deg^2) quantifies the constraining power for the joint inference of (Ω_m, σ_8).

To further assess the impact of PCA denoising on shear-field-based posterior inference, we compare the results of 2PCF-based and ML-derived SBI obtained from $\gamma^{\rm Avg}$ and $\gamma^{\rm PCA}$, respectively (Fig. 12). As shown in the left panels, the conventional 2PCF-based SBI shows little improvement from PCA denoising, as the process primarily enhances non-Gaussian components that are not effectively captured by the 2PCF method. In contrast, the ML-derived SBI (right panels) benefits substantially from PCA denoising, achieving a 25.7% increase in constraining power and demonstrating its effectiveness in strengthening parameter inference. Overall, compared with the traditional 2PCF method, the combination of PCA denoising and ML-based feature extraction further boosts the constraining power by 36.4%, highlighting the advantage of integrating these techniques for improved cosmological inference. The summarizing of the cosmological constraints is listed in Table 1.

6. Conclusions

We developed a ML-powered framework for shear-to-cosmology inference, motivated by the complementary strengths of simulation-based inference (SBI) and field-level inference (FLI). SBI offers a simple yet powerful approach for posterior estimation, capable of capturing complex, non-Gaussian, and even multi-modal parameter distributions. FLI networks can automatically extract informative and compact representations from cosmological fields. By integrating these two techniques, we establish an effective and robust shear-to-cosmology pipeline that harnesses the advantages of both SBI and FLI. This framework operates directly on the shear fields without requiring convergence reconstruction and comprises three main components: (1) a shear denoising technique to mitigate the noise from intrinsic galaxy shapes, (2) a transformer-based network for FLI, and (3) an SBI module.

Theoretically, the convergence and shear fields contain equivalent cosmological information. In practice, however, the shear-to-convergence inversion amplifies noise and introduces filtering artifacts, resulting in non-negligible information loss. A shear-based analysis therefore offers a cleaner alternative. To assess its feasibility, we conducted preliminary experiments (not shown here) demonstrating that a transformer-based U-Net can accurately reconstruct the convergence field from high-S/N shear data, indicating that the network is capable of learning the underlying reconstruction operator. Consequently, even with only shear inputs, the model can, in principle, recover information comparable to that in the true underlying convergence field. Consistently, SBI posteriors show that shear-based analysis nearly doubles the constraining power relative to the KS-convergence field, confirming that the FLI network does extract more cosmological information directly from shear.

In order to capture information encoded in cosmic evolution, the input to our framework is constructed from multi-redshift shear fields. In practice, increasing the number of redshift bins reduces the galaxy number density within each bin, thereby amplifying shape noise and making denoising an essential step in multi-redshift WL analyses. In this work, for shear-based analyses, we instead adopted PCA denoising along the redshift direction, which is a blind and training-free technique that suppresses noise while preserving the redshift evolution of the shear signal. To assess its impact, we evaluate the effect of PCA denoising using two types of summary statistics within the SBI framework: the 2PCFs and ML-derived features. For the 2PCF-based SBI, the improvement brought by PCA denoising is negligible. In contrast, when using ML-derived features, PCA denoising results in a 25.7% increase in constraining power. This demonstrates that the PCA-denoised shear fields retain richer cosmological information, predominantly non-Gaussian components, which are effectively exploited by the ML-derived features but not by the traditional 2PCF analysis.

Several aspects of this work warrant further in-depth investigation. WL analyses have traditionally relied on the convergence field, reconstructed using a variety of methods, including the Kaiser–Squires algorithm (Kaiser & Squires 1993), Bayesian forward modeling (Alsing et al. 2016), Wiener filtering (Jeffrey et al. 2018), wavelet-based techniques (Starck et al. 2006; Gatti et al. 2026), prior-free maximum-likelihood approaches (Shi et al. 2024, 2025), and machine-learning methods (Jeffrey et al. 2020; Shirasaki & Ikeda 2024). In this work, we adopted the standard KS algorithm for convergence reconstruction; however, a systematic comparison between shear-based inference and convergence-based results derived from alternative reconstruction techniques is still lacking and should be pursued in future studies.

