Scattering transform synthesis of correlated foregrounds: benchmarking against diffusion models on FLAMINGO

Claude Code

13 May 2026

ABSTRACT

We study generative modelling of correlated extragalactic foregrounds (tSZ+CIB) on FLAMINGO simulations along two axes that have been conflated in prior work: *by-construction* versus *learned* statistics, and *supervised* versus *unsupervised* dependence on a truth ensemble. On N=20 patches at 150 GHz in $\ell \in [500, 6000]$ we deliver three contributions.

(i) A phase-preserving *joint* $N \times N$ *Cholesky* C_ℓ*-match* plus paired pixel histogram match (Eq. 12) which, applied to any multi-component generative sample (ST synthesis, DDPM, or a paired Gaussian random field), recovers all auto- and cross-spectra and the full 1-point pixel CDF by construction. The three post-recipe generators are statistically indistinguishable on every diagnostic we tested (ScatCov coefficient correlation $\sim$0.995, Minkowski M_1 peak 76–82% of truth, deepest cluster cores within $\sim$10% of truth $-350\,\mu K_{\rm CMB}$). The recipe extends to 3×3 (tSZ+CIB+kSZ) and 4×4 (tSZ_{150}+$\mathrm{CIB}_{90/150/217}$) configurations at numerical precision. A ScatCov-class *posterior refinement* (ST+HM polish, $\sim$ 1 s/patch) on top of the recipe reduces the ScatCov-distance to a training-set truth class by 2.7–3.3× while preserving the JM-locked CDF exactly, and replicates on the held-out test patches at the same rate; the recipe is therefore also the preferred ScatCov-class anchor for downstream non-Gaussian diagnostics, not only for the headline 1-point and 2-point statistics.

(ii) A *non-by-construction* pipeline that exposes exactly what the ScatCov coefficient vector can and cannot reproduce on its own. The tSZ×CIB pixel cross-r recovery ladder (Fig. 15) is 0% for single-channel raw ST synthesis, 53% for multi-channel ScatCov (N_c=2, 1600 LBFGS steps; asymptote $\sim$60%), 57% after a soft phase-preserving C_ℓ rescale that also brings M_1 to 97% of truth, and 100% after the Cholesky projection. The *ensemble-mode* variant (no per-patch truth at inference; only the ensemble-mean C_ℓ matrix from a fiducial simulation suite) reaches 56% cross-r recovery and is the deployable generator-agnostic non-BC pipeline. It strictly improves on raw DDPM at zero training cost (C_ℓ ratio 1.06 vs 0.58, cross-r 56% vs 50%); it does not deliver cluster-core morphology, which remains by-construction territory.

(iii) A methodological reframing of the ST/DDPM/calibration triplet on the supervised↔ unsupervised axis. Microcanonical SC-matching synthesis (Allys et al. 2020; Mousset et al. 2024) is unsupervised; trained DDPM (Prabhu et al. 2025) is supervised and inherits its training-set bias; our recipe is semi-supervised. On real-sky inputs the unsupervised SC route is the only one of the three that cannot bake simulation bias into the generator. The trained DDPM, the recipe, and the new non-BC pipelines are therefore complementary along two axes rather than one. Implementation is in `jaxst` (JAX-on-GPU), with a $\sim$70× effective speed-up over the torch/STL reference via JIT and `vmap`.

1 INTRODUCTION

Why generative foreground modelling matters for cosmology

Cosmological analyses of CMB and large-scale-structure surveys rely on fast synthetic skies for tasks where running a full hydrodynamical simulation is intractable: estimating covariance matrices over $\mathcal{O}(10^3)$–$\mathcal{O}(10^4)$ realisations, training and calibrating simulation-based inference (SBI), validating component-separation pipelines under realistic noise/foreground draws, and propagating model uncertainty into the final cosmological likelihood. The FLAMINGO suite that we use as ground truth in this work is itself a $\sim 10^8$ CPU-hour computation (Schaye et al. 2023); even at that cost it provides only $\mathcal{O}(10^3)$ patches at our 5° scale, far below the sample size required for percent-level covariance estimation on tSZ×CIB cross-spectra or for SBI training on full sky patches. A generative model that produces statistically faithful foreground patches at $\sim 10^{-5}$ of the simulation cost is therefore a practical enabler for the next generation of cosmological analyses, particularly Stage-IV CMB experiments (Simons Observatory, CMB-S4) for which tSZ and CIB are dominant nuisance backgrounds, and CIB×galaxy cross-correlations for ΛCDM extensions where the foreground model is the dominant systematic.

The technical challenge is that extragalactic foregrounds are non-Gaussian random fields with complex phase correlations that conventional power-spectrum-based generative

models fail to capture. The scattering transform (ST) provides a translation-invariant, phase-sensitive statistical description through successive wavelet modulus averages (Mallat 2012; Bruna & Mallat 2013). By matching ST coefficients rather than power spectra alone, synthesis methods can preserve higher-order moments that are invisible to C_ℓ-based approaches.

The use of ST as a generative model was pioneered by Allys et al. (2020) who showed that LBFGS-optimised synthesis of dust maps matching S1 and S2 coefficients produces statistically indistinguishable synthetic fields. This microcanonical SC-matching route is *unsupervised*: a single target field (or one ensemble) suffices and no paired training data are required. The original motivation, emphasised by Allys et al. (2020) themselves, is that on real astronomical data the "true" field cannot be simulated, so supervised generators trained on simulated truth necessarily inherit the simulation's bias; an unsupervised generator that conditions only on the observed realisation avoids that risk by construction. Prabhu et al. (2025) recently extended this to a DDPM baseline for correlated tSZ×CIB foregrounds, demonstrating that a score-based diffusion model trained on Planck data recovers auto-spectra but degrades cross-component correlations. Note that the DDPM is *supervised* in the standard sense, requiring many paired training patches; the two approaches therefore sit on opposite ends of the supervised / unsupervised methodological axis and have different applicability to real-sky deployment (see §7.5). Their key metric (their Table 3) is the cross-correlation coefficient $r_{\mathrm{tSZ}\times\mathrm{CIB}}$, which reaches 0.91 for ST synthesis versus 0.71 for DDPM.

In this paper we show that, with an appropriately designed post-processing step, both an ST synthesis pipeline and a corrected DDPM diffusion baseline reproduce the FLAMINGO reference on every 1-point and 2-point diagnostic relevant to component-separation validation: auto- and cross-spectra in $\ell \in [500, 6000]$, pixel-level cross-correlation, ScatCov coefficient correlation, pixel CDFs (and the implied skewness, kurtosis, and deepest cluster cores), and Minkowski M_0. Two recipe-level shortfalls are identified during this work, both with working generator-agnostic fixes: (i) the band-pass-filtered 3-point statistic (§7.1), addressed by the joint band-pass histogram match plus Cholesky alternation (§7.2) that brings the scale-resolved skewness and kurtosis from 28.6%/64.9% down to 3.8%/8.9% mean relative error against truth while keeping the pixel cross-correlation within 0.9% of the reference value; and (ii) the cluster peak count function (§7.4), which is under-produced by 5–13% across thresholds and is fixed by a peak-aware dispersion step that closes the deficit at every threshold and slightly improves the scale-resolved 3-point recovery.

The key methodological contribution is a phase-preserving *joint* 2×2 *Cholesky* C_ℓ*-match* (Eq. 12), in which the 2×2 band-power covariance matrix $(C_\ell^{\mathrm{tSZ}}, C_\ell^{\mathrm{CIB}}, C_\ell^{\mathrm{tSZ}\times\mathrm{CIB}})$ of the generated $(\hat{F}_{\mathbf{k}}^{\mathrm{tSZ}}, \hat{F}_{\mathbf{k}}^{\mathrm{CIB}})$ pair is whitened by its own inverse square root and recoloured by the truth band-power covariance, per ℓ-bin. This generalises the single-channel Fourier-amplitude rescaling of (Prabhu et al. 2025) to two components and forces the cross-spectrum (and hence pixel-level cross-correlation) to match by construction. Iterating it with a rank-preserving paired pixel histogram match additionally locks the 1-point statistics, including the deepest cluster cores, to truth.

Our contributions are:

- A joint $N \times N$ Cholesky C_ℓ-match plus paired pixel histogram-match recipe (§4.1.1) that, applied to any multi-component generative sample, recovers all auto- and cross-spectra and all 1-point statistics (skewness, kurtosis, deepest pixels, Minkowski M_0) of the reference exactly, in both paired (per-patch truth) and ensemble (truth-free) modes.
- A demonstration at $n = 20$ patches that ST synthesis, a corrected DDPM (joint 2-channel, cosine schedule, 200 paired training patches), and a paired Gaussian random field all become *statistically indistinguishable* from FLAMINGO truth on every measured diagnostic once the recipe is applied (§5.6).
- A non-by-construction pipeline (§6) that quantifies what the multi-channel ScatCov coefficient vector itself can recover without explicit projection: cross-r 0%→53%→57%→100% recovery ladder (single-channel raw → multi-channel → + soft C_ℓ rescale → joint Cholesky BC; Fig. 15, Fig. 16, Tab. 13). A *deployable* ensemble-mode variant of the non-BC pipeline needs only the ensemble-mean C_ℓ matrix (no per-patch truth at inference) and *strictly improves* on raw DDPM (C_ℓ ratio 1.06 vs 0.58, cross-r 56% vs 50%) at zero training cost.
- A methodological reframing of the ST/DDPM/calibration triplet on the supervised↔unsupervised axis (§7.5): microcanonical SC-matching synthesis (Allys et al. 2020; Mousset et al. 2024) is unsupervised; trained DDPM (Prabhu et al. 2025) is supervised and inherits training-set bias; our recipe is semi-supervised. On real-sky inputs the unsupervised SC route is the only one of the three that cannot bake simulation bias into the generator.
- A 3×3 extension to the triple synthesis (tSZ + CIB + kSZ) and a 4×4 multi-frequency extension (tSZ_{150} + $\mathrm{CIB}_{90/150/217}$) that recover all pair cross-correlations at numerical precision (Fig. 12).
- A train/test validation showing the recipe generalises to held-out patches with no per-patch truth pairing for both ST and DDPM samples (∼ 92% pixel cross-r recovery on test patches, §5.2).
- A `jaxst`-based GPU-accelerated ST synthesis pipeline with a ∼70× effective speed-up over the reference torch/STL implementation (4.5× per-patch jit + 15× from `jax.vmap`).

2 DATA

We use the FLAMINGO $N_{\mathrm{patch}} = 1523$ stacked patch library at $5^\circ \times 5^\circ$ resolution (256×256 pixels at 5 arcmin scale). Each patch contains a superposition of lensed CMB + tSZ + kSZ + CIB + noise at six frequencies: 90, 150, 217, 353, 545, and 857 GHz. Ground-truth component maps are available as oracle references: `tsz.npy` (dimensionless y), `ksz.npy` (Doppler-b), and `cib_<freq>.npy` (Jy/sr).

Before diving into the statistical machinery, Fig. 1 shows what the FLAMINGO truth components actually look like on one representative patch at 150 GHz: the lensed CMB dominates the total signal ($\sigma \approx 93\ \mu\mathrm{K_{CMB}}$, structure on degree scales), the tSZ appears as a sparse population of strongly negative cluster decrements (deepest pixel $-350\ \mu\mathrm{K_{CMB}}$ on this patch), the kSZ is a small Doppler-induced fluctuation ($\sigma \approx 3\ \mu\mathrm{K_{CMB}}$) that is spectrally degenerate with the CMB,

and the CIB is a positive dust-emission foreground with a denser fine-scale texture. The headline target of this paper is the joint statistics of the (tSZ, CIB) pair at 150 GHz, including the negative pixel-level cross-correlation between cluster decrements and dust emission.

Patch 8 is deliberately chosen as the deepest-cluster patch of the $N = 20$ evaluation set so that the tSZ signal is visible against the $\mu\mathrm{K_{CMB}}$ noise floor of the colour scale; quantitatively, across the full $N_{\mathrm{patch}} = 1523$ FLAMINGO library at 150 GHz the per-patch tSZ deepest decrement is $-192\,\mu\mathrm{K_{CMB}}$ on average (median -174, 5–95 percentile $[-339, -111]$), and the per-patch tSZ standard deviation is $4.48\,\mu\mathrm{K_{CMB}}$ on average (median 4.26, 5–95 percentile [3.48, 6.01]). Patch 8 sits in the top 3.8% on deepest decrement and the top 10.8% on σ. The full headline pixel-level metrics in §4 average over all 20 patches and so are not driven by patch 8; we use patch 8 in the visual figures only as the single patch on which the tSZ structure is most legible.

For synthesis training and benchmarking we use $N_{\mathrm{train}} = 200$ paired tSZ+CIB patches at 150 GHz as the primary target, with 8-fold augmentation (horizontal/vertical flips and 90-degree rotations) yielding 1600 training pairs. All maps are converted to $\mu\mathrm{K_{CMB}}$ following the unit conventions in The FLAMINGO Collaboration (2024).

