

中国科学院上海天文台
Shanghai Astronomical Observatory,Chinese Academy of Sciences

LLM-Driven Search for Shear-Field Denoising Algorithms in KiDS-Legacy

Forward Modeling · LLM-Driven Algorithm Discovery · Simulation-Based Inference

Presenter: Jiacheng Ding (Postdoc @ SHAO)
Co-authors: Chen Su, Hekun Li, Ji Yao, Huanyuan Shan

2026.10.20 @ AI+Astro Workshop, Zhejiang Lab

Outline

01 · Motivation

From two-point statistics to field-level analysis
Pixel-level S/N and shear denoising

02 · Forward Modeling

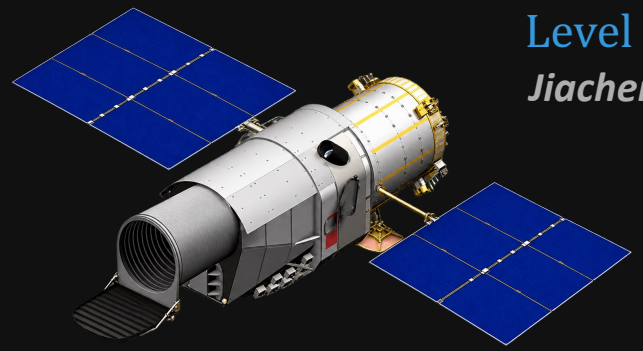
Observation-matched mock: noisy/true pairs

03 · Denoising Algorithm Search

Metrics + LLM × MCTS

04 · Cosmological Inference

raw / clean shear → $\xi_{\pm}$ → SBI

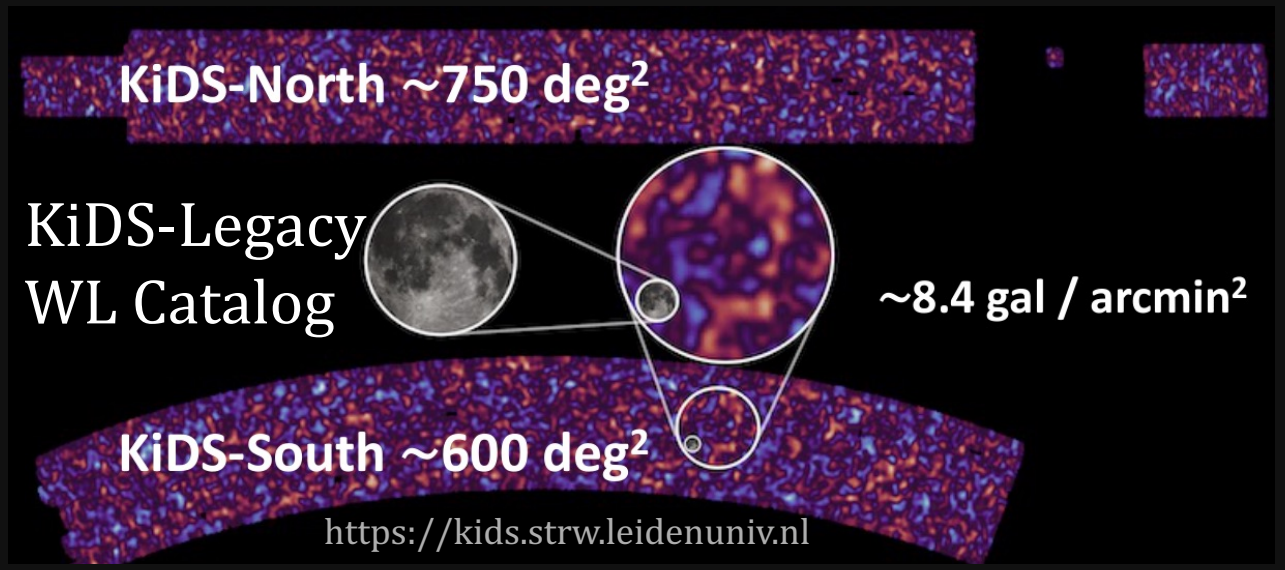


~26 gal / arcmin²

CSST-like simple mock → Field-Level Inference

***Jiacheng Ding** et al. A&A 712 (2026): A49.*

- Real specz-photoz ?
- Sky footprint ?
- Uneven sampling ?
- Other systematics ?



Weak-Lensing Information: Why Denoising Matters

Complementary Information

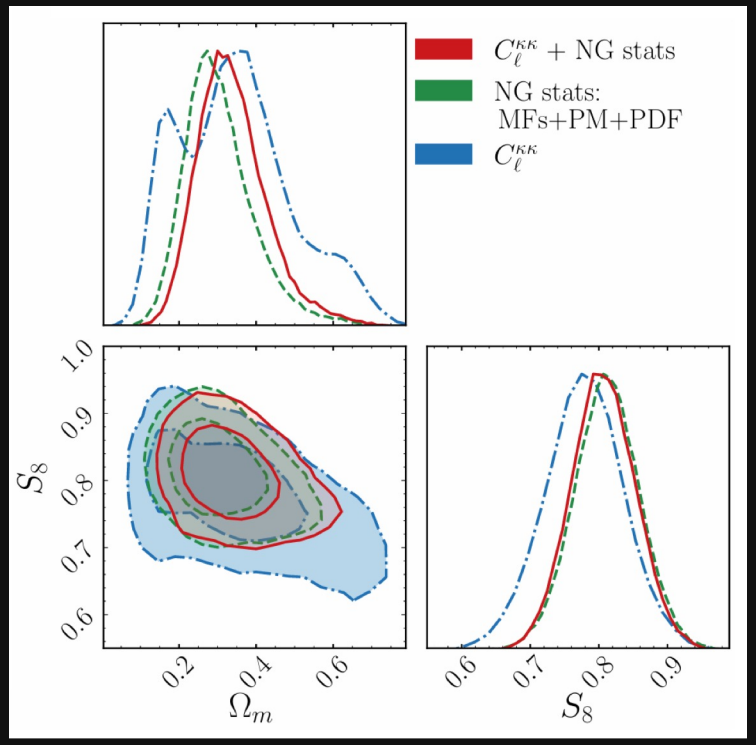
Two-point : $\xi\pm$, $C\ell$
Higher-order : peaks, PDF, Minkowski
Field-level / ML : spatial structure

The Pixel-Level Challenge

Shape noise, masks, uneven sampling obscure informative field structure.

Denoising Must Preserve

Amplitude • Morphology • Tomography

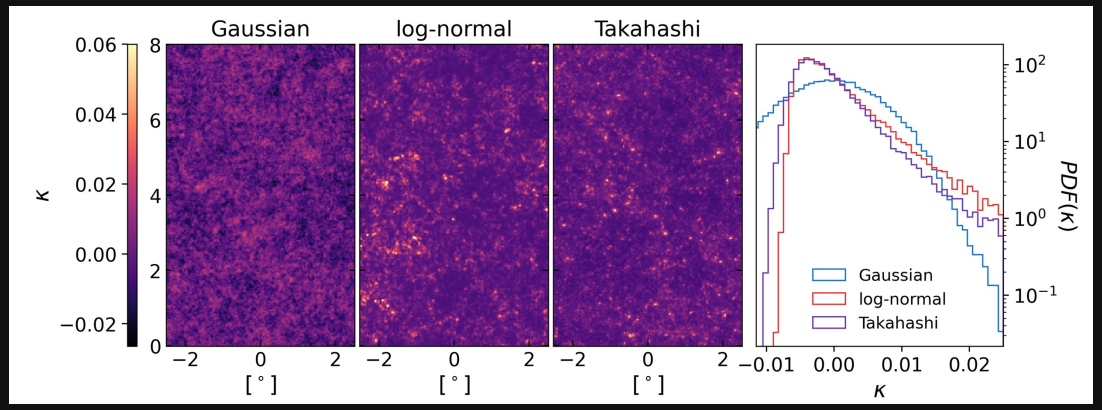


Posteriors from 2-point, higher-order and joint statistics

arXiv:2409.01301

Same statistics, different field

arXiv:2312.08934



Shear-Field Denoising: Feasibility & Scientific Value

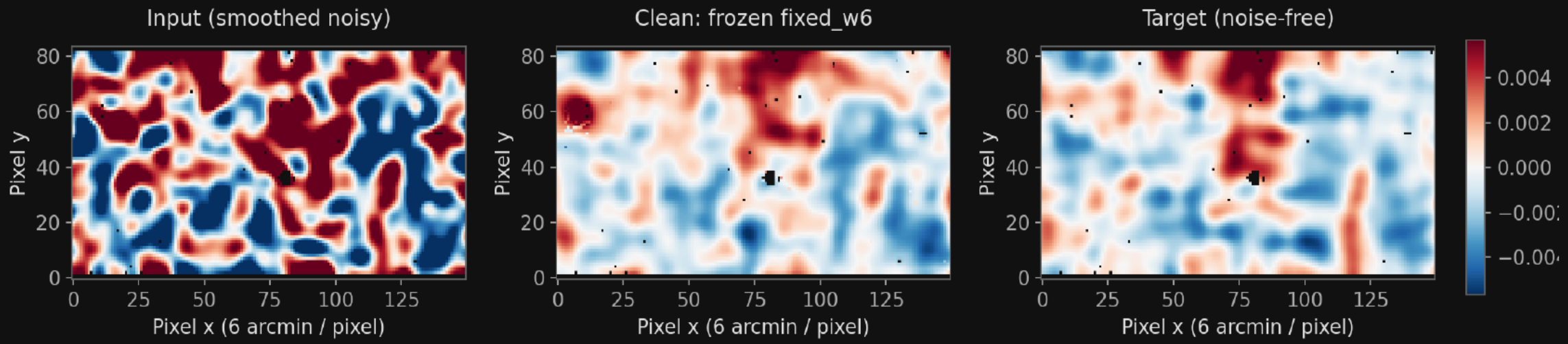
Why It Is Feasible

Spatial correlation: coherent signal vs. random shapes
Multiscale structure: sparsity and cross-scale coherence
E/B modes: E-dominated lensing vs. noise in both
Tomography: shared foregrounds across redshift bins