In addition, the impact of the denoising procedure has not yet been fully quantified. Our PCA-based approach, while simple and training-free, depends on several design choices (e.g., the number of subsets and the smoothing scale) and does not guarantee optimal separation between signal and noise. Moreover, as a linear decomposition method, PCA may not fully capture non-Gaussian or highly nonlinear features in the shear field. These limitations suggest that the current denoising strategy can be further improved. For example, wavelet-based methods may provide a more natural separation of structures across different scales, while more advanced techniques such as ICA (Hyvärinen & Oja 2000) and GMCA (Bobin et al. 2007) offer promising alternatives for blind source separation. A systematic exploration of these approaches, as well as their impact on cosmological inference, remains an important direction for future work. More generally, combining map-level denoising with redshift-direction denoising could further enhance performance, although such a strategy would require substantially greater computational resources and will therefore be considered as part of a future investigation.

Finally, due to computational limitations, both the angular resolution and the effective sky area of our input data are significantly lower than those expected for stage-IV WL surveys. In addition, several important observational systematics are not included in this preliminary study, such as PSF effects, intrinsic alignments, source clustering, and shear-measurement biases. Addressing these limitations will require the construction of more realistic mock datasets, incorporating spectroscopic and photometric galaxy catalogs together with corresponding weak-lensing fields, based on high-resolution N-body simulations, partial-sky ray tracing, and realistic image simulations.

In conclusion, we present a unified machine-learning framework that integrates FLI and SBI for direct shear-to-cosmology mapping. By combining PCA denoising, a transformer-driven feature extractor, and SBI posterior estimation, our approach enables accurate and efficient cosmological inference directly from shear fields. The results demonstrate that PCA denoising enhances the recovery of non-Gaussian information, while the transformer-based network effectively captures cosmological signals beyond the reach of shear two-point analyses. These findings highlight the potential of ML-powered hybrid FLI–SBI frameworks to fully exploit the statistical power of WL data, paving the way for more precise and information-rich cosmological parameter estimation in current and future surveys.

Acknowledgements. We acknowledge the support by National Key R&D Program of China (2022YFF0503403), National Natural Science Foundation of China (12533008, 12503010, 12203084, 12573006, 12473097), China Manned Space Program (CMS-CSST-2021-(A02, A03, B01), CMS-CSST-2025-(A03, A05)), Postdoctoral Fellowship Program of CPSF (GZC20252089), Guangdong Basic and Applied Basic Research Foundation (2024A1515012309). JCD would like to thank "Tree New Bee" club for supporting friendly discussion environment.

Software: TreeCorr (Jarvis et al. 2004; Jarvis 2015), DORIAN (Ferlito et al. 2024), PyTorch (Paszke et al. 2019), sbi (Tejero-Cantero et al. 2020), HEALPix/healpy (Górski et al. 2005; Zonca et al. 2019), NumPy (Van Der Walt et al. 2011; Harris et al. 2020), OpenMPI (Gabriel et al. 2004), mpi4py (Dalcin & Fang 2021), SciPy (Virtanen et al. 2020), Astropy (Astropy Collaboration et al. 2013, 2018, 2022), h5py (Collette 2013), Matplotlib (Hunter 2007), corner (Foreman-Mackey 2016)

References

Aihara, H., Arimoto, N., Armstrong, R., et al. 2018, PASJ, 70, S4
Alonso, D., Bull, P., Ferreira, P. G., & Santos, M. G. 2015, MNRAS, 447, 400