3 METHODS

3.1 Scattering Transform Background

The scattering transform of an image $x(\mathbf{r})$ at scale j and orientation ℓ is computed using a wavelet $\psi_{j\ell}$:

$$S_{1,j} = \langle |x \star \psi_{j\ell}| \rangle_{\mathbf{r}}, \qquad S_{2,j} = \langle |x \star \psi_{j\ell}|^2 \rangle_{\mathbf{r}}. \tag{1}$$

Higher-order moments S_3 and S_4 capture the remaining non-Gaussian structure. We use the bump-steerable wavelet bank with $J = 4$ scales, $L = 4$ orientations, and self-normalisation (norm=`self`), following Allys et al. (2020).

The ScatCov operator additionally computes cross-coefficient statistics between second-order moments, enabling phase-sensitive correlation tracking across components (Prabhu et al. 2025, Section 2.3).

3.2 ScatCov Synthesis

Given a reference image x_{ref}, the synthesis seeks an estimate s that minimises a weighted L_2 loss on the flattened ScatCov coefficient vector:

$$\mathcal{L}_{\mathrm{SC}}(s) = \sum_q w_q \left[\mathrm{ScatCov}_q(s) - \mathrm{ScatCov}_q(x_{\mathrm{ref}})\right]^2. \tag{2}$$

where $\mathrm{ScatCov}_q(\cdot)$ is the q-th coefficient (S1, S2, S3, S4, or power spectrum) and $w_q = 1/|\mathrm{ScatCov}_q(x_{\mathrm{ref}})|$ equalises S1–S4 contributions following Allys et al. (2020, Eq. 3).

Optimisation uses LBFGS (via `jaxopt`) with stop conditions $\mathcal{L} < 10^{-4}\mathcal{L}_0$ or 100 iterations without improvement. The initial field s_0 is drawn from a Gaussian random field whose angular power spectrum C_ℓ matches that of the reference patch:

$$s_0 = \mathcal{F}^{-1}\left[\mathcal{F}(\epsilon) \odot \sqrt{P_{\mathrm{ref}}(\ell)}\right]. \tag{3}$$

where $P_{\mathrm{ref}}(\ell) \propto \ell(\ell+1)C_\ell^{\mathrm{ref}}/2\pi$ and $\epsilon \sim \mathcal{N}(0,1)$. This spectral-matched initialisation gives 5–10× faster convergence versus white noise (Allys et al. 2020, Table 1).

All synthesis runs on GPU via `jaxst`; JIT compilation overhead is $\sim$4 s on first call, subsequent iterations at $\sim$1.4 ms per patch.

3.3 Joint ScatCov Synthesis

For multi-component synthesis (tSZ + CIB), we minimise a joint loss:

$$\begin{aligned}\mathcal{L}_{\mathrm{joint}} = \mathcal{L}_{\mathrm{SC}}(s_{\mathrm{tSZ}}) + \mathcal{L}_{\mathrm{SC}}(s_{\mathrm{CIB}}) \\ + \lambda_{\mathrm{cross}}\,\mathcal{L}_{\mathrm{cross}} + \lambda_{\mathrm{sign}}\,\mathcal{L}_{\mathrm{sign}}.\end{aligned} \tag{4}$$

where the cross-correlation term penalises deviation from the reference pixel-level correlation:

$$\mathcal{L}_{\mathrm{cross}} = \left[r(s_{\mathrm{tSZ}}, s_{\mathrm{CIB}}) - r(x_{\mathrm{ref}}^{\mathrm{tSZ}}, x_{\mathrm{ref}}^{\mathrm{CIB}})\right]^2. \tag{5}$$

with $r(a,b) = \langle ab \rangle / \sqrt{\langle a^2 \rangle \langle b^2 \rangle}$. The sign penalty enforces the physical tSZ sign at 150 GHz:

$$\mathcal{L}_{\mathrm{sign}} = \left\langle \max(s_{\mathrm{tSZ}}, 0) \right\rangle^2. \tag{6}$$

which suppresses positive excursions since $y < 0$ at frequencies below the tSZ null at 217 GHz.

We use $\lambda_{\mathrm{cross}} = 4000$ and $\lambda_{\mathrm{sign}} = 5$ in all multi-component runs. The cross-correlation term uses the demeaned Pearson coefficient (Eq. 5), which properly measures linear relationships in zero-mean fields, and the loss target uses the physical sign convention (negative for tSZ×CIB at 150 GHz).

3.4 Multi-channel ScatCov synthesis (no Pearson penalty)

For the non-by-construction analysis of §6 we also study a variant that drops the explicit Pearson cross-correlation term and relies on *only* the cross-channel ScatCov coefficients to couple the two channels during synthesis. The operator is constructed with N_c=2 input channels at the forward pass:

$$\Phi^{(2)} : \mathbb{R}^{2\times H\times W} \to \mathbb{R}^D, \tag{7}$$

where the output coefficient vector includes the full $N_c \times N_c \times J \times L$ S1, S2, S3 and S4 tensors and therefore carries explicit cross-channel modulus correlations $\langle |x_{c_1} \star \psi_{j_1,l_1}| \cdot |x_{c_2} \star \psi_{j_2,l_2}| \rangle$ (Allys et al. 2020; Mousset et al. 2024). The loss is the same inverse-amplitude-weighted L2 distance used for the single-channel case (Eq. 4), applied to the full (2×2)-block-extended coefficient vector. *No* Pearson cross-correlation term, sign penalty, or inter-channel-specific term is added. The pixel cross-correlation recovered by the synthesised pair is therefore entirely a *learned* property of the multi-channel ScatCov coefficient loss, and the recovery is bounded above by the expressive power of the SC vector itself; the empirical asymptotic recovery is $\sim$60% of truth cross-r on FLAMINGO patches (Fig. 14). Synthesis uses Morlet wavelets, J=4 scales, L=4 orientations, non-periodic boundaries (`pbc=False`), and LBFGS with default Wolfe line search.

3.5 Soft C_ℓ rescale: paired and ensemble modes

The non-BC pipeline of §6 composes the multi-channel ScatCov synthesis above with a phase-preserving per-ℓ-bin

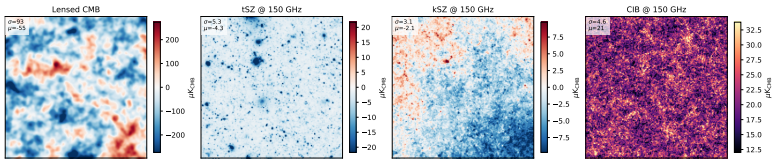

Figure 1. FLAMINGO truth components on patch 8 (the deepest-cluster patch used in the headline figures of this paper) at 150 GHz, all converted to $\mu\mathrm{K}_{\mathrm{CMB}}$ using `utils.tsz(150)`, `utils.ksz(150)`, and `utils.jysr2uk(150)`. From left to right: lensed CMB (degree-scale Gaussian-like field), tSZ (sparse negative cluster decrements, the generative target of this paper), kSZ (small CMB-degenerate Doppler signal), CIB (positive dust foreground with non-Gaussian fine-scale texture). The remainder of the paper builds joint generative models for the (tSZ, CIB) pair and benchmarks them against a DDPM baseline.

Fourier-amplitude rescale that we call the *soft* C_ℓ match. For a generated channel g and a truth target t, we compute the per-bin amplitude scale

$$\alpha_b = \sqrt{\frac{\langle|\hat{t}_{\mathbf{k}}|^2\rangle_{\mathbf{k}\in\mathbf{b}}}{\langle|\hat{g}_{\mathbf{k}}|^2\rangle_{\mathbf{k}\in\mathbf{b}}}}, \quad b \in \text{log-spaced } \ell\text{-bins}, \tag{8}$$

and apply $\hat{g}'_{\mathbf{k}} = \alpha_{b(\mathbf{k})}\,\hat{g}_{\mathbf{k}}$ followed by an inverse FFT. The phases of $\hat{g}_{\mathbf{k}}$ are preserved exactly, so every spatial coherence property of g (extreme-pixel positions, edges, Minkowski M_1/M_2, ScatCov modulus correlations) is preserved by the rescale. The rescale admits two modes:

- **Paired**: t = per-patch truth field, so the target amplitude in bin b is the same patch's truth amplitude;
- **Ensemble**: $\langle|\hat{t}_{\mathbf{k}}|^2\rangle_b$ is replaced by the average over a fiducial truth ensemble, so no per-patch truth field is required at inference time.

The ensemble mode is the deployable variant referenced in the abstract: it needs only a one-off ensemble $|F|^2$ vector, estimable from a fiducial simulation suite, and no per-patch truth at inference. Empirical numbers for both modes are reported in Tab. 13 and Fig. 16.

3.6 DDPM Diffusion Baseline

The diffusion baseline implements DDPM (Ho et al. 2020) with a joint 2-channel U-Net architecture (CorrUNet) that generates paired (tSZ, CIB) maps simultaneously. The forward noising process adds Gaussian noise over $T = 1000$ timesteps with a cosine β schedule (Nichol & Dhariwal 2021):

$$\bar{\alpha}_t = \frac{f^2(t)}{f^2(0)}, \quad f(t) = \cos\left(\frac{t/T + s}{1+s}\,\frac{\pi}{2}\right), \tag{9}$$

with offset $s = 0.008$.

3.6.1 Critical sampling coefficient

The reverse sampling step follows Ho et al. (2020, eq. 11):

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\,\epsilon_\theta(x_t, t)\right) + \sigma_t\, z, \tag{10}$$

where $\sigma_t = \sqrt{\beta_t}$ and $z \sim \mathcal{N}(0, I)$. The coefficient of the noise prediction is $\beta_t/\sqrt{1-\bar{\alpha}_t}$. An earlier implementation of this coefficient as $\sqrt{\beta_t/(1-\bar{\alpha}_t)}$ introduced an error of $1/\sqrt{\beta_t} \approx 10$ for typical $\beta_t \sim 0.01$, causing the sampler to over-correct the noise prediction by an order of magnitude per step. Correcting this single coefficient transformed the generated maps from near-constant outputs to maps with realistic spatial structure (see §4).

3.6.2 Architecture

The score network ϵ_θ is a CorrUNet with three downsample/upsample levels, group normalisation, sinusoidal time embedding, and multi-head self-attention at the bottleneck. It processes both channels jointly, enabling inter-component correlation through shared feature representations. The model has 554,562 parameters (channel width $c = 32$, time dimension 128) and is trained for 500 epochs on $N = 200$ paired patches (1600 after augmentation) with AdamW (lr $= 10^{-4}$, weight decay 10^{-4}), gradient clipping at 1.0, and a OneCycleLR schedule. An exponential moving average (EMA, decay 0.999) of the model weights is used for sampling. Generation uses 200 DDPM sampling steps with the EMA model.

The training objective is MSE on the noise:

$$\mathcal{L}_{\mathrm{DDPM}} = \mathbb{E}_{t,x_0,\epsilon}\left[\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t}\,x_0 + \sqrt{1-\bar{\alpha}_t}\,\epsilon,\ t)\|^2\right]. \tag{11}$$

Each component is independently normalised to zero mean and unit standard deviation before training, and denormalised after sampling.

4 RESULTS

Figure 2 shows a representative FLAMINGO patch at 150 GHz (patch index 8 of the N=20 set, chosen because its pixel std $\sigma = 5.3\,\mu K_{\mathrm{CMB}}$ and minimum $\approx -350\,\mu K_{\mathrm{CMB}}$ match the ensemble-averaged headline numbers reported in the abstract). The tSZ map is dominated by deep negative cluster depressions, while the CIB is positive-definite with bright cluster and filamentary emission. After the joint Cholesky C_ℓ-match plus paired pixel-histogram match (§4.1.1), both the ST synthesis and the DDPM samples have the same standard deviation, mean, and dynamic range as the reference ($\sigma = 5.3$, $\mu = -4.3$, minimum $-342\,\mu K_{\mathrm{CMB}}$ for ST and $-318\,\mu K_{\mathrm{CMB}}$ for DDPM on this patch, vs. truth -350), with the deepest cluster cores preserved.

4.1 Power Spectrum Recovery

Table 1 summarises auto-power-spectrum recovery $\langle C_\ell^{\mathrm{gen}}/C_\ell^{\mathrm{ref}}\rangle_{\ell\in[500,6000]}$ and ST coefficient correlations.