What It Enables

Higher effective S/N
Recovery of non-Gaussian structure
More stable higher-order and ML inference
More information from existing survey data

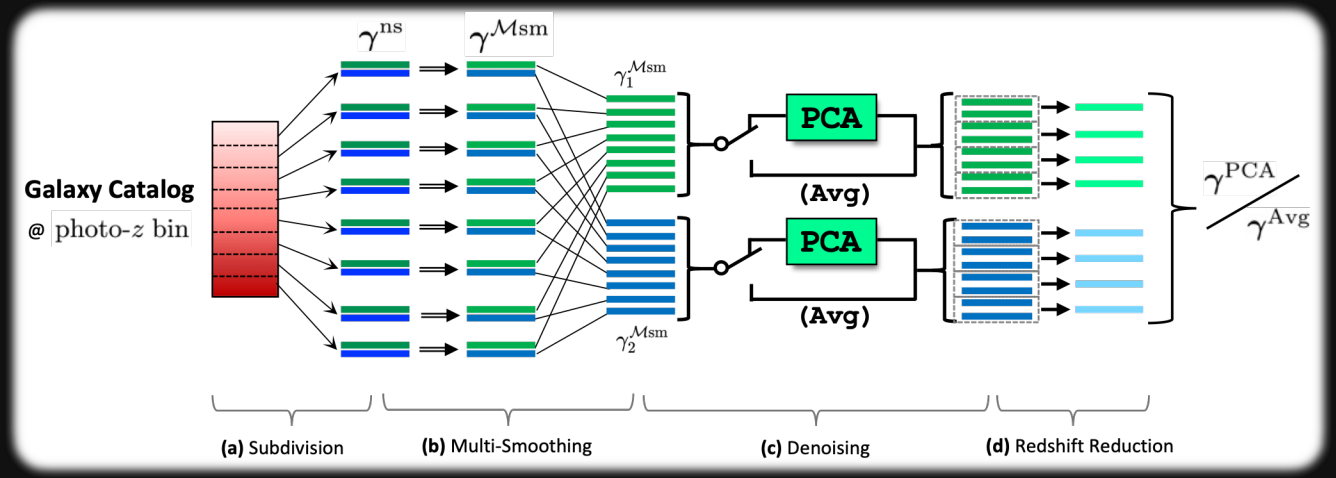
The mock data (noisy-truth pairs) support algorithm search and independent validation.
Truth is used only for development and validation.

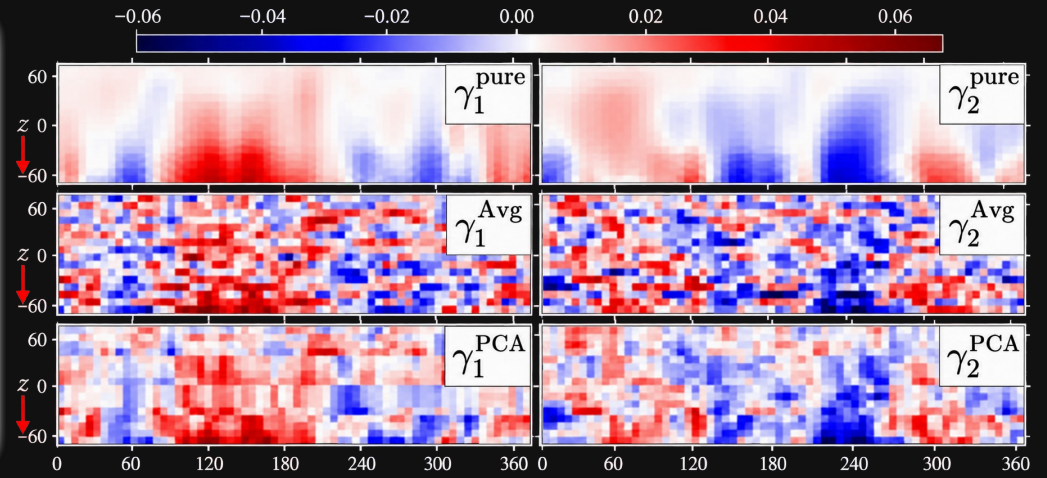


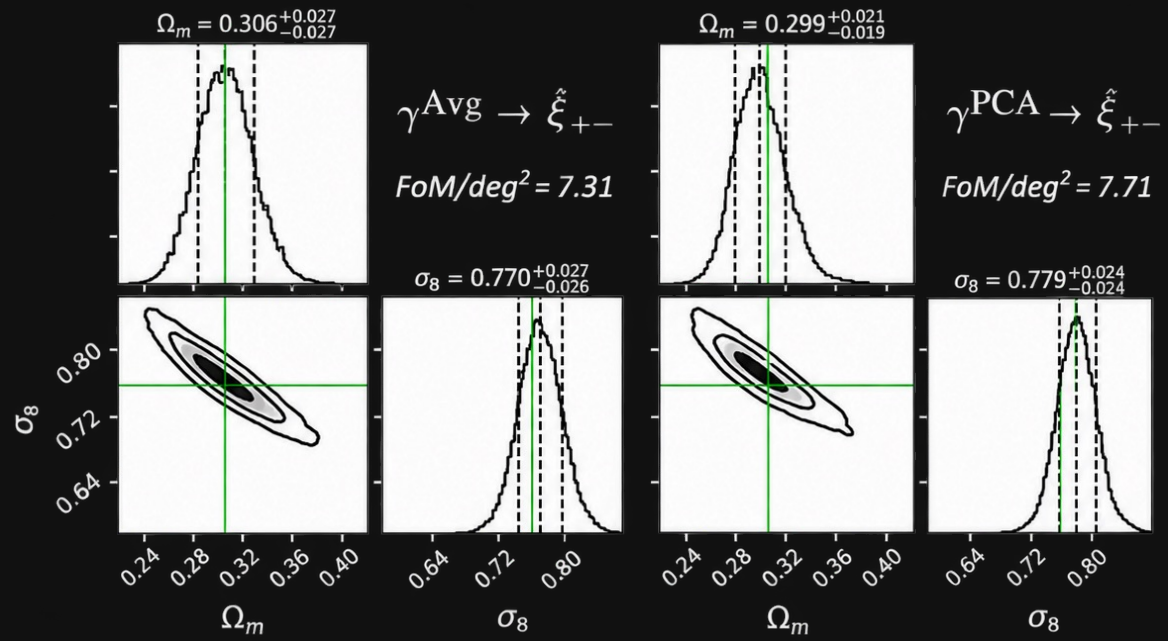
Core scientific objective: Suppress noise → Preserve amplitude, structure & tomography → Robust inference

Baseline Algorithms – GPCA (Guassian Smoothing + PCA)

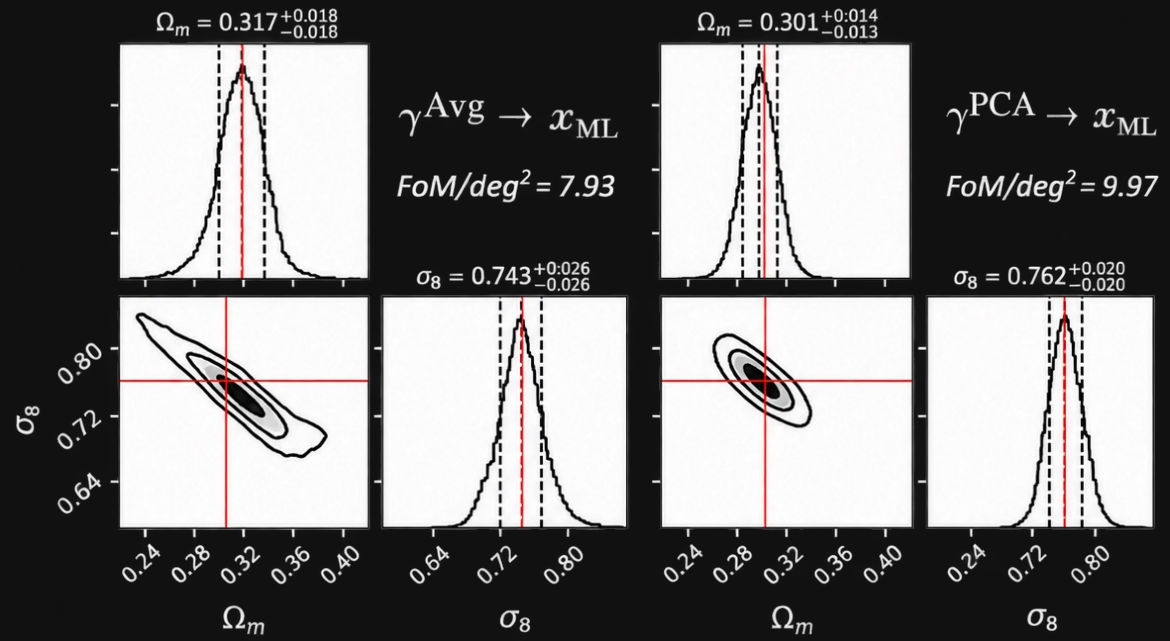
***Jiacheng Ding** et al. A&A 712 (2026): A49.*