Alsing, J., Heavens, A., Jaffe, A. H., et al. 2016, MNRAS, 455, 4452
Alsing, J. & Wandelt, B. 2018, MNRAS, 476, L60
Aoyama, S. D., Osato, K., & Shirasaki, M. 2026, PASJ, psag052
Astropy Collaboration, Price-Whelan, A. M., Lim, P. L., Earl, N., et al. 2022, AJ, 935, 167
Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B., Günther, H., et al. 2018, AJ, 156, 123
Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., Greenfield, P., et al. 2013, A&A, 558, A33
Benjamin, J., Van Waerbeke, L., Heymans, C., et al. 2013, MNRAS, 431, 1547
Bernstein, G. M. 2009, ApJ, 695, 652
Blake, C., Amon, A., Asgari, M., et al. 2020, A&A, 642, A158
Bobin, J., Starck, J.-L., Fadili, J., & Moudden, Y. 2007, IEEE Transactions on Image Processing, 16, 2662
Boelts, J., Deistler, M., Gloeckler, M., et al. 2025, JOSS, 10, 7754
Cao, Y., Gong, Y., Meng, X.-M., et al. 2018, MNRAS, 480, 2178
Cheng, S., Marques, G. A., Grandón, D., et al. 2025, J. Cosmology Astropart. Phys., 2025, 006
Cheng, S., Ting, Y.-S., Ménard, B., & Bruna, J. 2020, MNRAS, 499, 5902
Collette, A. 2013, Python and HDF5 (OŔeilly)
Cranmer, K., Brehmer, J., & Louppe, G. 2020, PNAS, 117, 30055
CSST Collaboration, Gong, Y., Miao, H., et al. 2026, Sci. China-Phys. Mech. Astron., 69, 239501
Dalcin, L. & Fang, Y.-L. L. 2021, Comput. Sci. Eng., 23, 47
Dark Energy Survey Collaboration, Abbott, T., Abdalla, F. B., et al. 2016, MNRAS, 460, 1270
Dieleman, S., Willett, K. W., & Dambre, J. 2015, MNRAS, 450, 1441
Dietrich, J. P. & Hartlap, J. 2010, MNRAS, 402, 1049
Euclid Collaboration, Mellier, Y., Abdurro'uf, et al. 2025, A&A, 697, A1
Ferlito, F., Davies, C. T., Springel, V., et al. 2024, MNRAS, 533, 3209
Fluri, J., Kacprzak, T., Lucchi, A., et al. 2019, Phys. Rev. D, 100, 063514
Fluri, J., Kacprzak, T., Lucchi, A., et al. 2022, Phys. Rev. D, 105, 083518
Foreman-Mackey, D. 2016, JOSS, 1, 24
Gabriel, E., Fagg, G. E., Bosilca, G., et al. 2004, in Recent Advances in Parallel Virtual Machine and Message Passing Interface: 11th European PVM/MPI Users' Group Meeting Budapest, Hungary, September 19–22, 2004. Proceedings 11 (Springer), 97–104
Gatti, M., Anbajagane, D., Chang, C., et al. 2026, Phys. Rev. D, 113, L081303
Gong, Y., Liu, X., Cao, Y., et al. 2019, ApJ, 883, 203
Gong, Y., Miao, H., Zhou, X., et al. 2025, Sci. China-Phys. Mech. Astron., 68, 280402
Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, ApJ, 622, 759
Harris, C. R., Millman, K. J., Van Der Walt, S. J., et al. 2020, Nature, 585, 357
Hildebrandt, H., Köhlinger, F., van den Busch, J. L., et al. 2020, A&A, 633, A69
Hoekstra, H., Yee, H. K. C., & Gladders, M. D. 2004, ApJ, 606, 67
Hortúa, H. J., Volpi, R., Marinelli, D., & Malagò, L. 2020, Phys. Rev. D, 102, 103509
Hu, W. 1999, ApJ, 522, L21
Hunter, J. D. 2007, Comput. Sci. Eng., 9, 90
Hyvärinen, A. & Oja, E. 2000, Neural Networks, 13, 411
Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, ApJ, 873, 111
Jarvis, M. 2015, TreeCorr: Two-point correlation functions, Astrophysics Source Code Library, record ascl:1508.007
Jarvis, M., Bernstein, G., & Jain, B. 2004, MNRAS, 352, 338
Jeffrey, N., Abdalla, F. B., Lahav, O., et al. 2018, MNRAS, 479, 2871
Jeffrey, N., Lanusse, F., Lahav, O., & Starck, J.-L. 2020, MNRAS, 492, 5023
Kacprzak, T., Fluri, J., Schneider, A., Refregier, A., & Stadel, J. 2023, J. Cosmology Astropart. Phys., 2023, 050
Kaiser, N. & Squires, G. 1993, ApJ, 404, 441
Kilbinger, M. 2015, Reports on Progress in Physics, 78, 086901
Kratochvil, J. M., Lim, E. A., Wang, S., et al. 2012, Phys. Rev. D, 85, 103513