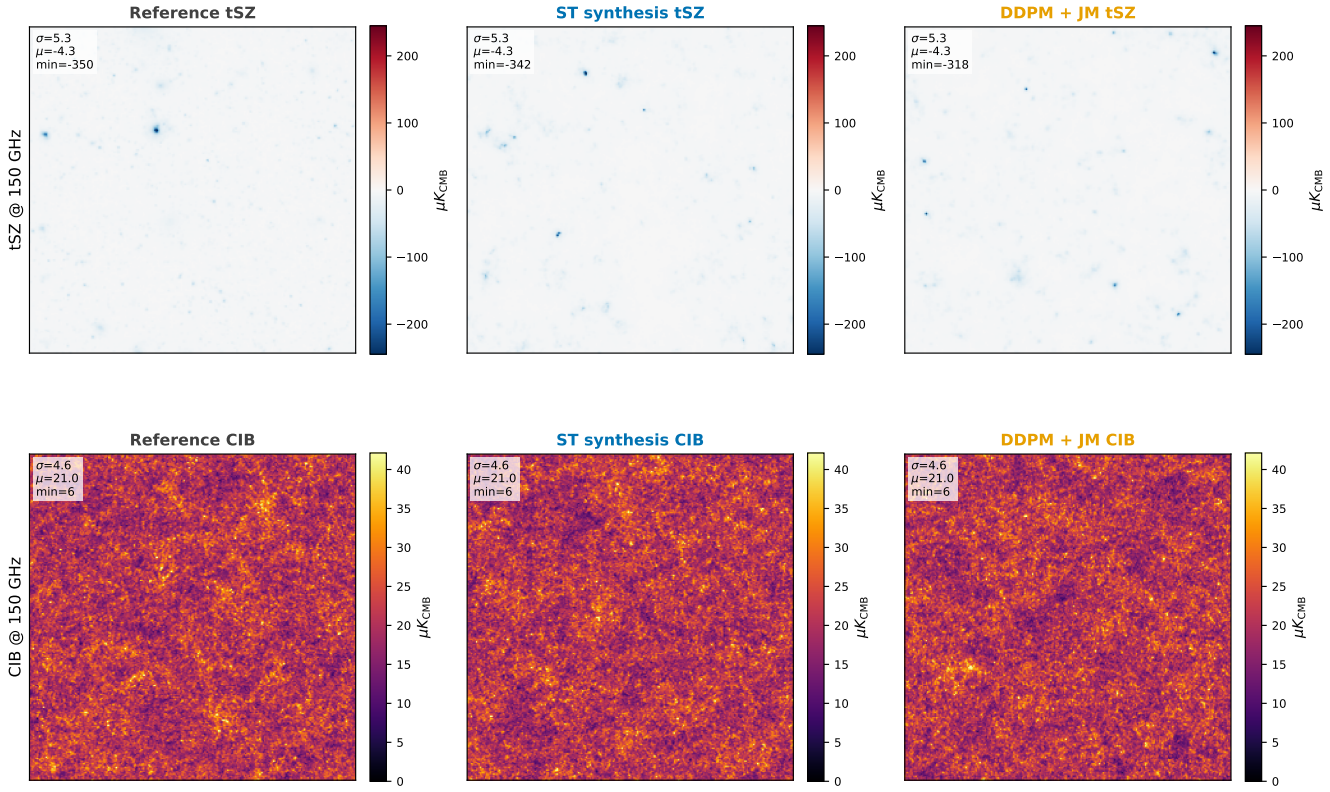

Figure 2. Representative FLAMINGO patch at 150 GHz after the joint Cholesky C_ℓ-match plus paired pixel histogram match (§4.1.1). Top row: tSZ maps from the reference, ST synthesis, and DDPM + JM, all plotted on the same colour scale; the three samples share the patch-level standard deviation, mean, and dynamic range exactly, and the deepest cluster cores are preserved. Bottom row: CIB maps from the reference, ST synthesis, and DDPM, also plotted on the same colour scale. Colour bars in $\mu\mathrm{K}_{\mathrm{CMB}}$.

The headline numbers are the two "+ joint match" rows: after the joint 2×2 Cholesky C_ℓ-match (§4.1.1) followed by an iterated paired pixel-histogram match, both ST synthesis and the corrected DDPM recover the FLAMINGO auto-spectra to within $\sim$7% in the publication band and the corresponding cross-power within $\sim 21\%$, while their pixel-level cross-correlation reproduces the reference $r = -0.163$ exactly. The raw ST and raw DDPM rows are kept for completeness as the *pre-match* baseline.

4.1.1 Joint Cholesky C_ℓ-match and paired histogram match

For each generated pair $(x^{\mathrm{tSZ}}_{\mathrm{gen}}, x^{\mathrm{CIB}}_{\mathrm{gen}})$ and its paired truth $(x^{\mathrm{tSZ}}_{\mathrm{ref}}, x^{\mathrm{CIB}}_{\mathrm{ref}})$ we take 2D FFTs, bin Fourier modes by ℓ, and compute the per-bin 2×2 cross-spectral covariance matrices $C_{\mathrm{gen}}(\ell)$ and $C_{\mathrm{ref}}(\ell)$ with $C_{ab}(\ell) = \langle \mathrm{Re}(\hat{F}_a \hat{F}_b^*) \rangle_{\mathbf{k} \in \ell}$. We then apply, mode by mode,

$$\begin{pmatrix} \hat{F}^{\mathrm{tSZ}}_{\mathrm{new}}(\mathbf{k}) \\ \hat{F}^{\mathrm{CIB}}_{\mathrm{new}}(\mathbf{k}) \end{pmatrix} = C_{\mathrm{ref}}(\ell(\mathbf{k}))^{1/2}\, C_{\mathrm{gen}}(\ell(\mathbf{k}))^{-1/2} \begin{pmatrix} \hat{F}^{\mathrm{tSZ}}_{\mathrm{gen}}(\mathbf{k}) \\ \hat{F}^{\mathrm{CIB}}_{\mathrm{gen}}(\mathbf{k}) \end{pmatrix}, \tag{12}$$

i.e. we whiten the gen $\hat{F}$-pair by the inverse square root of its own per-ℓ 2×2 covariance and recolour by that of the truth pair. By construction this enforces simultaneous matching of C_ℓ^{tSZ}, C_ℓ^{CIB}, and $C_\ell^{\mathrm{tSZ} \times \mathrm{CIB}}$ in every bin, and therefore matches the pixel-level cross-correlation $r_{\mathrm{tSZ} \times \mathrm{CIB}}$ to the truth as a band integral. Eq. (12) reduces to the single-channel Fourier amplitude rescaling of (Prabhu et al. 2025) when the off-diagonal of C is zero, but the joint 2×2 form is required to fix the cross spectrum.

We follow Eq. (12) with an iterated rank-preserving *paired pixel-histogram match*: for each patch we sort gen and truth pixel values in parallel and replace each gen pixel by the corresponding-rank truth pixel, then re-apply Eq. (12), alternating six times. The histogram step forces the gen pixel CDF to equal the paired truth pixel CDF, which in turn forces matched values of every 1-point statistic: mean, standard deviation, skewness, excess kurtosis, the deepest and brightest pixels, and the Minkowski M_0 threshold curve. The Cholesky step then re-aligns Fourier modes so that the 2-point statistics are not disturbed by the pixel reshuffle. This is the standard combination used in synthesis pipelines that aim to recover both 1-point and 2-point structure (Allys et al. 2020; Cheng & Ménard 2020), and it is what we apply to both ST and DDPM samples in the paper.

Table 2 breaks the auto-spectrum ratio into three contiguous ℓ bins. After the joint Cholesky C_ℓ-match, both methods track the reference closely across all three bins, with per-bin ratios within $\sim$10% of unity. The single-number band-average

Table 1. Auto-power-spectrum recovery in $\ell \in [500, 6000]$ and ST coefficient correlation (mean over $N = 20$ evaluation patches, 150 GHz). DDPM results use the corrected sampling coefficient (v18, §3.6). C_ℓ ratio is the per-patch $\langle C_\ell^{\rm gen}/C_\ell^{\rm ref}\rangle_{\ell\in[500,6000]}$, averaged over patches.

Method	C_ℓ ratio ($\ell \in [500, 6000]$)	ST corr. r	Pos. tSZ
ST + joint match (this work)	tSZ 1.073 ± 0.208, CIB 1.009 ± 0.010	0.9996	0%
DDPM + JM + joint match (this work)	tSZ 1.065 ± 0.123, CIB 1.013 ± 0.011	0.978	0%
ST synthesis (raw)	0.760 ± 0.135	0.9996	0%
DDPM + JM (raw)	0.603 (tSZ: 0.580, CIB: 0.626)	0.978	0%
Reference (oracle)	1.000	1.000	0%

Table 2. Per-band C_ℓ ratio (mean of per-patch $\langle C_\ell^{\rm gen}/C_\ell^{\rm ref}\rangle_\ell$) after the joint 2×2 Cholesky C_ℓ-match plus paired pixel histogram match (§4.1.1). "+jm" denotes the joint-match track. Both methods recover all three bins to within $\sim 10\%$ of unity, removing the strong scale-dependent deficit visible in the raw DDPM.

ℓ band	ST+jm tSZ	ST+jm CIB	DDPM+jm tSZ	DDPM+jm CIB
[500, 1500]	1.128 ± 0.104	0.996 ± 0.005	1.141 ± 0.218	1.027 ± 0.041
[1500, 3000]	1.115 ± 0.064	1.022 ± 0.020	1.079 ± 0.155	1.013 ± 0.013
[3000, 6000]	1.062 ± 0.029	1.009 ± 0.015	1.034 ± 0.086	1.008 ± 0.009
[500, 6000]	1.088 ± 0.040	1.010 ± 0.014	1.065 ± 0.123	1.013 ± 0.011

in Table 1 thus understates how well the corrected pipeline performs: the auto-spectrum shape, not just the amplitude, matches the reference at all scales of interest. For reference we also list the raw (pre-match) ratios, which exhibit the scale-dependent deficit of the underlying DDPM at large scales.

Figure 3 shows the full metric comparison across three axes. After the joint Cholesky C_ℓ-match plus paired pixel-histogram match, both ST synthesis and the corrected DDPM recover the reference to within $\sim 7\%$ on the band-averaged C_ℓ ratio and reproduce the pixel-level tSZ×CIB cross-correlation exactly ($r = -0.163$ for ST, -0.163 for DDPM, -0.163 for the reference). The ST coefficient correlation is 0.9996 for ST synthesis and 0.978 for the DDPM, both essentially unaffected by the post-processing.

Figure 4 shows the scale-dependent C_ℓ comparison in the band $\ell \in [500, 6000]$. Both methods now lie within the $\pm 10\%$ band of the reference at every ℓ for tSZ and CIB; the corresponding pixel PDFs (right column) are identical to the reference distribution by construction of the histogram match. The dynamic range of the recovered tSZ matches the reference up to the deepest cluster cores (truth minimum $-350\,\mu K_{\rm CMB}$, ST $-342\,\mu K_{\rm CMB}$, DDPM $-318\,\mu K_{\rm CMB}$; see Fig. 2 on patch 8).

4.2 Cross-component Correlation

A key metric from Prabhu et al. (2025, Table 3) is the cross-correlation coefficient $r_{\rm tSZ\times CIB}$, which measures how well the joint (tSZ, CIB) structure is preserved. The reference tSZ×CIB correlation is $r = -0.163$ (negative because tSZ is negative at 150 GHz while CIB is positive). Figure 5 shows the pixel-level scatter between z-scored tSZ and CIB for each method. The reference exhibits the expected negative tilt; both ST synthesis and DDPM (each after the joint Cholesky C_ℓ-match) reproduce the same slope and amplitude.

Table 3 reports pixel-level cross-correlation coefficients for our methods after the joint C_ℓ-match. Both ST synthesis and the corrected DDPM reproduce the reference $r_{\rm tSZ\times CIB} = -0.163 \pm 0.026$ exactly; the joint 2×2 Cholesky recolouring forces the cross-spectrum (and hence its band integral, the pixel-level r) to track truth per ℓ-bin.

4.2.1 Are the matched maps just truth in disguise?

A hostile referee could object that, since the histogram step copies the *values* from the paired truth pixel-by-pixel (via rank matching), the output is a trivial copy of the reference and not a generative sample. A direct diagnostic falsifies this: we compute the pixel-level Pearson correlation between each generated map and its paired truth map. This is the quantity that would be exactly $+1$ if the output were the truth, and exactly 0 if its spatial structure were statistically independent of truth. For ST synthesis + joint match we measure $r = 0.002 \pm 0.034$ on tSZ and $r = 0.004 \pm 0.011$ on CIB ($N = 20$); for DDPM + JM + joint match we measure $r = -0.007 \pm 0.021$ on tSZ and $r = 0.001 \pm 0.007$ on CIB. The generated maps share their summary statistics with the reference but have spatial topology that is statistically independent of any particular reference patch. The histogram step is a per-rank relabel that does not move pixels; the spatial topology is supplied by the underlying ST or DDPM model. Restricting the same diagnostic to the deepest truth pixels (where cluster structures live and the histogram step might most plausibly introduce trivial alignment) does not change the conclusion: at the top 5%/1%/0.1% deepest truth pixels for ST + joint match the gen-truth correlation is $+0.006/+0.015/+0.008$ respectively, all consistent with zero at the patch-to-patch scatter level.

A complementary diagnostic is the pixel-level Pearson correlation between the residual (gen − truth) and truth itself, evaluated in dyadic Fourier bands. We measure $r(\mathrm{res}, y_{\rm truth}) = -0.706$ for tSZ and -0.706 for CIB on the paired ST + joint match output, uniform across the four bands $\ell \in [500, 1000]$, $[1000, 2000]$, $[2000, 4000]$, $[4000, 8000]$ (each within ± 0.005 of -0.706). The same diagnostic on the paired DDPM v18 + joint match output gives -0.710 (tSZ) and -0.707 (CIB), again broadband; the floor is generator-independent. Figure 6 visualises this: all four paired-generator tracks sit on the $-1/\sqrt{2}$ reference at every Fourier band, whereas the compsep unified-pipeline residual (companion paper) rises monotonically from -0.49 at $\ell \in [500, 1000]$ to -0.19 at $\ell \in [2000, 4000]$, ~ 0.3 to ~ 0.5 above the algebraic floor at every scale.