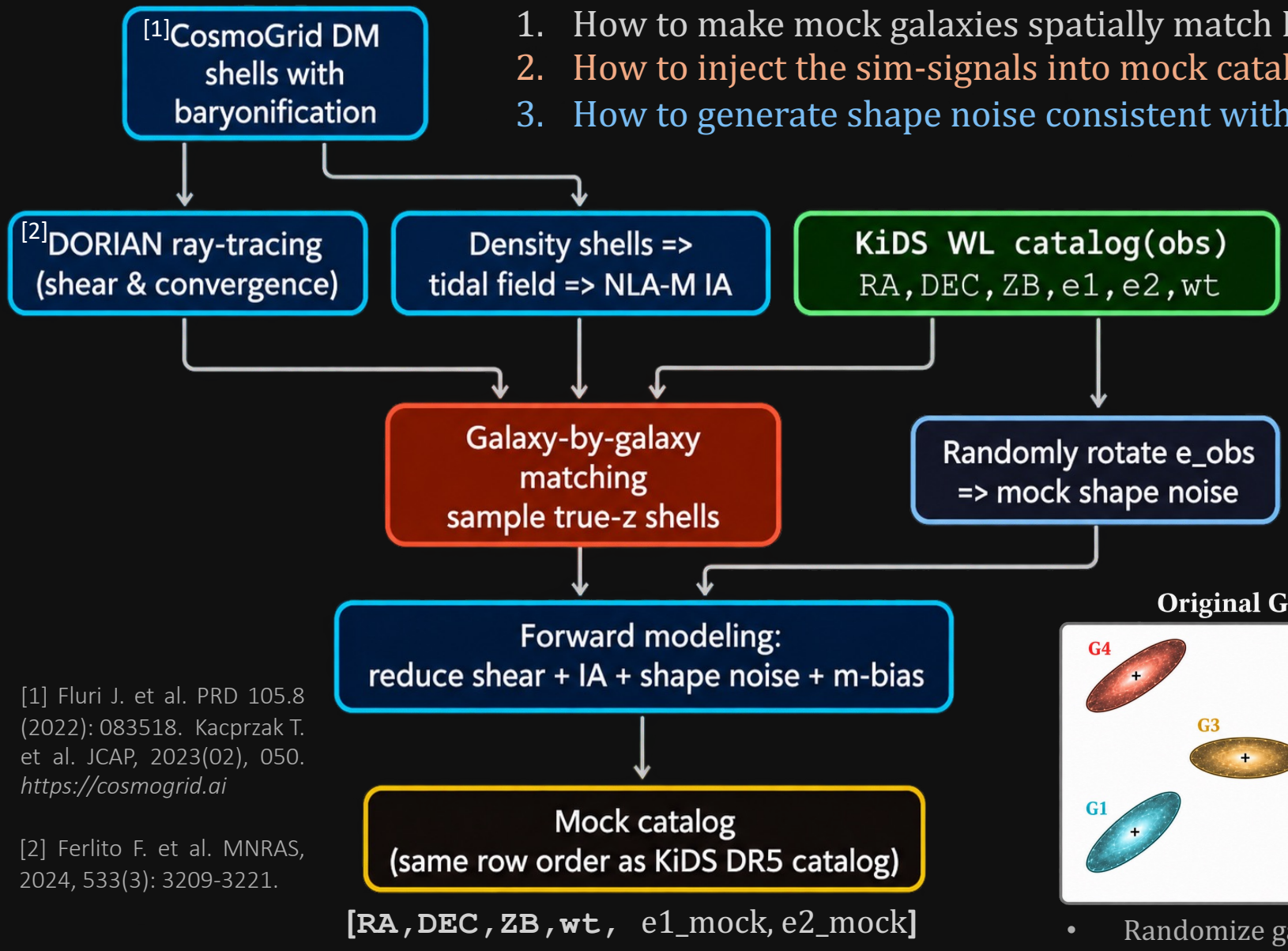
Denoise + 2-pt statistics => FoM + 5%

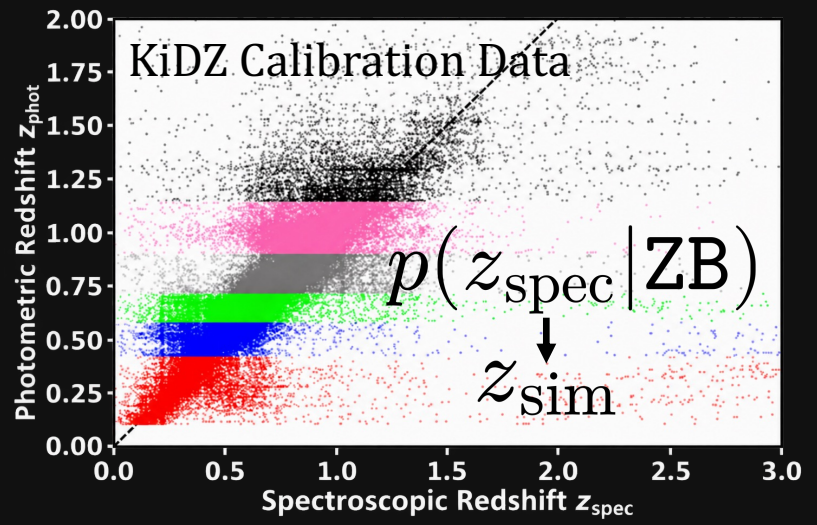


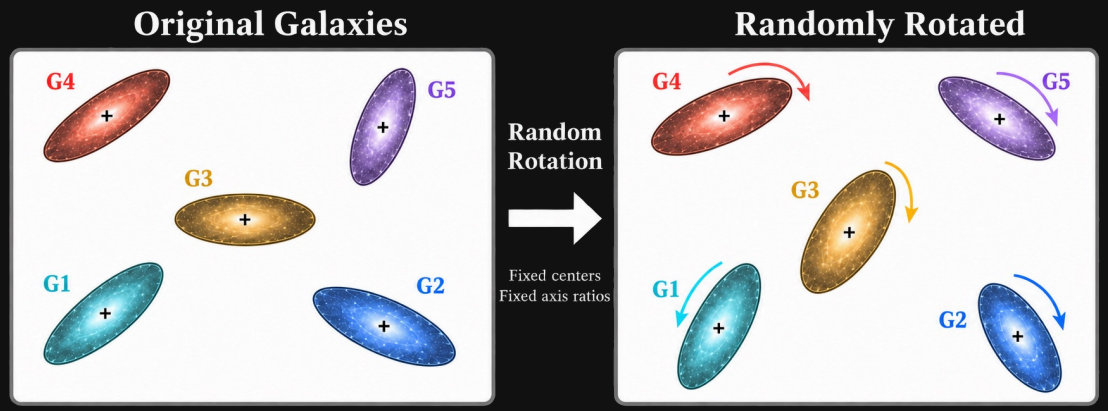
Denoise + Field-level => FoM + 26 %

Forward Modeling: Matching KiDS-Legacy WL catalog

1. How to make mock galaxies spatially match KiDS observation ?
2. How to inject the sim-signals into mock catalogs ?
3. How to generate shape noise consistent with the KiDS weight ?







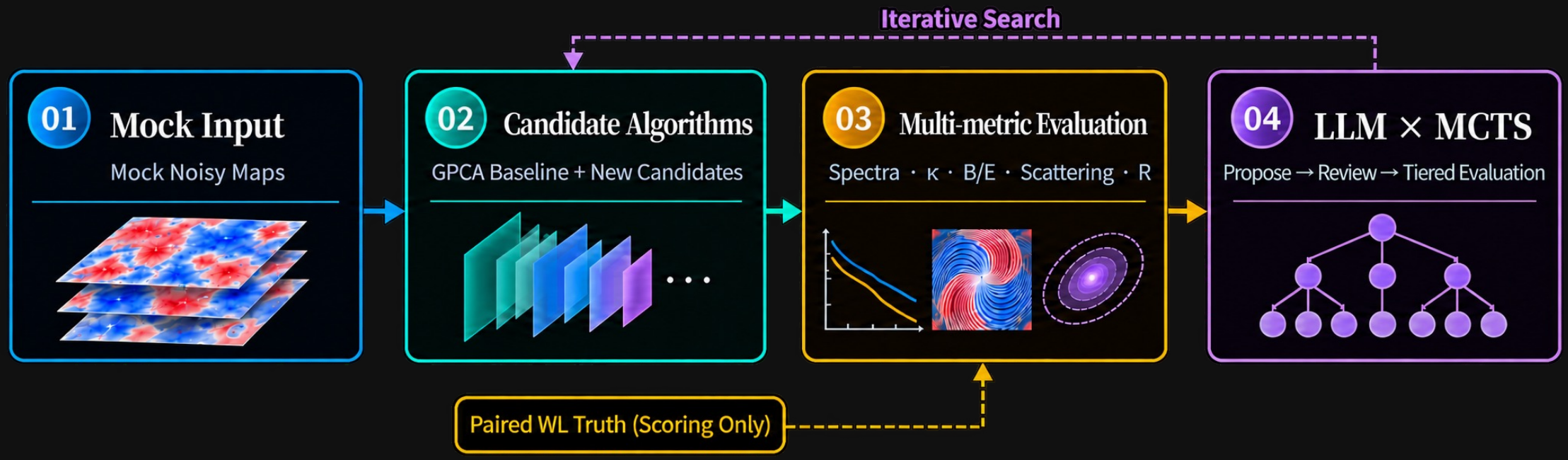
- Randomize galaxy orientations to remove coherent shear;
- Preserve shape noise & survey sampling.