Lanzieri, D., Zeghal, J., Lucas Makinen, T., et al. 2025, A&A, 697, A162
Leclercq, F. 2025, arXiv preprint arXiv:2509.13435
Leclercq, F. & Heavens, A. 2021, MNRAS, 506, L85
Lemos, P., Cranmer, M., Abidi, M., et al. 2023, Mach. Learn.: Sci. Technol., 4, 01LT01
Liu, J. & Haiman, Z. 2016, Phys. Rev. D, 94, 043533
Liu, T., Quan, Y., Su, Y., et al. 2025, Nat. Astron., 9, 608
Liu, Z., Zhang, J., Li, H., Shen, Z., & Liu, C. 2024, Sci. China-Phys. Mech. Astron., 67, 270413
Lu, T., Haiman, Z., & Li, X. 2023, MNRAS, 521, 2050
Mandelbaum, R. 2018, ARA&A, 56, 393
Massey, R., Rhodes, J., Ellis, R., et al. 2007, Nature, 445, 286
Meneghetti, M. 2022, Introduction to Gravitational Lensing: With Python Examples
Min, Z., Xiao, X., Ding, J., et al. 2024, Phys. Rev. D, 110, 063531
Oguri, M., Miyazaki, S., Hikage, C., et al. 2018, PASJ, 70, S26
Padilla, L. E., Tellez, L. O., Escamilla, L. A., & Vazquez, J. A. 2021, Universe, 7, 213
Papamakarios, G. & Murray, I. 2016, in Advances in Neural Information Processing Systems, ed. D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, & R. Garnett, Vol. 29 (Curran Associates, Inc.)
Paszke, A., Gross, S., Massa, F., et al. 2019, in Advances in Neural Information Processing Systems, ed. H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, & R. Garnett, Vol. 32 (Curran Associates, Inc.)
Petri, A., Haiman, Z., Hui, L., May, M., & Kratochvil, J. M. 2013, Phys. Rev. D, 88, 123002
Petrillo, C. E., Tortora, C., Chatterjee, S., et al. 2017, MNRAS, 472, 1129
Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, A&A, 641, A6
Schawinski, K., Zhang, C., Zhang, H., Fowler, L., & Santhanam, G. K. 2017, MNRAS, 467, L110
Schneider, P., van Waerbeke, L., Kilbinger, M., & Mellier, Y. 2002, A&A, 396, 1
Schrabback, T., Hartlap, J., Joachimi, B., et al. 2010, A&A, 516, A63
Shan, H., Kneib, J.-P., Tao, C., et al. 2012, ApJ, 748, 56
Shan, H., Liu, X., Hildebrandt, H., et al. 2018, MNRAS, 474, 1116
Shan, H. Y., Kneib, J.-P., Comparat, J., et al. 2014, MNRAS, 442, 2534
Shi, Y., Zhang, P., Deng, F., et al. 2025, J. Cosmology Astropart. Phys., 2025, 038
Shi, Y., Zhang, P., Sun, Z., & Wang, Y. 2024, Phys. Rev. D, 109, 123530
Shirasaki, M. & Ikeda, S. 2024, Open J. Astrophys., 7, 42
Shirasaki, M., Moriwaki, K., Oogi, T., et al. 2021, MNRAS, 504, 1825
Spergel, D., Gehrels, N., Baltay, C., et al. 2015, arXiv e-prints, arXiv:1503.03757
Starck, J.-L., Pires, S., & Réfrégier, A. 2006, A&A, 451, 1139
Stopyra, S., Peiris, H. V., Pontzen, A., Jasche, J., & Lavaux, G. 2024, MNRAS, 527, 1244
Su, C., Shan, H., Zhao, C., Xu, W., & Zhang, J. 2025, A&A, 699, A174
Tang, J. Y. & Fan, Z. H. 2005, ApJ, 635, 60
Taylor, P. L. & Markovič, K. 2022, Phys. Rev. D, 106, 063536
Tejero-Cantero, A., Boelts, J., Deistler, M., et al. 2020, JOSS, 5, 2505
Van Der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, Comput. Sci. Eng., 13, 22
Van Waerbeke, L., Benjamin, J., Erben, T., et al. 2013, MNRAS, 433, 3373
Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, Nat. Methods, 17, 261
Wu, Y., Li, N., Shan, H.-Y., et al. 2025, Res. Astron. Astrophys., 25, 125003
Yao, J., Shan, H., Li, R., et al. 2024, MNRAS, 527, 5206
Yao, J., Shan, H., Zhang, P., et al. 2023, A&A, 673, A111
Zeghal, J., Lanzieri, D., Lanusse, F., et al. 2025, A&A, 699, A327
Zhan, H. 2021, Chinese Science Bulletin (Chinese Version), 66, 1290
Zhang, Z., Shan, H., Li, N., et al. 2024, A&A, 683, A209
Zhong, F., Luo, R., Napolitano, N. R., et al. 2025, ApJS, 277, 12
Zonca, A., Singer, L., Lenz, D., et al. 2019, JOSS, 4, 1298