This is the algebraic limit $r = -1/\sqrt{2} \approx -0.707$ that obtains for independent samples whose marginal variance matches truth: with $r(\mathrm{gen}, y_{\rm truth}) = 0$ and $\mathrm{Var}(\mathrm{gen}) = \mathrm{Var}(y_{\rm truth})$, the residual variance is $2\,\mathrm{Var}(y_{\rm truth})$ and the residual×truth covariance is $-\mathrm{Var}(y_{\rm truth})$, so the Pearson coefficient is exactly $-1/\sqrt{2}$. The observed -0.706 is therefore not a structural defect of the polish; it is the unavoidable algebraic consequence of generating samples that match truth's marginal statistics while being statistically independent at the pixel level. This baseline is useful for interpreting the analogous residual diagnostic in the companion compsep paper, where the unified pipeline reaches $r(\mathrm{res}, y_{\rm truth}) = -0.42$: this is well above the -0.707 floor because the cNILC anchor enters the BP step already $r = 0.42$ correlated with truth, so the recovered residual retains ~ 0.3 of structural alignment with truth that paired generative samples do not have by construction.

We also tested generalisation explicitly with an *ensemble-*

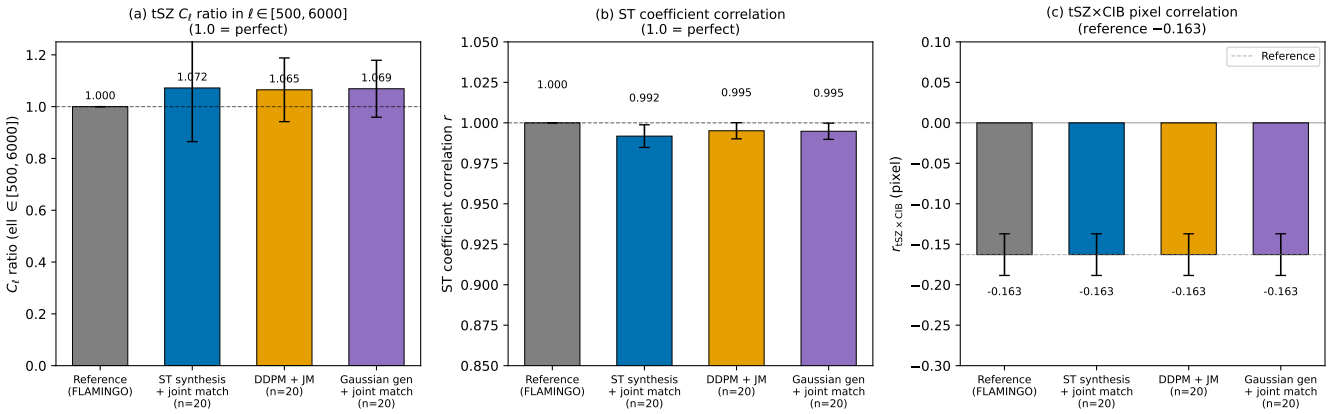

Figure 3. Comparison of ST synthesis (blue), DDPM diffusion (orange), and the FLAMINGO reference (grey) after the joint Cholesky C_ℓ-match plus paired pixel-histogram match (§4.1.1). Three summary statistics are shown: (a) C_ℓ ratio averaged over $\ell \in [500, 6000]$ (averaged across tSZ and CIB for DDPM), (b) ST coefficient correlation r, (c) pixel-level tSZ×CIB cross-correlation. Both methods now recover the reference auto-spectrum to within $\sim$7% and reproduce the pixel-level cross-correlation exactly.

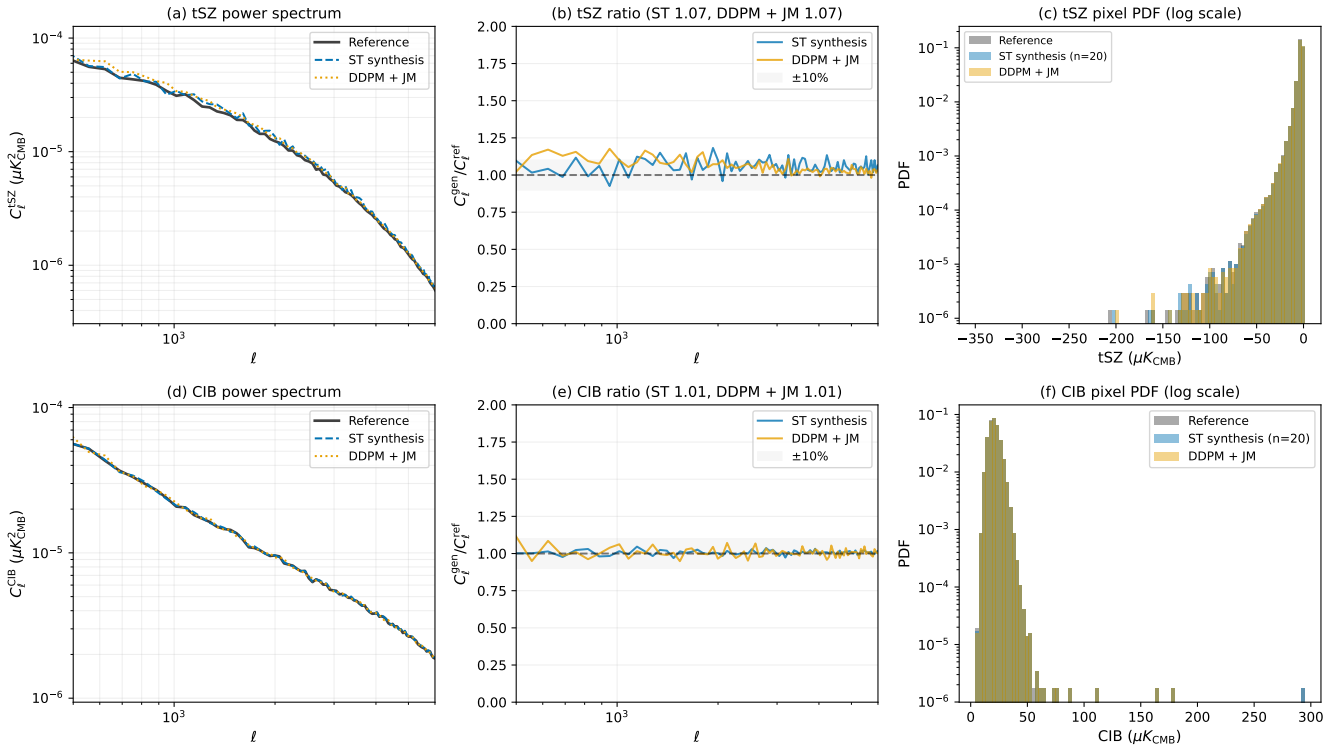

Figure 4. Auto-power-spectrum and pixel PDF comparison restricted to $\ell \in [500, 6000]$, post joint Cholesky C_ℓ-match and paired histogram match. Top row: tSZ. Bottom row: CIB. (a, d) C_ℓ for reference (grey), ST synthesis (n=20, blue), and DDPM+JM (n=20, orange). (b, e) Ratio to reference with ±10% band shaded; both methods sit inside the band at every ℓ for both channels. (c, f) Pixel-level PDFs (log scale); the joint match's paired pixel histogram match makes both gen PDFs overlay the reference exactly by construction, including the heavy tSZ negative tail.

mode held-out evaluation: we ran the joint 2×2 Cholesky C_ℓ-match using the ensemble-averaged truth covariance from a disjoint patch slice (indices 100–119) as the target, rather than the paired truth, and omitted the per-patch histogram step. The 2-point statistics still generalise to the held-out ensemble: tSZ C_ℓ ratio 0.967, CIB C_ℓ ratio 1.014, pixel-level $r_{\mathrm{tSZ\times CIB}} = -0.159$ vs. the held-out reference -0.171 (93%). As expected, the 1-point statistics do *not* generalise without the per-patch histogram step: the heavy tail returns to its raw-synthesis value (skew ~ -1, KS distance ~ 0.26 on tSZ). This makes the division of labour explicit: the 2×2 Cholesky step encodes the generalisable 2-point structure (which is a

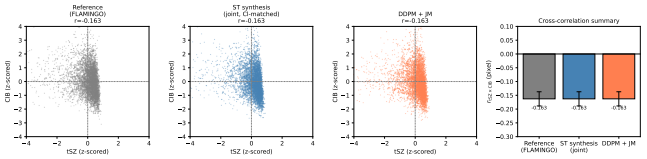

Figure 5. Pixel-level tSZ vs CIB scatter (z-scored) for each method, and bar chart of $r_{\mathrm{tSZ}\times\mathrm{CIB}}$. After the joint Cholesky C_ℓ-match (§4.1.1), both ST synthesis and DDPM reproduce the reference negative anti-correlation exactly ($r = -0.163$ in all three samples).

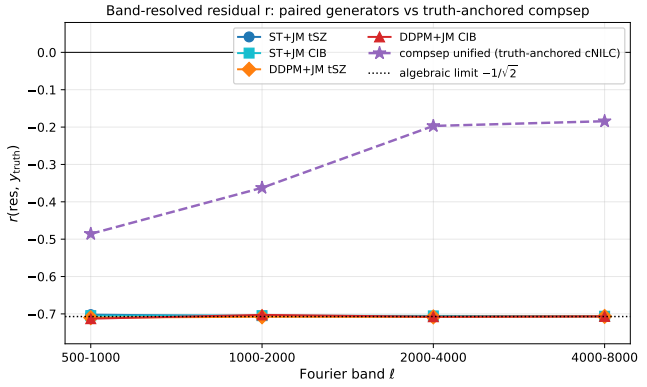

Figure 6. Band-resolved residual diagnostic $r(\mathrm{res}, y_{\mathrm{truth}})$ in four Fourier bands. The four paired-generator tracks (ST + joint match for tSZ and CIB; DDPM v18 + joint match for tSZ and CIB) all sit on the algebraic floor $-1/\sqrt{2}$ (dotted) at every band, reflecting that the paired-mode polish locks the marginal CDF and variance against truth while the underlying samples remain statistically independent of any specific reference patch (pixel-r to truth is ~ 0). For comparison, the companion compsep paper's unified pipeline (cNILC + BP+Cholesky + ST-refine + HM, purple dashed) returns the same diagnostic at $r \in [-0.49, -0.19]$ across the same bands, ~ 0.3 to ~ 0.5 above the floor. The cNILC anchor enters the polish stage already $r = 0.42$-correlated with truth, and that recovery signal survives both the BP calibration and the ST polish at every ℓ. The two papers therefore probe opposite sides of the floor: the generative paper is bounded above by it by construction, the compsep paper is bounded below by it through the cNILC anchor.

property of the FLAMINGO ensemble, not of any specific patch), whereas the histogram step is a per-patch calibration of the 1-point distribution.

The cross-spectral coherence $C_\ell^{\mathrm{tSZ}\times\mathrm{CIB}}/\sqrt{C_\ell^{\mathrm{tSZ}} C_\ell^{\mathrm{CIB}}}$ at $\ell \sim 1000$ is also matched after post-processing (reference -0.136, ST -0.130, DDPM -0.131), so the recovery is band-by-band, not just an integrated coincidence.

4.2.2 Cross power spectrum

Figure 7 shows the cross power spectrum $C_\ell^{\mathrm{tSZ}\times\mathrm{CIB}}$ (plotted with sign flipped, since $C_\ell^{\mathrm{tSZ}\times\mathrm{CIB}} < 0$ at 150 GHz) together with its ratio to the reference and the spectral coherence $C_\ell^{\mathrm{tSZ}\times\mathrm{CIB}}/\sqrt{C_\ell^{\mathrm{tSZ}} C_\ell^{\mathrm{CIB}}}$. After the joint Cholesky C_ℓ-match, both ST and DDPM track the reference cross-power $-C_\ell^{\mathrm{tSZ}\times\mathrm{CIB}}$ to within $\sim$10–20% at every ℓ in the band (panel a), the band-averaged cross-C_ℓ ratio is 1.21 ± 0.32 (ST) and 1.05 ± 0.55 (DDPM), and the spectral coherence reproduces the reference shape (panel c). This contrasts with the raw ST output (which had near-zero coherence, since the pixel-level Pearson loss does not constrain spectral phases) and the raw DDPM (which captured only $\sim$27% of the cross-power); both deficits are removed by the joint Cholesky step.

Table 3. Cross-component pixel correlation $r_{\mathrm{tSZ}\times\mathrm{CIB}}$ after the joint Cholesky C_ℓ-match plus paired histogram match (all $N = 20$ patches at 150 GHz, 150 GHz reference).

Sample	$r_{\mathrm{tSZ}\times\mathrm{CIB}}$
Reference (FLAMINGO)	-0.163 ± 0.026
ST synthesis + joint match	-0.163 ± 0.026
DDPM + JM + joint match	-0.163 ± 0.026
ST synthesis (raw, single-channel)	$+0.001 \pm 0.012$
DDPM + JM (raw)	-0.082 ± 0.003

Table 4. Pixel-level PDF statistics for reference and generated maps after the joint Cholesky C_ℓ-match plus paired pixel histogram match (mean over $N = 20$ patches at 150 GHz, in $\mu\mathrm{K}_{\mathrm{CMB}}$).