[1] Fluri J. et al. PRD 105.8 (2022): 083518. Kacprzak T. et al. JCAP, 2023(02), 050. *https://cosmogrid.ai*

[2] Ferlito F. et al. MNRAS, 2024, 533(3): 3209-3221.

How to make an LLM find scientifically valuable denoising algorithms?

Define metrics first, then search



Metrics for Assessing Output–Target Map Consistency

$$J = 0.35\,\Delta\mathrm{Spec} + 0.15\,\Delta B/E + 0.25\,\Delta\kappa + 0.25\,S - 0.05\,V - 0.20\,|R-1|$$

Relative to GPCA · Aggregate over samples and seeds

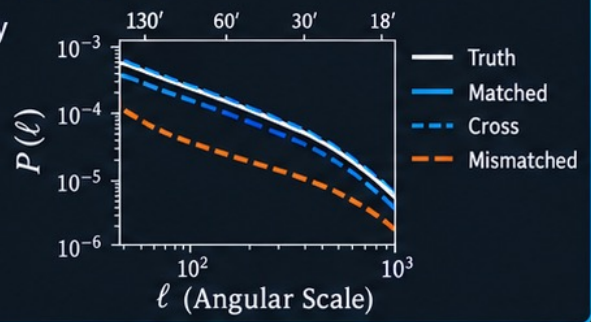
35% | Spectral Coherence

Band-limited spectral consistency

18′ – 130′ · Lower is better (ideal 0)

$$\left\langle \left| \ln \frac{P_{\mathrm{pred}}\,P_{\mathrm{true}}}{P_{\mathrm{cross}}^2} \right| \right\rangle$$

Scale-invariant

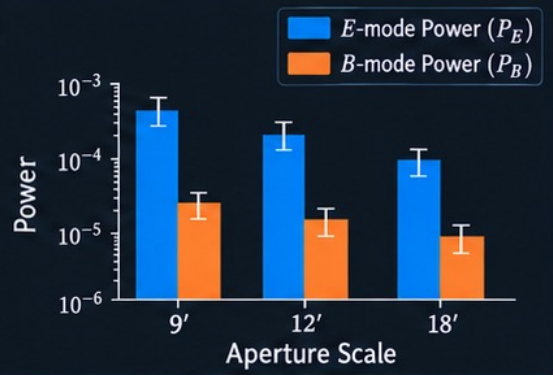


15% | B/E Leakage

Compensated-aperture B/E power

9′ · 12′ · 18′ · Lower is better

Scale-invariant



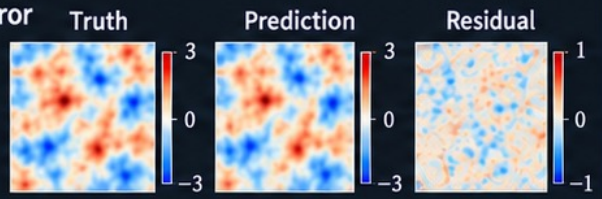
25% | Convergence Accuracy

Kaiser–Squires reconstruction error

Lower is better (ideal 0)

$$\left\langle \left(\kappa_{E,\mathrm{pred}} - \kappa_{E,\mathrm{true}}\right)^2 \right\rangle$$

Amplitude + spatial errors

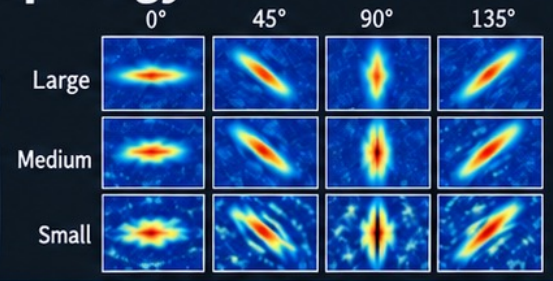


25% | Multiscale Morphology

Wavelet-modulus correlation

18′ – 120′ · Higher is better (ideal 1)

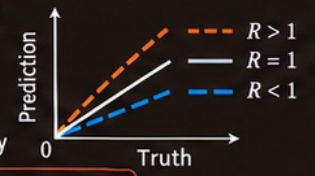
$$S = \mathrm{clip}\left[\frac{(r_{\mathrm{cand}} - r_{\mathrm{base}}}{(1 - r_{\mathrm{base}}}, -1, 1\right]$$



Amplitude Response R

$$R = \left\langle \frac{\mathrm{Cov}(\mathrm{truth}, \mathrm{pred})}{\mathrm{Var}(\mathrm{truth})} \right\rangle$$

$R = 1$: ideal · $R < 1$: shrink · $R > 1$: amplify



Penalty: $-0.20\,|R-1|$

$0.35 \le R \le 1.5$

Invalid / out of bounds → $J = -10^6$

Stability V

Variation across samples / seeds

$$V = \mathrm{mean}_k \ \min\left[\frac{\mathrm{std}_k}{\max(|\mu_k|, 10^{-12})}, 10\right]$$

Lower is better · Penalty: $-0.05\,V$

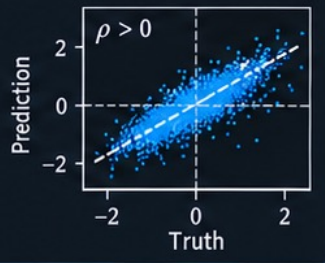
Across all enabled metrics

Spatial Correlation ρ

Diagnostic · Higher is better

$$\rho = \frac{\langle (\hat{x} - \bar{x})(\hat{y} - \bar{y}) \rangle}{\sqrt{\langle (\hat{x} - \bar{x})^2 \rangle \langle (\hat{y} - \bar{y})^2 \rangle}}$$

- No direct reward weight
- Included in stability penalty



LLM Search Framework

(Multi-Agent Monte Carlo Tree Search for Algorithm Discovery)

- `DeepSeek-V4-Pro` `(Deep Reasoning)`
- `DeepSeek-V4.1-Flash` `(Fast & Economical)`
- `GPT 5.6,6.1` `(Usage-Capped)`

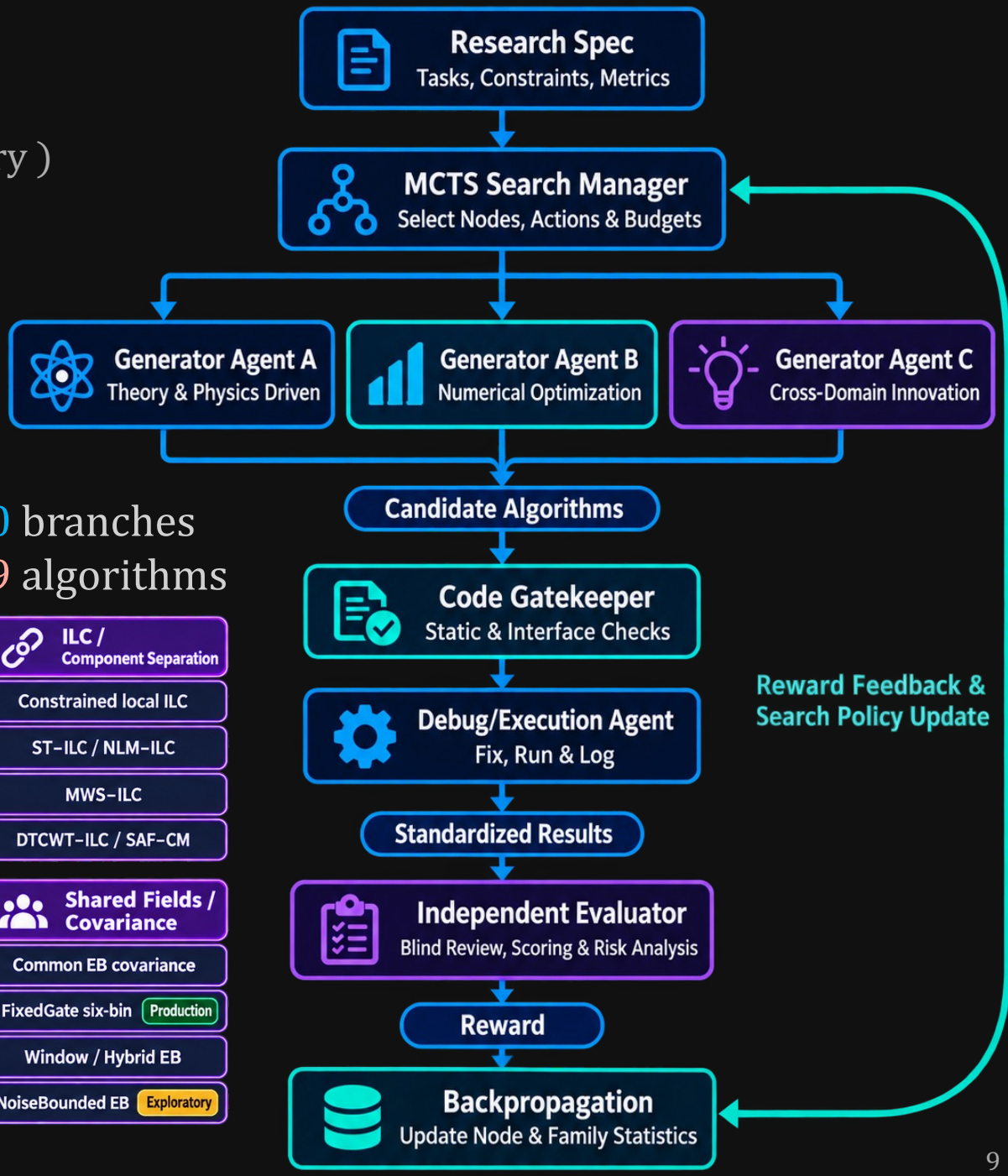
Inputs

- Fixed metrics
- KiDS–Legacy mock catalog
- Baseline denoising (GPCA)