Appendix A: SBI results

This appendix presents the full set of inferred posteriors from 10 independent SBI realizations, together with 6 types of summary statistics (Table 1), for a fiducial cosmology with $\Omega_m = 0.3062$ and $\sigma_8 = 0.7615$ over 655.4 deg^2.

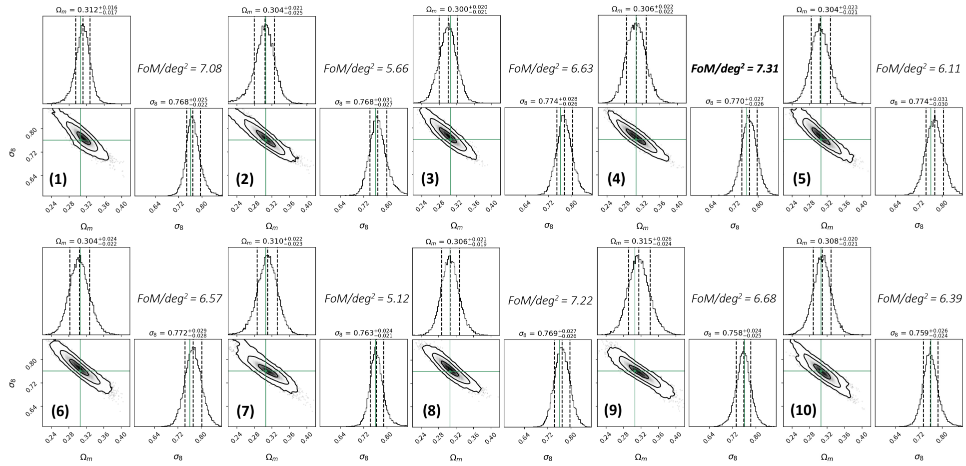


Fig. A.1: SBI posteriors based on 2PCF ($\hat{\xi}_{+-}$) of $\gamma^{\rm Avg}$.