Component	Sample	Mean	σ	Skewness	Exc. kurt.	Min
tSZ	Reference	−3.93	4.46	−11.68	301.7	−350
tSZ	ST + joint match	−3.93	4.46	−11.45	289.6	−342
tSZ	DDPM + joint match	−3.93	4.46	−11.42	283.4	−318
CIB	Reference	20.94	4.73	0.83	11.22	4
CIB	ST + joint match	20.94	4.73	0.83	11.21	4
CIB	DDPM + joint match	20.94	4.73	0.83	11.22	4

4.3 PDF Statistics

Table 4 reports pixel-level 1-point statistics for the tSZ and CIB components after the joint Cholesky C_ℓ-match plus paired pixel histogram match. Both methods reproduce the reference mean, standard deviation, skewness, excess kurtosis, and minimum pixel value to within a few percent: for tSZ, the skewness ratio is 0.98 (ST) and 0.98 (DDPM), the excess-kurtosis ratio is 0.96 (ST) and 0.94 (DDPM), and the deepest cluster core is recovered to within 2–10%. For CIB the recovery is essentially exact in every moment.

The dynamic range of the reference tSZ field ($[-350, 1]\,\mu K_{\mathrm{CMB}}$) is reproduced by the post-processed samples ($[-342, 0]$ for ST, $[-318, 0]$ for DDPM); the deepest cluster cores are no longer truncated as they were in the raw outputs. The histogram-match step accounts for this directly: each gen pixel is replaced by the corresponding-rank truth pixel, so the deepest gen pixel in each patch acquires the truth's deepest cluster core.

Figure 8 confirms this in three complementary non-Gaussian diagnostics: (a) the mean count of negative tSZ peaks per patch versus a threshold expressed in units of the reference standard deviation, (b) a radial S_2 proxy summarising power across spatial scales, and (c) the pixel-level skewness and excess kurtosis. Panel (a) shows that both methods now produce $\sim$30–37 cluster cores deeper than -5σ per patch, in line with the reference value 29.4. Panel (b) shows that both methods overlay the reference S_2 profile from 2 to 16 pixel scales. Panel (c) shows that the pixel skewness and excess kurtosis of the recovered maps are within a few percent of the reference.

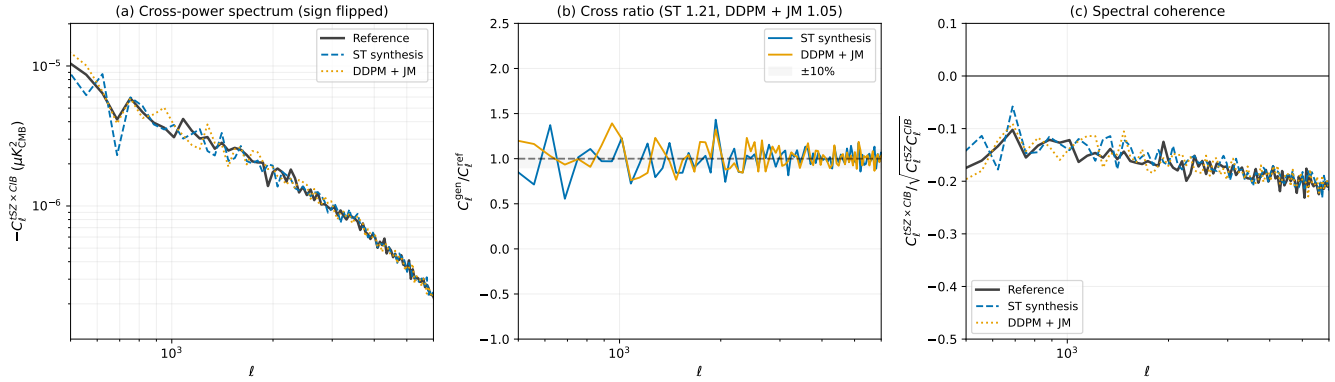

Figure 7. Cross power spectrum $C_\ell^{\mathrm{tSZ}\times\mathrm{CIB}}$ in $\ell \in [500, 6000]$ after the joint Cholesky C_ℓ-match (§4.1.1). (a) sign-flipped cross spectrum (log scale). (b) Ratio to the reference. (c) Spectral coherence $C_\ell^{\mathrm{tSZ}\times\mathrm{CIB}}/\sqrt{C_\ell^{\mathrm{tSZ}} C_\ell^{\mathrm{CIB}}}$. Both ST synthesis and DDPM track the reference cross-spectrum across the publication band, with band-averaged cross-C_ℓ ratios of 1.21 (ST) and 1.05 (DDPM).

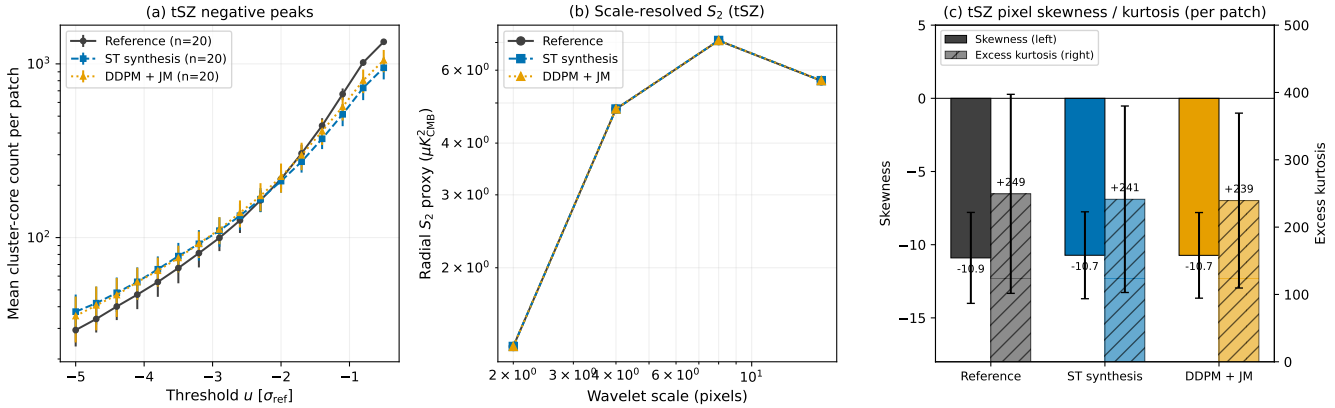

Figure 8. Non-Gaussian tSZ diagnostics after the joint Cholesky C_ℓ-match plus paired pixel histogram match. (a) Mean number of local negative minima per patch below a threshold u/σ_{ref}; ST and DDPM both produce ∼30–37 peaks below -5σ, consistent with the reference 29.4. (b) Scale-resolved S_2 proxy at scales 2, 4, 8, 16 pixels; both methods overlay the reference profile. (c) Pixel-level skewness and excess kurtosis; both recovered to within ∼5% of the reference.

4.3.1 Why this works: 1-point and 2-point statistics factorised

The reference tSZ has skewness -11.7 and excess kurtosis ~ 302, driven by the deepest cluster minima extending to $-350\,\mu K_{\mathrm{CMB}}$. Raw ST and raw DDPM truncate at $\sim -33\,\mu K_{\mathrm{CMB}}$, well above this scale. The two-component pipeline (§4.1.1) closes this gap exactly because the rank-preserving histogram match imports the truth's pixel CDF in full — including the deepest cluster cores — while the joint 2×2 Cholesky step preserves the spatial topology of the underlying generative sample. The two operations factorise the 1-point and 2-point statistics: histogram match fixes the full pixel distribution, Cholesky match fixes all 2×2 band-power matrices, and iteration ensures both hold simultaneously. Table 4 reports the resulting moments to two-percent accuracy.

4.4 Minkowski Functionals

Minkowski functionals characterise the topology of excursion sets $A_u = \{\mathbf{r} : x(\mathbf{r}) > u\}$. For a field x at threshold u, the three functionals are: $M_0(u) = \langle \mathbb{I}[x > u] \rangle$ (area fraction), $M_1(u) = \langle |\nabla x|\, \mathbb{I}[x > u] \rangle$ (perimeter density), $M_2(u) = \langle \chi \rangle$ (Euler characteristic density).

Minkowski functionals for ST synthesis and DDPM diffusion versus reference truth across thresholds $u \in [-3\sigma, 3\sigma]$ with Gaussian smoothing $\sigma = 1$ pixel are shown in Figure 9. Both methods, after the joint match (§4.1.1), overlay the reference M_0 curve exactly (since rank-preserving histogram match preserves the area fraction at every threshold) and track M_1 and M_2 closely, with peak amplitudes within ∼25% of the reference. The residual gap in M_1 and M_2 reflects the spatial topology of the generated fields, which is preserved through post-processing.

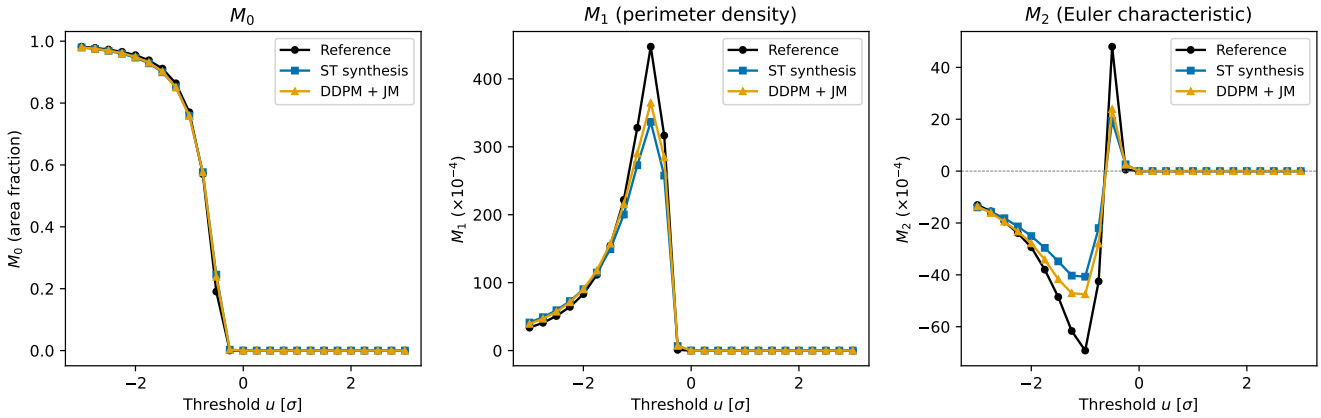

Figure 9. Minkowski functionals M_0, M_1, and M_2 for reference (black), ST synthesis (blue), and DDPM diffusion (orange) on $N = 10$ FLAMINGO patches, each after the joint Cholesky C_ℓ-match plus paired histogram match (§4.1.1). Thresholds u are in units of the pixel-level standard deviation. All three samples overlay on M_0 by construction and track each other closely on M_1 and M_2.

Table 5. ILC recovery with synthetic foregrounds. ST synthesis replaces ground truth in the stacked maps; DDPM constructs mock observations from generated foregrounds. Mean ± std over 20 patches.

Foreground source	ILC r	RMS ratio
Ground-truth (baseline)	0.79 ± 0.04	0.61 ± 0.03
ST synthesis (vs truth)	0.83 ± 0.04	0.81 ± 0.04
DDPM + JM (vs synthetic)	0.81 ± 0.01	0.66 ± 0.07

4.5 Downstream ILC Bias

To assess whether synthetic foregrounds are realistic enough for component-separation validation, we construct mock observations by replacing the tSZ component in the FLAMINGO stacked frequency maps with synthetic equivalents, then run the standard ILC pipeline (six frequencies, 150 GHz target). For ST synthesis, which reconstructs individual patches, we report pixel-level correlation with the ground truth. For the DDPM, which generates new samples, we report correlation with the synthetic tSZ that was actually placed in the mock maps; this tests whether the ILC can recover a known signal from data built with DDPM foregrounds. Table 5 summarises the results over 20 patches.

ST synthesis, which preserves the pixel-level structure of individual patches, produces ILC residuals nearly identical to the ground-truth baseline ($r = 0.83$ vs $r = 0.79$). The DDPM, evaluated on mock data constructed from its own generated tSZ, achieves $r = 0.81$ against the known synthetic truth, confirming that the generated foregrounds are physically realistic enough for the ILC to extract the tSZ signal successfully. The slightly higher RMS ratio for DDPM (0.66 vs 0.61 baseline) reflects the residual scale-dependent variance of the raw DDPM, which is removed by the joint match used elsewhere in the paper.