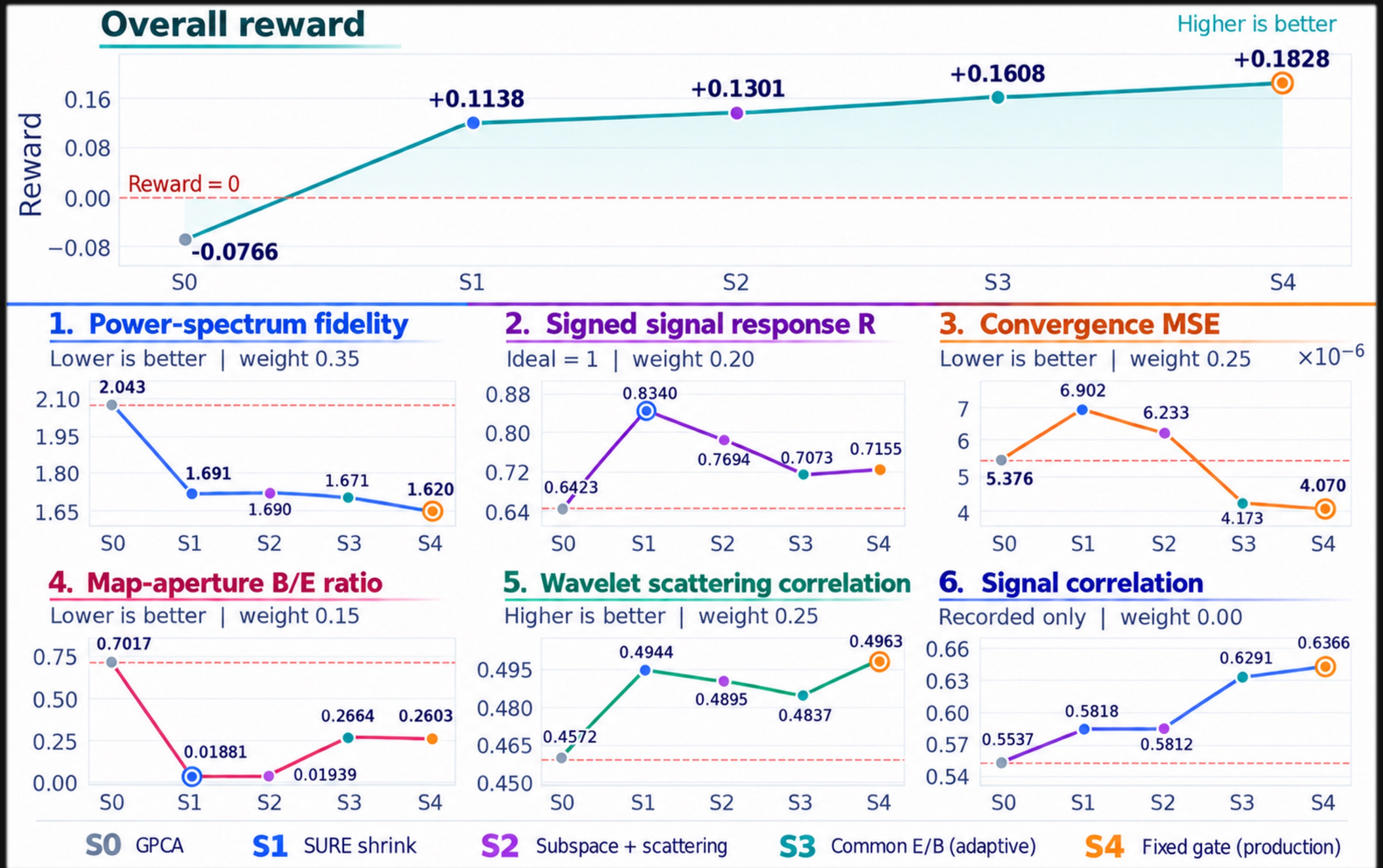
10 branches
49 algorithms

Smoothing / Local Statistics	Transforms / Shrinkage	Cross-Bin / Low-Rank Structure	Bayesian / Risk Shrinkage	ILC / Component Separation
Gaussian smoothing	Fourier / E–B Wiener	Redshift PCA	GP Wiener / EB–MAP	Constrained local ILC
Density-adaptive Gaussian	DCT sparse / GSM–BLS	Wavelet PCA / Tucker	James–Stein	ST–ILC / NLM–ILC
DoG Wiener / Bilateral	DWT / SWT / DTCWT	Nuclear norm	SURE / Stein shrinkage	MWS–ILC
Lee–Kuan / Whittaker	Starlet / Morlet Gating	Cadzow / Patch SVD	EB ensemble	DTCWT–ILC / SAF–CM

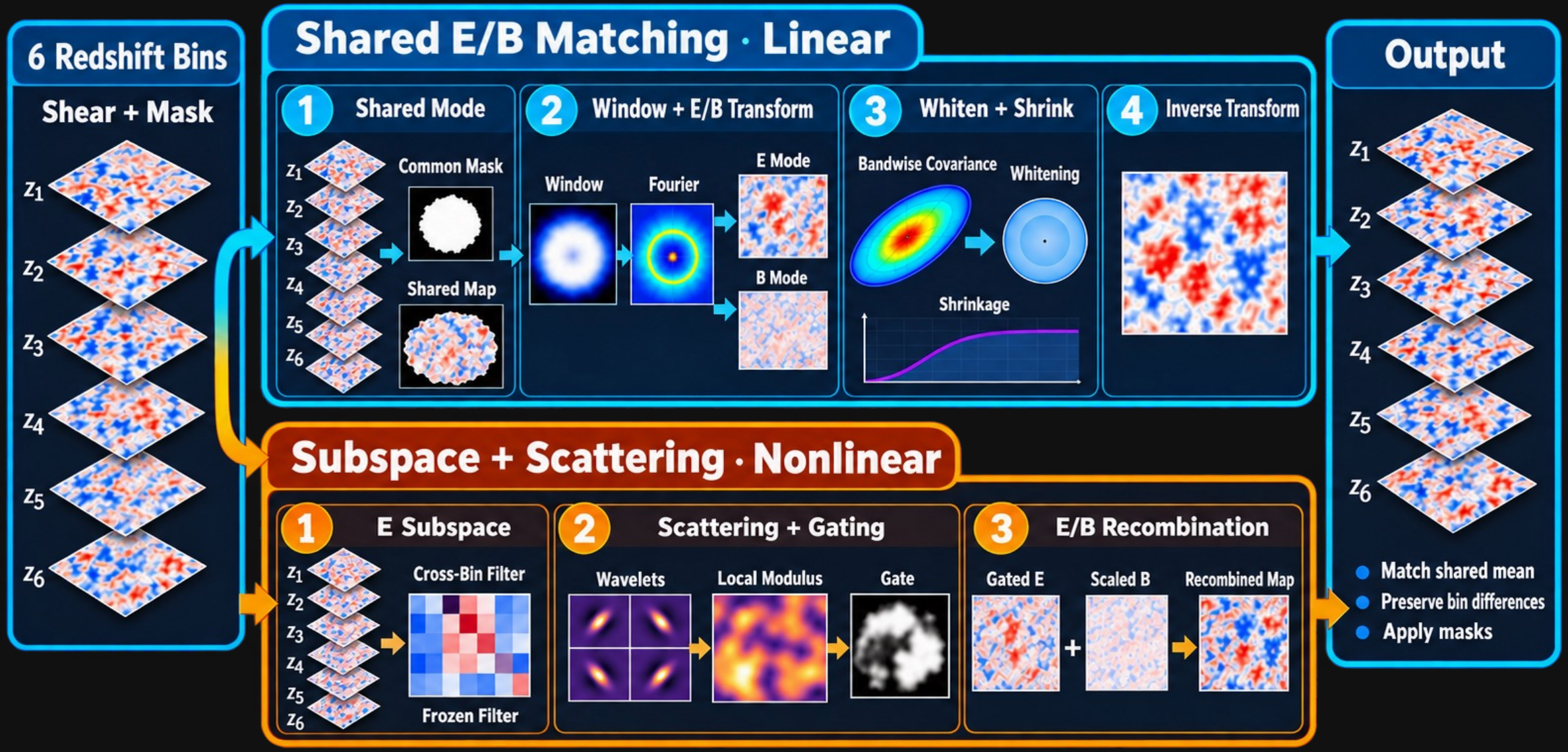
Optimization / Regularization	Diffusion / Decomposition	State Space / Prediction	Nonlocal / Patches	Shared Fields / Covariance
Masked pure–E CG	Anisotropic diffusion	Kalman filtering	NLM tomo	Common EB covariance
Tikhonov inversion	Spin–2 diffusion	RTS smoothing	NLM shear	FixedGate six-bin Production
PnP–ADMM + DTCWT	Graph diffusion	f–x prediction	Scattering NLM	Window / Hybrid EB
Group sparse + FISTA	EMD sifting	Ridge prediction	Patch SVD	NoiseBounded EB Exploratory