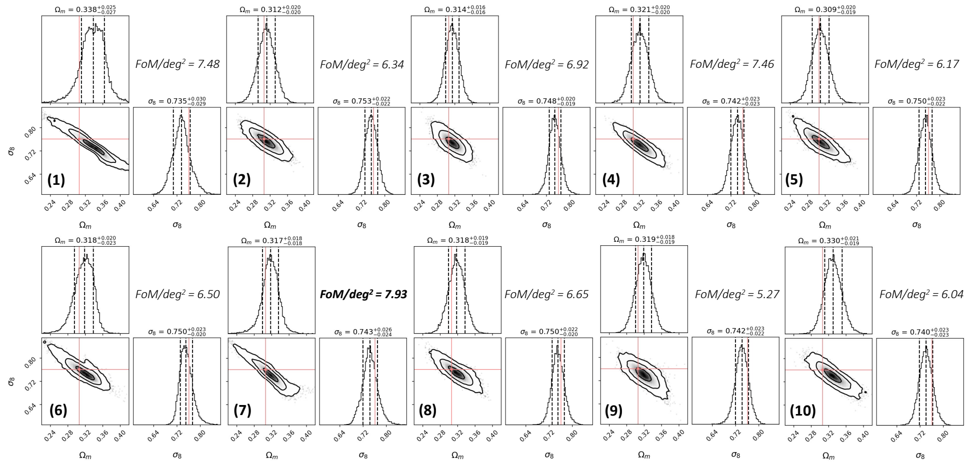


Fig. A.2: SBI posteriors based on ML-derived features ($x_{\tt ML}$) of $\gamma^{\rm Avg}$.

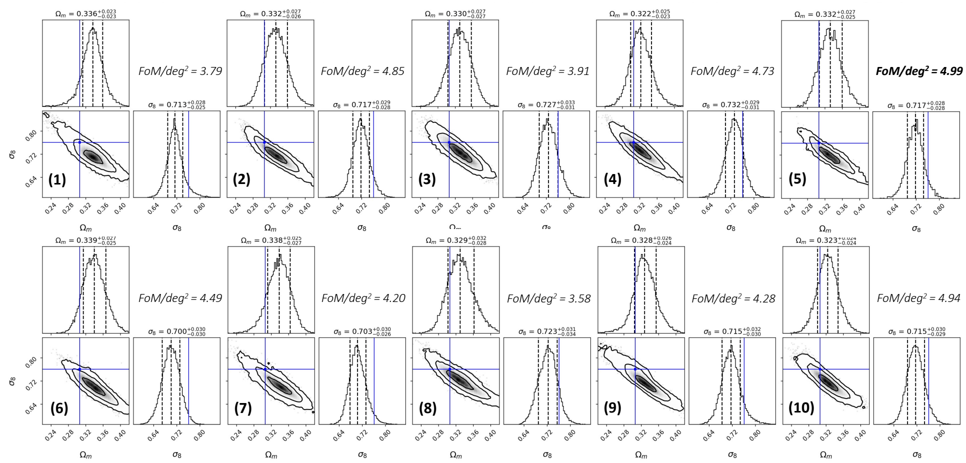


Fig. A.3: SBI posteriors based on ML-derived features ($x_{\tt ML}$) of $\kappa^{\rm Avg}$.