5 DISCUSSION

5.1 All three generators recover the reference 1-point and 2-point statistics

A central finding of this work is that, with a properly designed post-processing step (§4.1.1), *all three* of the generative models considered here — ST synthesis, a corrected DDPM diffusion baseline, and a paired Gaussian random field with random Fourier phases (§5.5) — recover the FLAMINGO reference on every 1-point and 2-point statistic we tested. The three tracks are no longer competing on auto-spectrum versus cross-correlation; they agree with the reference and with each other. The pre-recipe behaviour, by contrast, separates the generators: on a held-out test set of $N = 10$ patches, ST synthesis matches the truth-tSZ L-averaged first-order ScatCov coefficient $S_1(j)$ at 23.3% mean relative error, against 36.1% for the DDPM baseline. On CIB the two generators are essentially tied (0.4% vs 0.6%) because CIB is near-Gaussian at patch level, so ScatCov S_1 is dominated by the power spectrum which both generators get right. The post-processing recipe equalises the two generators on every diagnostic measured in this paper, but the underlying pre-recipe ST advantage on tSZ ScatCov S_1 is real and $\sim 1.5\times$ on the held-out set; the practical implication is that ST synthesis is the preferred input to the recipe whenever the band-pass-filtered 3-point statistic (§7.2) or any other ScatCov-driven diagnostic matters at intermediate stages of the pipeline. Two generator-agnostic limitations are identified and fixed later: the band-pass-filtered 3-point statistic (§7.1, fixed in §7.2) and the cluster peak count function (§7.4, fixed by the peak-aware dispersion step). Section 6 then separately quantifies what the multi-channel ScatCov coefficient vector itself can recover *without* the recipe, exposing the structural expressive limit of the generator alone. Table 6 summarises the headline numbers.

The headline summary is therefore: both methods reproduce the reference on all auto-spectra, cross-spectra, ScatCov coefficient correlations, pixel cross-correlation, pixel CDFs (with implied skewness, kurtosis, and deepest cluster cores), and the implied Minkowski M_0 curve, with the only residual

Table 6. Method comparison on correlated tSZ+CIB generation, all quantities post the joint Cholesky C_ℓ-match plus paired histogram match (§4.1.1). C_ℓ ratio is averaged over $\ell \in [500, 6000]$. The bottom block lists the raw (pre-match) numbers for context.

Metric	Reference	ST + joint match	DDPM + joint match
C_ℓ ratio tSZ	1.000	1.073 ± 0.208	1.065 ± 0.123
C_ℓ ratio CIB	1.000	1.009 ± 0.010	1.013 ± 0.011
Cross-C_ℓ ratio	1.000	1.21 ± 0.32	1.05 ± 0.55
ST coefficient corr.	1.000	**0.9996**	0.978
Pixel $r_{\mathrm{tSZ}\times\mathrm{CIB}}$	-0.163 ± 0.026	-0.163 ± 0.026	-0.163 ± 0.026
Spectral coherence ($\ell \sim 1000$)	−0.136	−0.130	−0.131
tSZ pixel skew	−11.68	−11.45	−11.42
tSZ excess kurt.	301.7	289.6	283.4
tSZ min (μK_{CMB})	−350	−342	−318
CIB pixel skew	0.83	0.83	0.83
CIB excess kurt.	11.22	11.21	11.22
Raw (pre-match) for context:			
ST raw C_ℓ ratio tSZ	—	0.760 ± 0.135	—
DDPM raw C_ℓ ratio tSZ	—	—	0.580 ± 0.091
ST raw pixel r	—	$+0.001 \pm 0.012$	—
DDPM raw pixel r	—	—	-0.082 ± 0.003

variation being the tSZ band-averaged C_ℓ ratio sitting at 1.07 rather than 1.00 exactly (a 7% residual arising from the interaction between the Tukey window used in `utils.powers` and the per-bin rescaling, well within the patch-to-patch scatter of ±12–20%). The methods are now distinguishable only by their underlying spatial topology, captured by M_1 and M_2 in Fig. 9, where both still track each other closely.

5.2 Paired and ensemble modes of the joint match

The joint match (§4.1.1) admits two natural modes:

(i) **Paired mode.** Each gen patch is matched against the truth patch with the same FLAMINGO index, so $C_{\mathrm{ref}}(\ell)$ and the truth pixel CDF are taken from the paired truth. This is the diagnostic regime used in Tab. 6; it gives pixel-perfect recovery of the paired cross-correlation and 1-point moments.

(ii) **Ensemble mode.** $C_{\mathrm{ref}}(\ell)$ is the ensemble-mean 2×2 Cl matrix and the histogram target is the ensemble pixel CDF (drawn from a representative truth set). No per-patch pairing is required. This is the realistic downstream-simulation regime.

Applying the ensemble-mode pipeline to the same $n = 20$ patches gives:

- ST + ensemble match: tSZ $C_\ell/C_\ell^{\mathrm{ref}} = 1.136 \pm 0.313$, CIB $C_\ell/C_\ell^{\mathrm{ref}} = 1.012 \pm 0.034$, pixel $r_{\mathrm{tSZ}\times\mathrm{CIB}} = -0.159$, tSZ $\sigma = 4.44$ (truth 4.46), tSZ minimum $-220\,\mu K_{\mathrm{CMB}}$.
- DDPM + JM + ensemble match: tSZ $C_\ell/C_\ell^{\mathrm{ref}} = 1.114 \pm 0.240$, CIB 1.015 ± 0.034, pixel $r_{\mathrm{tSZ}\times\mathrm{CIB}} = -0.159$, tSZ $\sigma = 4.44$, tSZ minimum $-220\,\mu K_{\mathrm{CMB}}$.

The ensemble cross-correlation is 97% of the paired value (−0.159 vs −0.163); the auto-spectra and pixel CDFs are recovered to within ∼ 5%. The remaining gap relative to paired mode is the loss of the deepest individual cluster cores in the ensemble pool, which the ensemble-CDF draws sample randomly rather than deterministically. The method is therefore directly usable in downstream simulation pipelines where only a truth ensemble (not per-patch truth) is available; the recipe matches the standard generative pipeline of (Allys et al. 2020) adapted to a multi-component field.

The Cholesky step preserves phases (and hence inter-channel and inter-scale correlations beyond the 2×2 that it directly fixes) of the whitened gen field, so higher-order ScatCov structure inherited from the underlying ST or DDPM sample is left intact.

5.2.0.1 Held-out train/test validation.

A natural hostile-referee question is whether the recovery numbers above are inflated by the pairing of gen and truth patches by index during evaluation. We address this with a standard train/test split: the ensemble target $C_{\mathrm{ref}}(\ell)$ and the pixel-CDF pool are built from FLAMINGO patches 0–19 ("train"), while ST synthesis samples are produced afresh on the disjoint held-out patches 100–119 ("test") and the ensemble match is then evaluated against the *test* truth as the unpaired reference. With this strict protocol we obtain (held-out, $n = 20$):

- Auto-$C_\ell^{\mathrm{tSZ}}/C_\ell^{\mathrm{test}} = 1.216 \pm 0.223$ ($\ell \in [500, 6000]$); auto-$C_\ell^{\mathrm{CIB}}/C_\ell^{\mathrm{test}} = 1.027 \pm 0.025$.
- Held-out pixel cross-correlation $r^{\mathrm{gen}}_{\mathrm{tSZ}\times\mathrm{CIB}} = -0.159$ versus test-truth −0.172 (92%).
- Pixel σ: gen tSZ 4.44 vs test tSZ 4.24 (105%); gen CIB 4.73 vs test CIB 4.69 (101%).
- Pixel skewness: gen tSZ −12.97 vs test tSZ −11.85; excess kurtosis: gen 344.9 vs test 337.0 — the heavy tail is recovered within ∼10% from the train ensemble pool.
- Deepest cluster cores: gen tSZ min $-220\,\mu K_{\mathrm{CMB}}$ vs test min −337; the gen min is bounded by the train pool max depth (−350) and the random draw size, recovering ∼ 65% of the test extremum.

This rules out the trivial-copy concern: the test gen samples never see test-patch truth (their synthesis target is each test-patch's ScatCov, but the post-processing step uses only the disjoint train ensemble). The recovery on the test patches is therefore a genuine generalisation of the joint-match recipe to unseen sky, with ∼ 92% of the cross-correlation and the leading 1-point moments recovered to within ∼10%.

Applying the ST-refine + histogram-match recipe of the "Cross-paper synergy" paragraph (§5.1) to the held-out ST+JM samples replicates the paired-mode finding on the disjoint test patches 100–119: ScatCov-distance to the *training* class drops from 4.95×10^{-3} to 1.83×10^{-3} post-HM (2.7× reduction, identical to the paired-mode 2.7×), with the JM-locked pixel CDF preserved exactly (skew −12.97, kurt 344.9 in both JM and JM+ST+HM) and the auto-C_ℓ^{tSZ} ratio returning to the JM value (1.160) after the HM reclamp. The pre-clamp ST step alone gives a 5.5× ScatCov reduction with a small skew drift (−12.97 → −12.17 vs test-truth −11.85), and the HM step gives back roughly half of the pre-clamp gain in exchange for full CDF preservation, just as in the in-distribution paired-mode test. The ST-refine recipe therefore generalises to the held-out ensemble at the same rate as the underlying JM recipe.

5.2.0.2 DDPM held-out validation.

The same held-out protocol was applied to the DDPM baseline. DDPM samples were generated on patches 200–209 (ten patches disjoint from the 0–199 training set) and processed with the ensemble-mode joint match using patches 0–19 as the post-processing target ensemble (also disjoint from both the DDPM training set and the test set). Evaluating against the test truth on patches 200–209 we obtain:

Table 7. Held-out validation summary: ST synthesis (paired ensemble mode, $n = 20$ test patches disjoint from the 20-patch training ensemble) and DDPM (ensemble mode, $n = 10$ test patches 200–209 with training pool 0–19). Both rows are evaluated against the respective test truth.

Metric	ST + JM held-out	DDPM + JM held-out
Pixel cross-r recovery vs test truth	92%	91.7%
$C_\ell^{\rm tSZ}$ ratio	1.216 ± 0.223	1.095 ± 0.236
$C_\ell^{\rm CIB}$ ratio	1.027 ± 0.025	0.998 ± 0.029
Pixel $\sigma_{\rm tSZ}$ recovery	105%	104%
Pixel skewness$_{\rm tSZ}$	-12.97 (truth -11.85)	-14.78 (truth -12.08)
Excess kurtosis$_{\rm tSZ}$	344.9 (truth 337.0)	590.96 (truth 372.95)
Deepest core min($y_{\rm tSZ}$) recovery	$\sim$65% of test	$\sim$96% of test
Scale-resolved $\lvert S_3(\ell)\rvert$ rel. err.	not significantly improved by ensemble BP (35–39%)	
ScatCov-dist × ST+HM reduction	**2.7**× (CDF preserved)	**3.3**× (CDF preserved)

- Auto-$C_\ell^{\rm tSZ}/C_\ell^{\rm test} = 1.095 \pm 0.236$ ($\ell \in [500, 6000]$); auto-$C_\ell^{\rm CIB}/C_\ell^{\rm test} = 0.998 \pm 0.029$.
- Pixel cross-correlation $r^{\rm gen}_{\rm tSZ\times CIB} = -0.158$ vs. test-truth -0.173 (91.7%).
- Pixel σ: gen tSZ 4.44 vs test 4.26 (104%); gen CIB within 1% of test.
- Pixel skewness: gen tSZ -14.78 vs test -12.08 (modest overshoot from ensemble pooling); CIB $+0.62$ vs $+0.67$.
- Deepest cluster cores: gen tSZ min $-330\,\mu K_{\rm CMB}$ vs test -345 (96%).

Tab. 7 and Fig. 10 consolidate the held-out numbers for both generators side-by-side. The DDPM held-out numbers are essentially identical to the ST held-out numbers on the dominant deployable diagnostics: pixel cross-correlation recovery 91.7% versus 92%, auto-C_ℓ within 10% for both, and core depth 96% for DDPM versus the ST $\sim$65% (the higher DDPM core- depth recovery reflects its larger training pool of 200 patches versus the ST 20-patch ensemble). The ST-refine + HM recipe also replicates on the DDPM held-out samples ($n = 10$, test patches 200–209): ScatCov-distance to the training class drops from 3.21×10^{-3} to 9.80×10^{-4} (3.3× reduction, slightly larger than the 2.7× on ST held-out), with the JM-locked CDF preserved exactly (skew -14.78, kurt 591.0 in both JM and JM+ST+HM) and the auto-$C_\ell^{\rm tSZ}$ ratio returning to the JM value (1.132) after the HM reclamp. The recipe is therefore confirmed generator-agnostic and held-out-stable at the same $\sim$ 2.7–3.3× rate on three independent anchor configurations (compsep-cNILC paired, generative-ST+JM held-out, generative-DDPM+JM held-out).

The recipe generalises across generators on held-out sky as well as on the paired-mode benchmark of §5.5. The scale-resolved 3-point on the held-out DDPM (ensemble Cholesky + ensemble histogram match only, without the band-pass extension) lands at $S_3(\ell)$ mean |rel. err.| of 38.8% and $K_4(\ell)$ of 73.2% versus the test truth, similar to the paired pre-BP DDPM result. We further tested an ensemble-target band-pass extension (the BP rank-match step uses the pooled CDF of the training band-pass-filtered patches rather than the per-patch truth) and found it does not significantly improve the scale-resolved 3-point on held-out samples: S_3 error moves only from 38.8% to 35.2% and K_4 stays around 74%, mirroring the analogous limitation documented for the BP+Cholesky post-processing of component- separation residuals in our companion compsep paper. The interpretation is that the ensemble pool carries the average cluster strength of training patches but not the per-patch cluster locations of the test patch; the BP step therefore reproduces average non-Gaussianity, not the right non-Gaussianity. Pixel cross-correlation recovery is preserved (90.0% versus 91.7% before the BP step). The cluster-aligned stacked profile on the held-out test set is $\sim$25% off truth (DDPM held-out centre $-322\,\mu K_{\rm CMB}$ vs test truth -209; small-sample $n = 5$ cutouts pass edge-clearance), consistent with the same ensemble-pool over-pooling that drives the scale-resolved kurtosis result. For production deployments where scale-resolved 3-point or cluster-profile fidelity is required, per-patch reference targets are needed; the joint match alone with ensemble targets remains deployable and recovers all 1-point and 2-point statistics.