Algorithm Search Results: 5 milestone algorithms



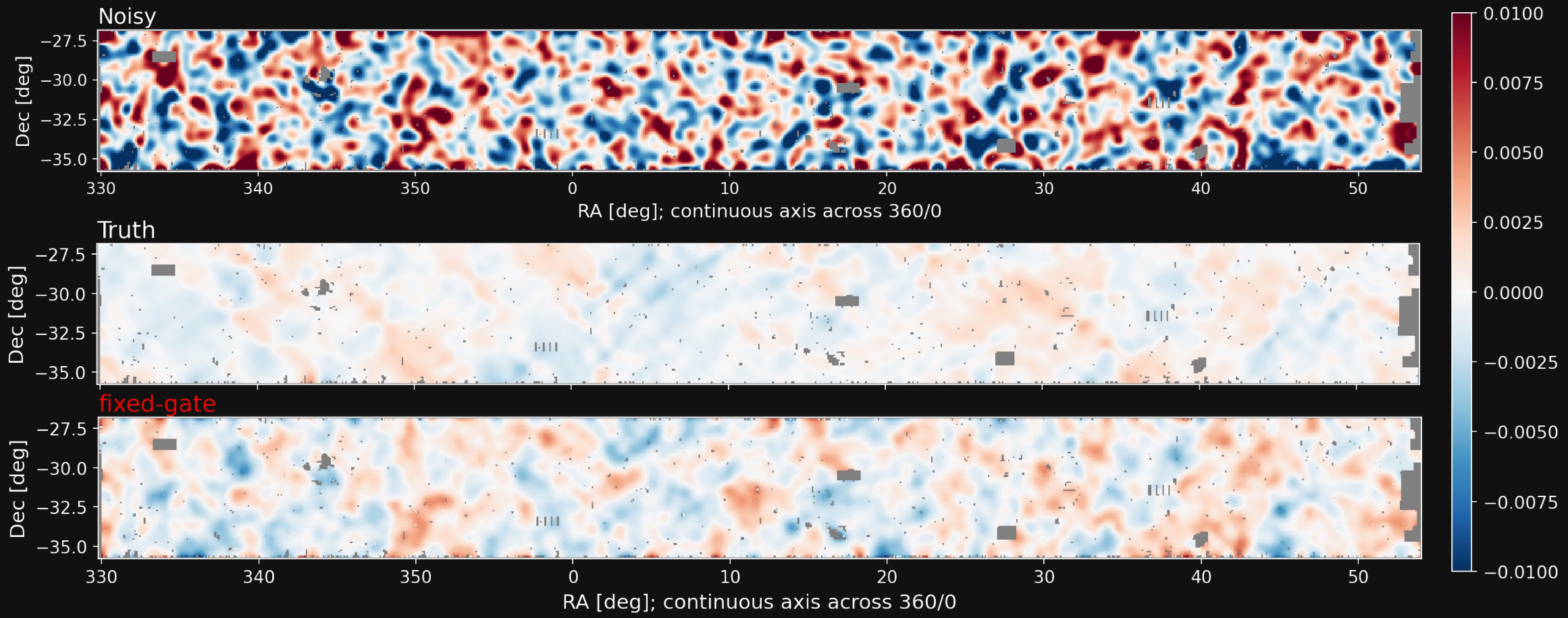
Current Best Denoising Algorithm: “FixedGate six-bin”



KiDS-like Mock => Denoising

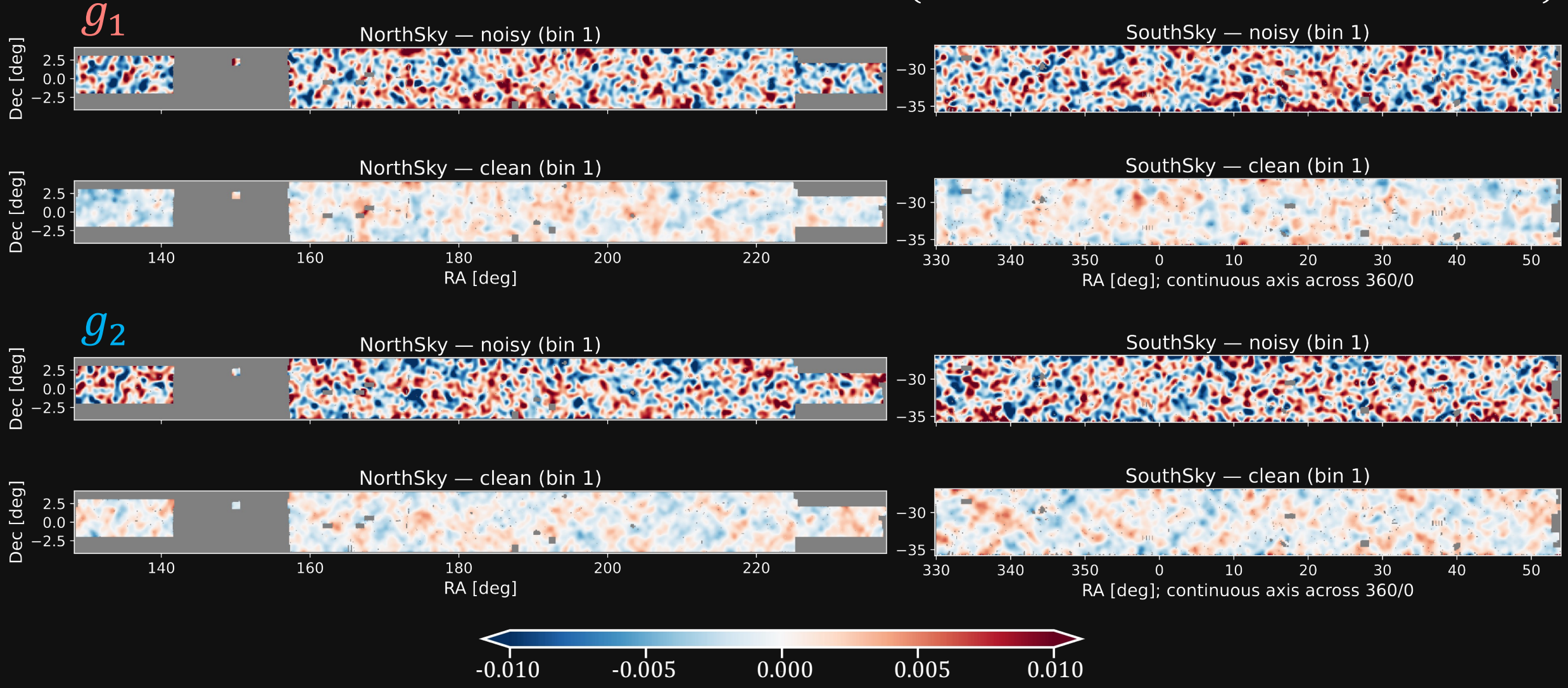
(SouthSky, the 1st tomo-bin of KiDS: $0.11 < \mathrm{ZB} < 0.43$)

g_2



KiDS–Legacy Observation => Denoising

(The 1st tomo-bin of KiDS: $0.11 < \mathrm{ZB} < 0.43$)