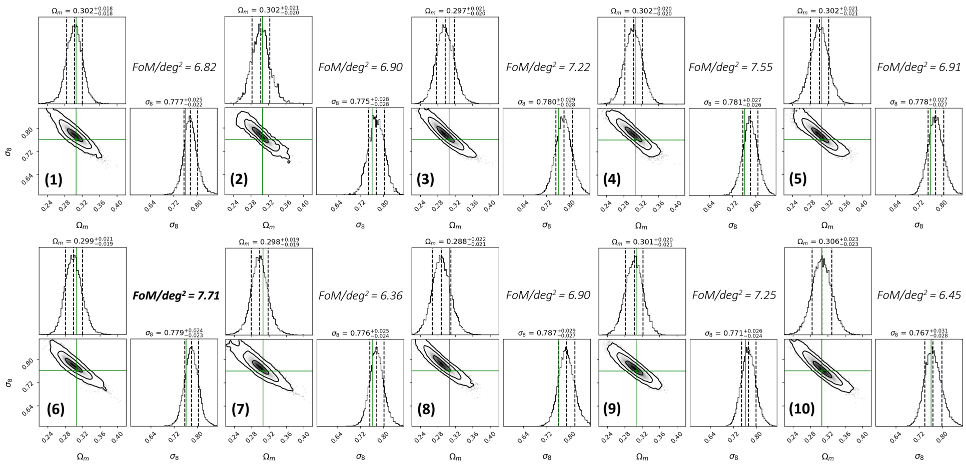


Fig. A.4: SBI posteriors based on 2PCF ($\hat{\xi}_{+-}$) of $\gamma^{\rm PCA}$.

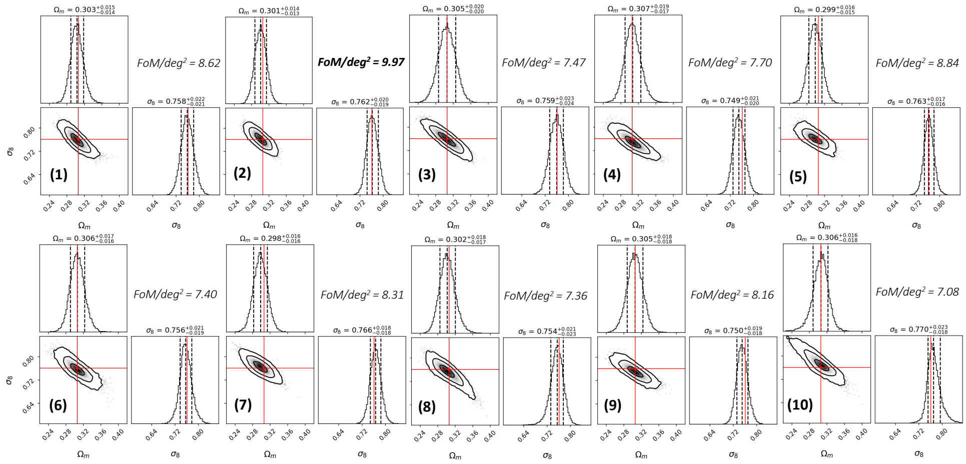


Fig. A.5: SBI posteriors based on ML-derived features ($x_{\tt ML}$) of $\gamma^{\rm PCA}$.

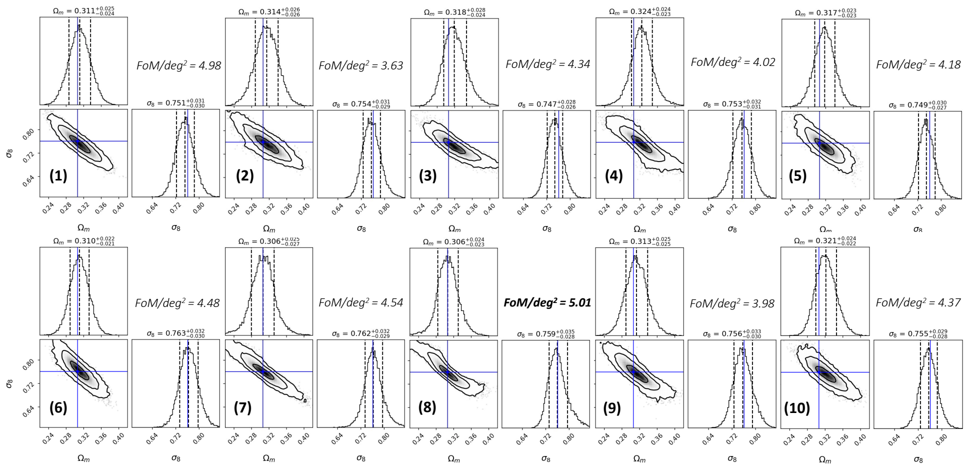


Fig. A.6: SBI posteriors based on ML-derived features ($x_{\tt ML}$) of $\kappa^{\rm PCA}$.