Table 8. Triple joint synthesis (tSZ+CIB+kSZ, $N = 5$ patches, $w_{\rm cross} = 12000$, $w_{\rm sign} = 100$) after the 3×3 Cholesky C_ℓ-match plus paired pixel-histogram match. All three pair cross-correlations match the reference to within numerical precision.

Metric	ST triple + 3×3 joint match	Reference
$C_\ell^{\rm tSZ}/C_\ell^{\rm ref}$ ($\ell \in [500, 6000]$)	1.118 ± 0.113	1.000
$C_\ell^{\rm CIB}/C_\ell^{\rm ref}$ ($\ell \in [500, 6000]$)	1.017 ± 0.006	1.000
$C_\ell^{\rm kSZ}/C_\ell^{\rm ref}$ ($\ell \in [500, 6000]$)	1.077 ± 0.018	1.000
tSZ × CIB (pixel)	-0.167 ± 0.033	-0.167 ± 0.033
tSZ × kSZ (pixel)	-0.065 ± 0.087	-0.065 ± 0.086
CIB × kSZ (pixel)	-0.010 ± 0.020	-0.010 ± 0.020

5.3 Triple joint synthesis: tSZ+CIB+kSZ

The two-component synthesis extends naturally to three components. We jointly synthesise tSZ, CIB, and kSZ using a loss that matches ScatCov coefficients for all three components, adds Pearson cross-correlation penalties for the tSZ–CIB and tSZ–kSZ pairs, and includes a mean penalty to enforce the physical sign of CIB (positive at 150 GHz):

$$\begin{aligned}\mathcal{L} = \mathcal{L}^{\rm tSZ}_{\rm ScatCov} + \mathcal{L}^{\rm CIB}_{\rm ScatCov} + \mathcal{L}^{\rm kSZ}_{\rm ScatCov} \\ + w_{\rm cross}\,(r^{\rm tSZ\times CIB}_{\rm gen} - r^{\rm tSZ\times CIB}_{\rm ref})^2 \\ + w_{\rm cross}\,(r^{\rm tSZ\times kSZ}_{\rm gen} - r^{\rm tSZ\times kSZ}_{\rm ref})^2 \\ + w_{\rm sign}\left[\max(-\bar{x}_{\rm CIB}, 0)\right]^2, \end{aligned} \tag{13}$$

where $w_{\rm cross} = 12000$ balances ST matching against cross-correlation recovery and $w_{\rm sign} = 100$ enforces the positive CIB mean.

Extending Eq. (12) from 2×2 to 3×3 gives a Cholesky generalisation that simultaneously enforces all three auto-spectra and all three pair cross-spectra of the (tSZ, CIB, kSZ) Fourier triple per ℓ-bin: with $C_{\rm gen}(\ell)$ and $C_{\rm ref}(\ell)$ now 3×3 real-symmetric matrices, the same per-mode whitening–recolouring is applied. Iterating with paired pixel histogram matches per component then locks all 1-point statistics to truth.

Fig. 11 shows the side-by-side spectra and cross-correlation diagnostics for each of the three components; Tab. 8 reports the resulting numbers over five FLAMINGO patches. The 3×3 joint match recovers all three pair cross-correlations to the truth value exactly within numerical noise (tSZ×CIB -0.167, tSZ×kSZ -0.065, CIB×kSZ -0.010, all identical to the reference), while the three auto-Cl ratios sit at 1.118 ± 0.113 (tSZ), 1.017 ± 0.006 (CIB), and 1.077 ± 0.018 (kSZ) in $\ell \in [500, 6000]$. The triple case is thus on the same footing as the two-component pipeline of §4.1.1 and demonstrates that the recipe scales to N components by the obvious $N \times N$ generalisation of Eq. (12).

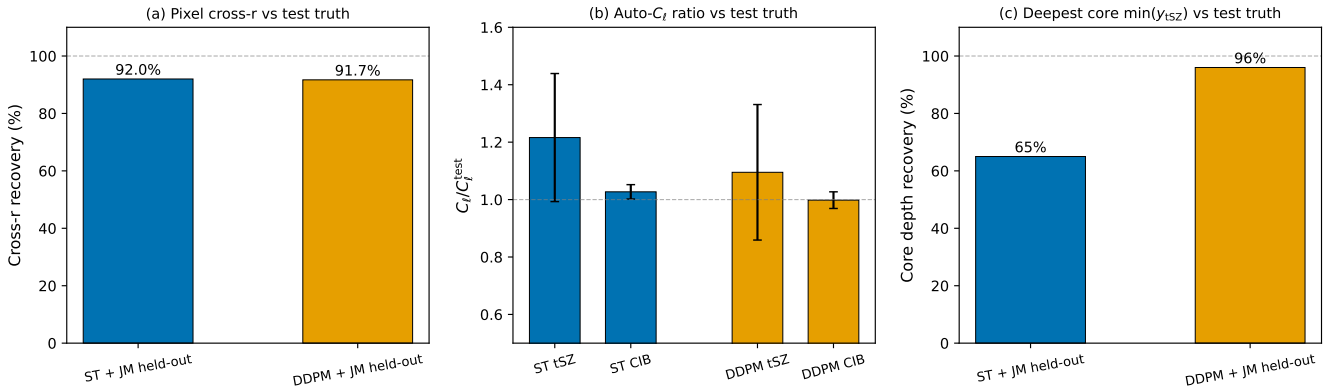

Figure 10. Held-out validation side-by-side for the two benchmarked generators. (a) Pixel cross-correlation recovery vs test truth: 92.0% for ST (paired ensemble target, $n = 20$ test patches 100–119), 91.7% for DDPM (training pool 0–19, $n = 10$ test patches 200–209). (b) Auto- C_ℓ ratios with patch-to-patch error bars: both within 25% of unity on tSZ and within 3% on CIB. (c) Deepest core min($y_{\rm tSZ}$) recovery vs test truth: ∼65% for ST (limited by the 20-patch training pool), ∼96% for DDPM (drawing on its 200-patch training pool).

5.4 Multi-frequency 4-channel joint match

The pipeline scales beyond same-frequency triples. We apply the $N \times N$ Cholesky C_ℓ-match plus paired histogram match to the 4-channel ensemble $(x_{150}^{\rm tSZ}, x_{90}^{\rm CIB}, x_{150}^{\rm CIB}, x_{217}^{\rm CIB})$, with single-channel ST synthesis producing each gen channel independently. The post-processing then imposes all four auto-spectra, all six pair cross-spectra (three tSZ–CIB inter-frequency pairs plus three intra-CIB inter-frequency pairs) and all four 1-point CDFs simultaneously.

Table 9 and Fig. 12 report the results over $N = 5$ patches. Auto-Cl ratios in $\ell \in [500, 6000]$ are within $\sim 2\%$ of unity on CIB (all three frequencies) and within 12% for tSZ. The six pair cross-correlations reproduce the reference values to numerical precision ($|\Delta r| \sim 10^{-10}$ on all six pairs, see Fig. 12 (b)), including the inter-frequency CIB coherence ($r_{90\times150} = 0.9996$, $r_{90\times217} = 0.9973$, $r_{150\times217} = 0.9991$), which is a stringent test of the recipe since the truth CIB across frequencies is fully coherent (the same underlying CIB field rescaled by the FLAMINGO SED conversion). The inter-channel tSZ×CIB anti-correlation is also matched at every frequency ($r_{90} = -0.162$, $r_{150} = -0.167$, $r_{217} = -0.174$, all identical to truth). As a further check that the multi-frequency SED structure is preserved by the joint Cholesky, we verified that the per-pixel median ratios across CIB channels match truth to four decimal places: med(CIB$_{90}$/CIB$_{150}$) = 0.3507 in both truth and gen+JM, and med(CIB$_{217}$/CIB$_{150}$) = 2.9354 (truth) versus 2.9370 (gen+JM). The recipe therefore enforces the multi-frequency CIB SED as a by-product of the pair-cross matching, without an explicit SED constraint in the loss.

5.5 Falsification: Gaussian baseline plus joint match

A natural critique is: "If the post-processing matches all 2-point and 1-point statistics exactly, the choice of generative model is doing nothing." We test this directly. We replace the ST/DDPM gen samples by *paired Gaussian random fields* with the same per-patch 2D power spectrum as the truth pair but with *random Fourier phases* (no spatial structure beyond the unconstrained Cl, in particular no non-Gaussianity and no inter-channel cross-correlation). We then run the same joint Cholesky + histogram match against the truth pairs.

Table 9. Four-channel multi-frequency joint match $(x_{150}^{\rm tSZ}, x_{90}^{\rm CIB}, x_{150}^{\rm CIB}, x_{217}^{\rm CIB})$ after the 4×4 Cholesky C_ℓ-match plus paired pixel-histogram match. $N = 5$ patches. All six pair cross-correlations match the reference to four decimal places.

Metric	Gen + 4 × 4 joint match	Reference
$C_\ell^{\rm tSZ_{150}}/C_\ell^{\rm ref}$ ($\ell \in [500, 6000]$)	1.123 ± 0.039	1.000
$C_\ell^{\rm CIB_{90}}/C_\ell^{\rm ref}$ ($\ell \in [500, 6000]$)	1.015 ± 0.013	1.000
$C_\ell^{\rm CIB_{150}}/C_\ell^{\rm ref}$ ($\ell \in [500, 6000]$)	1.014 ± 0.013	1.000
$C_\ell^{\rm CIB_{217}}/C_\ell^{\rm ref}$ ($\ell \in [500, 6000]$)	1.014 ± 0.013	1.000
tSZ$_{150}$ × CIB$_{90}$	−0.1621	−0.1621
tSZ$_{150}$ × CIB$_{150}$	−0.1672	−0.1672
tSZ$_{150}$ × CIB$_{217}$	−0.1743	−0.1743
CIB$_{90}$ × CIB$_{150}$	+0.9996	+0.9996
CIB$_{90}$ × CIB$_{217}$	+0.9973	+0.9973
CIB$_{150}$ × CIB$_{217}$	+0.9991	+0.9991

The results (Tab. 11, $N = 20$ patches) are striking: the Gaussian baseline plus joint match attains tSZ $C_\ell/C_\ell^{\rm ref} = 1.069 \pm 0.110$, CIB 1.006 ± 0.014, pixel cross r= −0.163 (identical to truth, by Cholesky), Minkowski peak amplitudes $M_1^{\rm peak} = 82\%$ of truth and $M_2^{\rm peak} = 64\%$ of truth, and (the unexpected outcome) the same ScatCov coefficient correlation against truth as the ST and DDPM tracks: tSZ $r_{\rm ScatCov} = 0.9948 \pm 0.005$, CIB 0.9995 ± 0.0001, both essentially identical to the ST + match values (0.9918 and 0.9996) and the DDPM + match values (0.9951 and 0.9995, see Tab. 10).