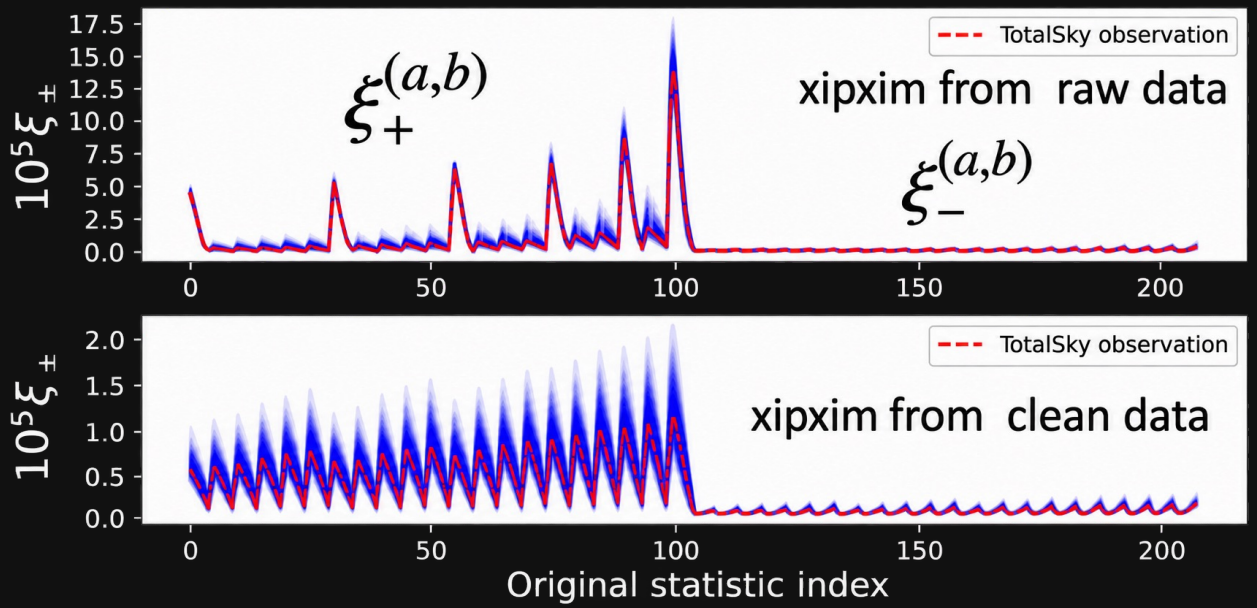


KiDS-obs. TotalSky: raw / clean $\xi_{\pm}$ → Simulation-Based Inference

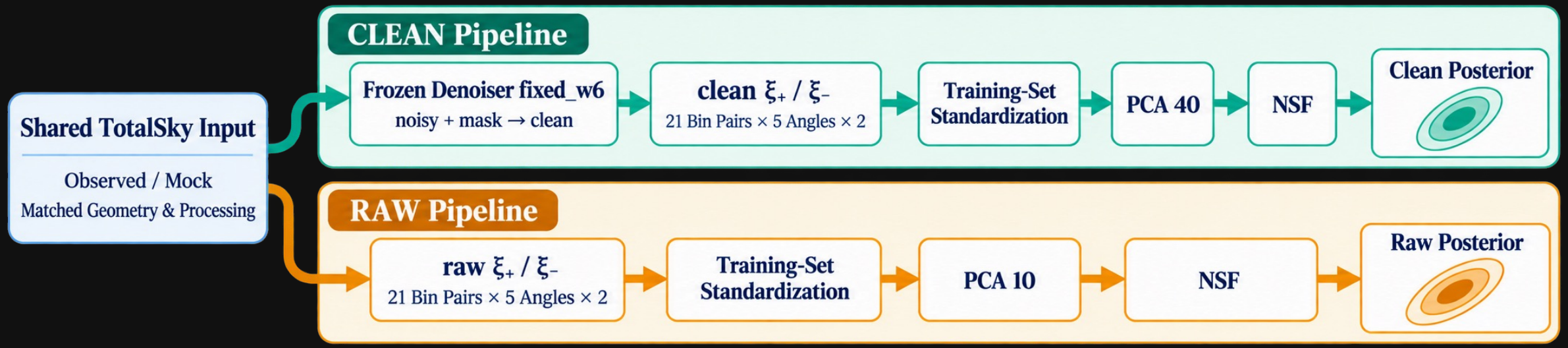
$$\xi_{+}^{(a,b)}(\vartheta) = \left\langle \gamma_{\mathrm{t}}^{a}\,\gamma_{\mathrm{t}}^{b} \right\rangle(\vartheta) + \left\langle \gamma_{\times}^{a}\,\gamma_{\times}^{b} \right\rangle(\vartheta)$$

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

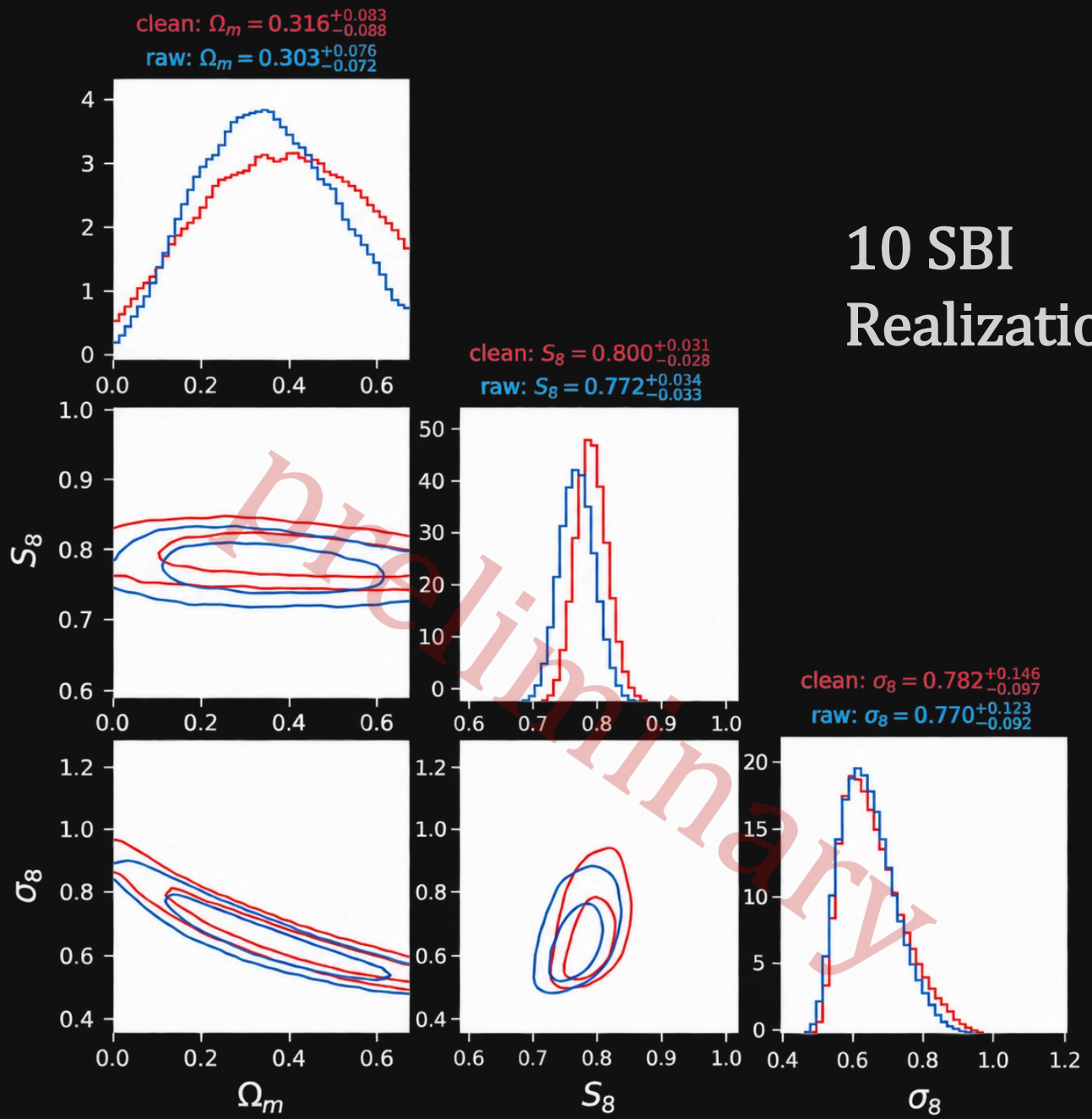
($\xi_{\pm}$ theta-bin: 10'~120' 10 bins)



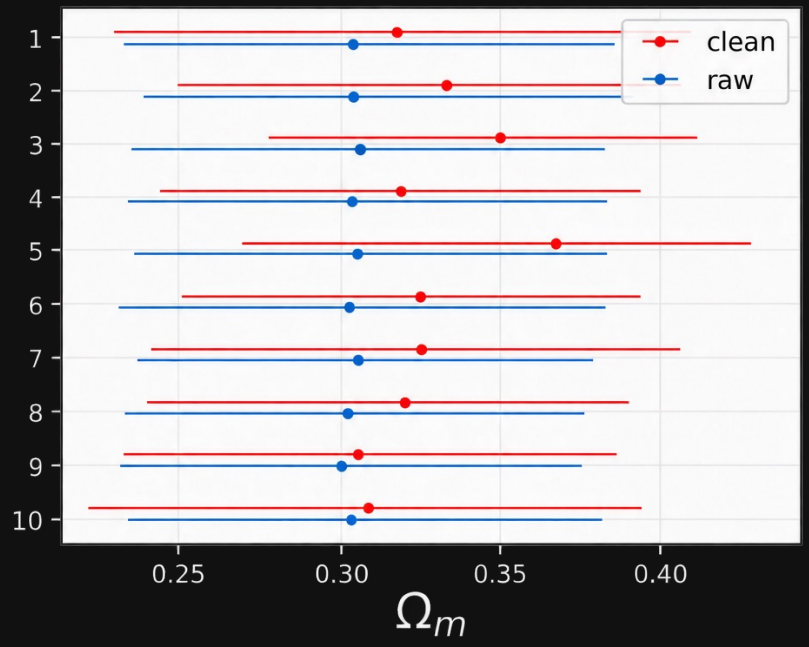
- raw $\xi_{\pm}$ / clean $\xi_{\pm}$ → SBI

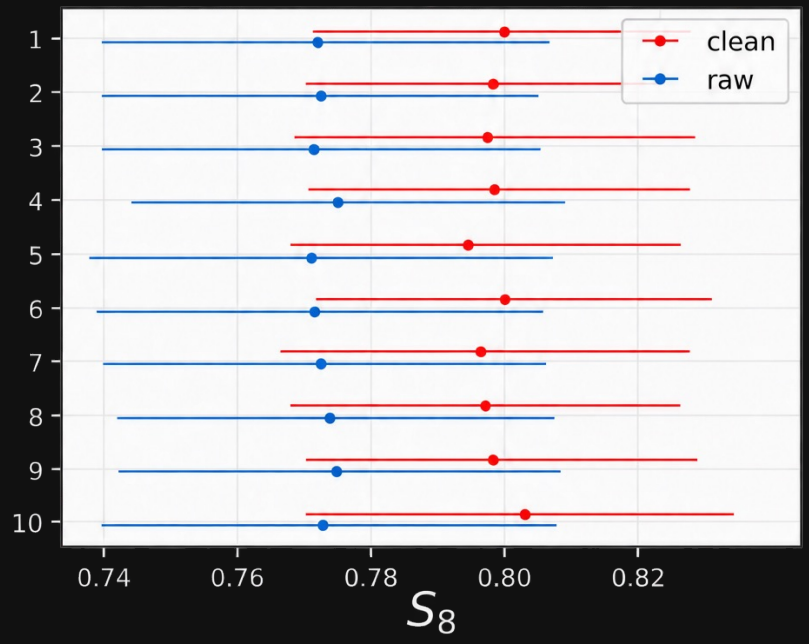


KiDS-obs. TotalSky: raw / clean $\xi_\pm$ → Cosmo-Inference



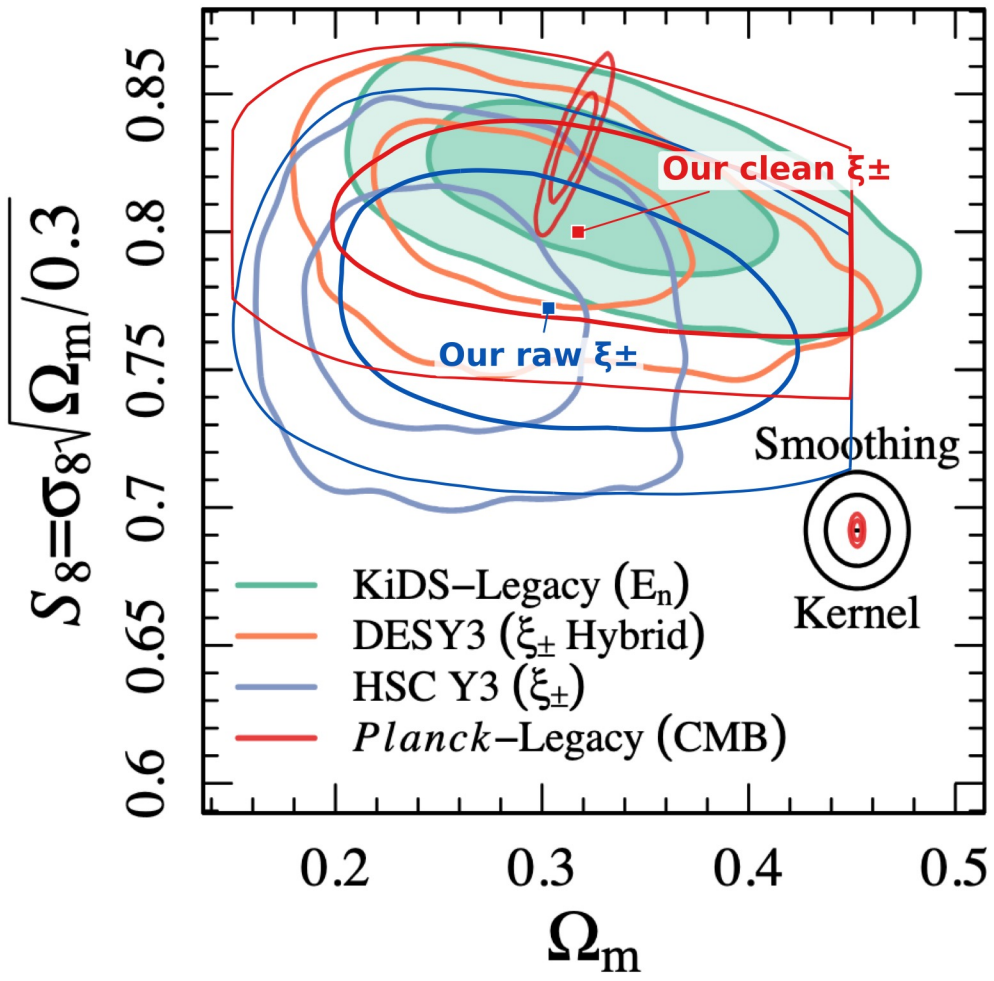
10 SBI Realizations:

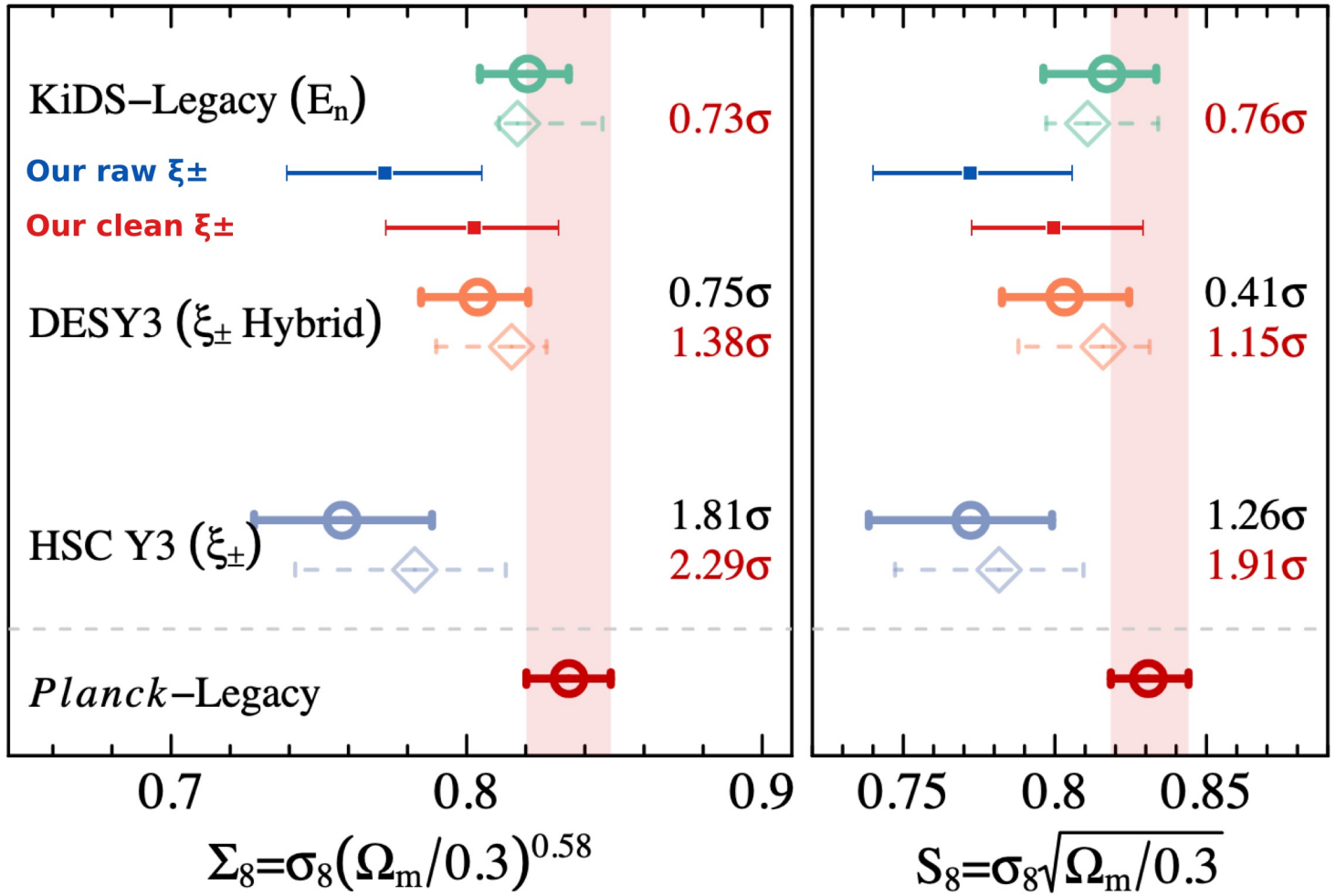




preliminary

Comparison with KiDS-Legacy and Other Surveys





A. H. Wright et al. (2025), A&A, 703, A158 · KiDS-Legacy: cosmological constraints from cosmic shear

raw (T=1.00): Om=0.3034 (+0.0745/-0.0708) S8=0.7721 (+0.0337/-0.0320) Sigma8=0.7725 (+0.0328/-0.0332)

clean (T=1.00): Om=0.3177 (+0.0815/-0.0862) S8=0.7997 (+0.0295/-0.0271) Sigma8=0.8027 (+0.0285/-0.0299)

Summary

Denoising keeps more than meets the eye

Forward Modeling

Reliable, data-matched mock pipeline

Metrics + LLM × MCTS

Cross-domain algorithms → Candidates
→ Estimate

Raw/clean 2-pt statistics → SBI

Cosmo-Inference → validates the denoiser

Next step:

- 2-pt statistics → High-order statistics
- Field-Level Inference
- Reconstruction of DM Field