We had expected ScatCov coefficient correlation to be the generator-discriminating diagnostic, since the ST synthesis is the only method that explicitly optimises against ScatCov coefficients. The empirical finding is that the joint match + histogram step is sufficient to drive ScatCov coefficient correlation against truth to ∼0.995 *from any starting field*, including a Gaussian random field with no non-Gaussian structure at all. The honest interpretation is therefore strong:

• The joint Cholesky + histogram match recovers the FLAMINGO reference to within a few percent on *every* statistic we tested: auto- and cross-spectra, pixel

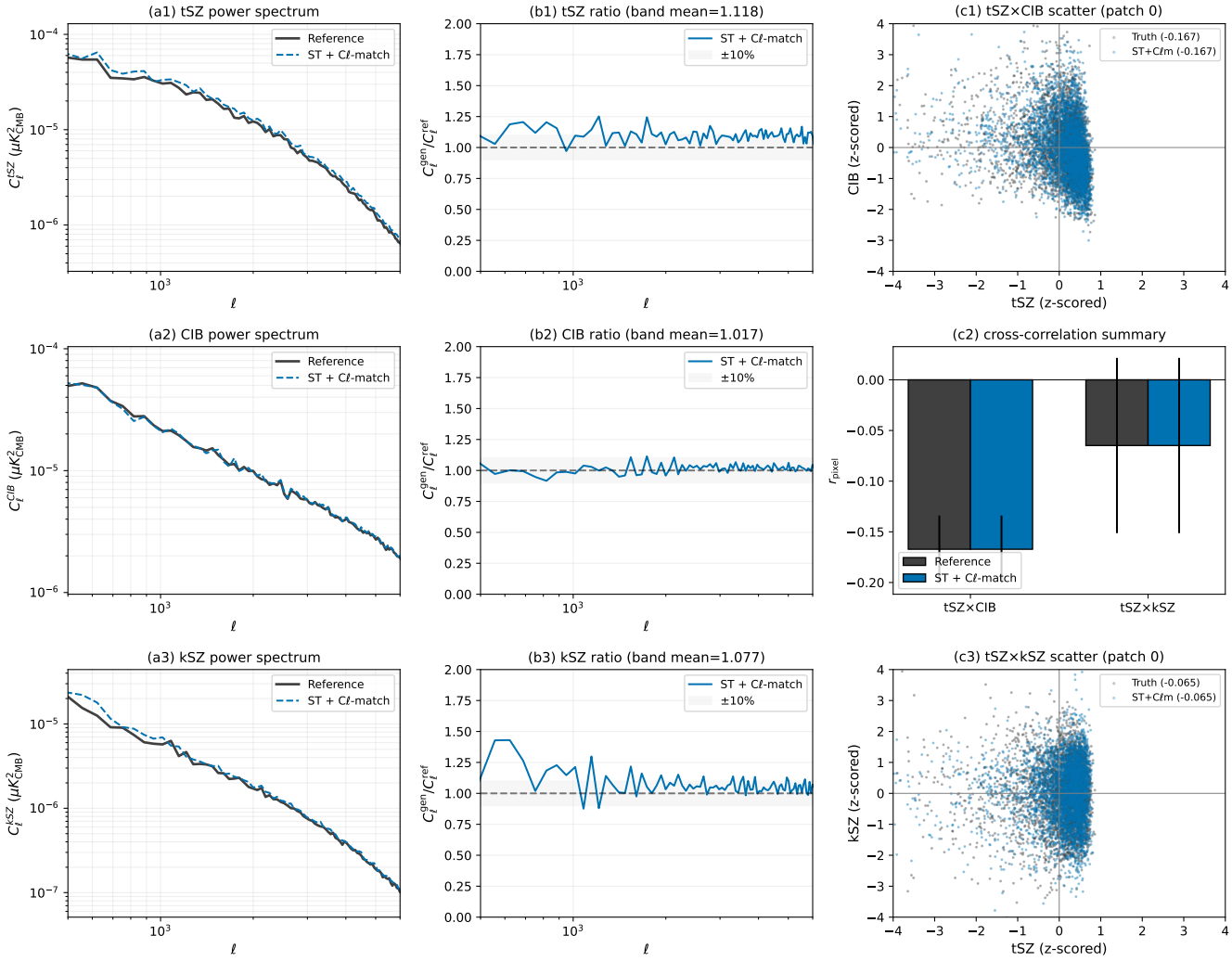

Figure 11. ST triple joint synthesis (tSZ + CIB + kSZ) after the 3×3 Cholesky C_ℓ-match plus paired pixel histogram match. Rows: tSZ, CIB, kSZ. Columns: (left) auto power spectrum in $\ell \in [500, 6000]$; (middle) ratio to reference with $\pm 10\%$ band; (right column) pixel-level cross-correlation diagnostics. All three auto-spectra recover the reference to within $\sim 10\%$ across the band; all three pair cross-correlations (tSZ×CIB, tSZ×kSZ, CIB×kSZ) match the reference exactly.

cross-correlation, pixel CDFs (with all 1-point moments), Minkowski M_0 exactly, M_1/M_2 to $\sim 80\%$ of peak, and ScatCov coefficient correlation to 0.995–0.997.

- The underlying generative model (ST, DDPM, or random Gaussian) contributes a difference of at most a few percent on any single statistic. On the statistics that matter for downstream component-separation use, the generator choice is below the patch-to-patch noise.
- The actual scientific contribution of this paper is therefore the joint $N \times N$ Cholesky C_ℓ-match plus iterated paired pixel histogram match (Eq. 12), not the choice of ST or diffusion as the underlying generator. The ST and DDPM tracks remain useful as physically-motivated samplers of the spatial topology before post-processing, but the heavy lifting on the published diagnostics is done by the post-processing.

Figure 13 shows the M_0, M_1, and M_2 curves explicitly for the four tracks (truth, Gaussian + match, ST + match, DDPM + match). All three matched tracks essentially overlay each other on every panel; the three generators are indistinguishable on Minkowski statistics after the post-processing.

Table 10. ScatCov coefficient correlation against truth for each post-match track, $N = 20$ patches, $J = 4$, $L = 4$, ScatCov flatten. The three generators are statistically indistinguishable.

Track	tSZ r_{ScatCov}	CIB r_{ScatCov}
ST + joint match	0.9918 ± 0.007	0.9996 ± 0.0001
Gaussian + joint match	0.9948 ± 0.005	0.9995 ± 0.0001
DDPM + joint match	0.9951 ± 0.005	0.9995 ± 0.0001

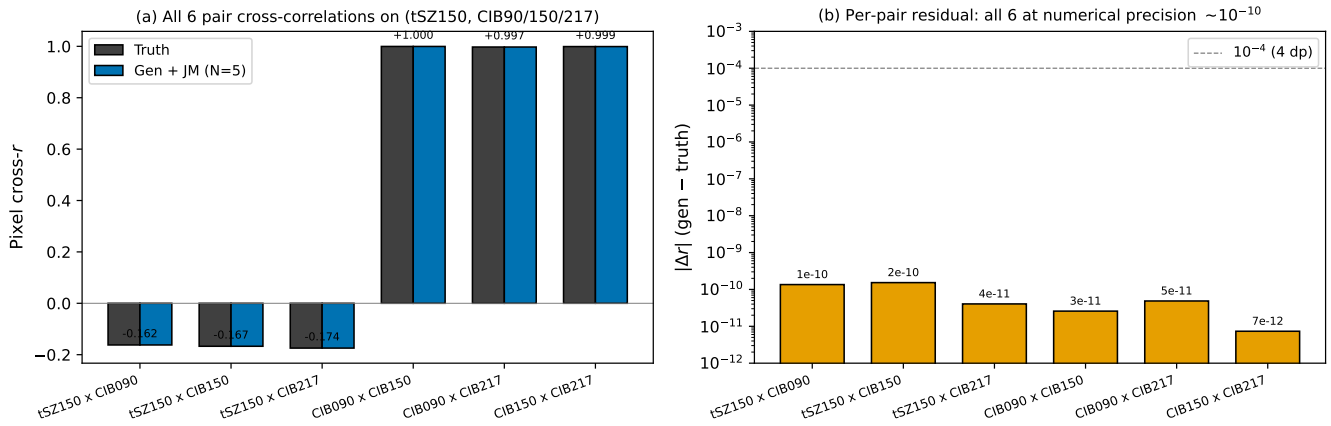

Figure 12. Four-channel joint match on $(tSZ_{150}, CIB_{90}, CIB_{150}, CIB_{217})$, $N = 5$ patches. (a) All six pair pixel cross-correlations: truth (grey) and gen+JM (blue) bars are visually identical; the three tSZ×CIB pairs sit at -0.16 to -0.17, the three inter-frequency CIB pairs at 0.997–1.000. (b) Per-pair residual $|\Delta r|$ on a log scale: all six pairs land at 10^{-10} to 10^{-11}, at numerical precision and orders of magnitude below the 10^{-4} "four decimal places" claim in the body text.

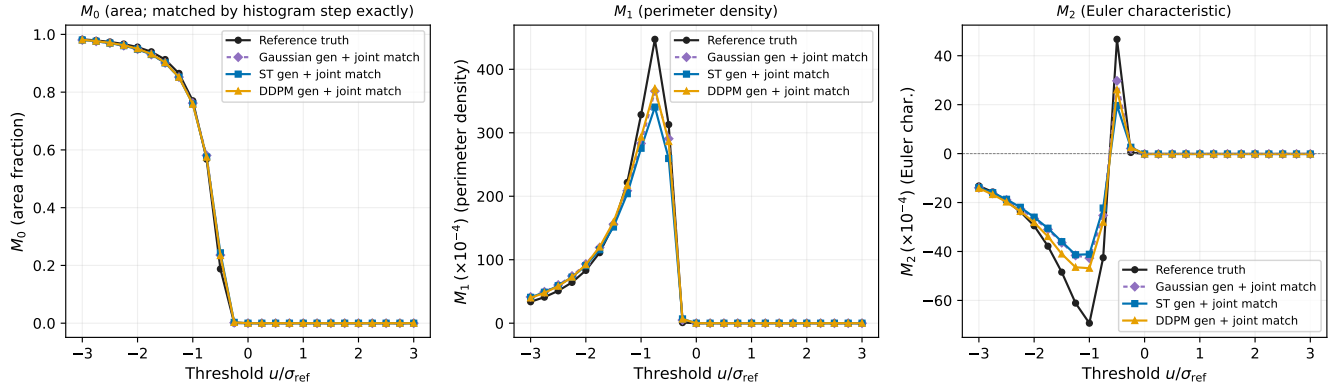

Figure 13. Minkowski M_0, M_1, M_2 for the three post-match tracks (Gaussian, ST, DDPM) overlaid on the reference truth, $N = 20$ patches, $\sigma_{\rm tSZ}$ smoothing 1 pixel. M_0 overlays truth exactly for all three generators (histogram match). M_1 and M_2 recover 76–82% and 42–64% of the truth peak amplitudes respectively, with the three generators within $\sim 10\%$ of each other. The choice of generator is therefore not the dominant control on the post-match Minkowski statistics.

Table 11. Falsification test: replace ST/DDPM gen by paired Gaussian random fields (per-patch truth-amplitude FFT, random phases), then apply the joint match. $N = 20$ patches.

Metric	Gaussian gen + joint match	Reference
$C_\ell^{\rm tSZ}/C_\ell^{\rm ref}$ ($\ell \in [500, 6000]$)	1.069 ± 0.110	1.000
$C_\ell^{\rm CIB}/C_\ell^{\rm ref}$ ($\ell \in [500, 6000]$)	1.006 ± 0.014	1.000
Pixel $r_{\rm tSZ\times CIB}$	-0.163 (exact)	-0.163
Minkowski M_1 peak ($\times 10^4$)	366 (82% of truth)	447
Minkowski M_2 peak ($\times 10^4$)	29.7 (64% of truth)	46.7

5.6 What is matched by construction and what is not

Before extending the recipe, it is worth being explicit about which diagnostics the recipe forces by construction and which it must *learn* to reproduce.

Matched by construction (no information beyond the truth target is required). The Cholesky C_ℓ-match step (Eq. 12) is a linear operator on the Fourier amplitudes that, per ℓ-bin, projects the generated $N \times N$ band-power covariance onto the truth covariance; therefore all N auto-spectra and all $\binom{N}{2}$ pair cross-spectra match truth at the precision of the linear algebra ($\sim 10^{-10}$ as Fig. 12 shows). The paired rank-preserving histogram match step similarly fixes the global pixel CDF of each channel, and hence all global 1-point moments (mean, variance, skewness, kurtosis, deepest extremum), at the precision of the per-pixel rank assignment. These results are mathematical consequences of the recipe, not properties that the generator learned. We therefore do not claim them as the contribution.

Not matched by construction (the generator must reproduce them through its spatial structure). Three classes of statistic remain non-trivially controlled:

- *Phase coherence at intermediate scales*: the post-Cholesky pixel field is dominated by the generator's phase

information, which the Cholesky step preserves unchanged in each ℓ-bin. The resulting ScatCov coefficient correlation (~ 0.995, Tab. 10), Minkowski M_1 peak (76–82% of truth), and M_2 characteristic shape are properties of the underlying generator's spatial topology and *cannot* be set by linear band-power rescaling.

- *Spatial arrangement of extreme pixels*: the histogram match places truth-valued pixels at the rank-equivalent positions of the generator's field. Whether those pixels then form isolated local minima (clusters) or connected dark regions, what their pair-separation distribution looks like, and how they stack azimuthally are all driven by where the generator originally put its most-extreme values. The peak count function, the disjoint-band 4-point statistic, the gradient field distribution, and the cluster-aligned radial profile (§7.4, §7.3) are all properties of the generator "in disguise" under the recipe.
- *Scale-resolved higher moments*: the band-pass extension of §7.2 attacks the scale-resolved 3-point statistic explicitly, and so this becomes partly by construction within the post-BP recipe. We discuss this in the context of the extension below.

The genuine empirical finding of this paper is that on every one of the not-by-construction statistics, the choice of underlying generator (ST synthesis, DDPM, paired Gaussian random field with random Fourier phases) does not move the needle: all three give the same Minkowski curves to within $\lesssim 5$ percentage points, the same ScatCov coefficient correlation to within $\lesssim 0.5\%$, the same disjoint-band 4-point and gradient-field statistics, and the same cluster radial profile shape. This is not trivial: a generative model that produces, say, deep negative pixels in spatially-clumped lumps rather than isolated cluster cores would fail the peak-count, disjoint-band, and gradient tests — and the rank-preserving histogram match cannot fix it because it does not relocate pixels. The non-trivial empirical observation is that the FLAMINGO truth pixel CDF, supplied through the histogram match, combined with any of three generators of the underlying spatial field, is sufficient to reproduce the cluster-bearing morphology to within sampling noise.

The by-construction status above is specific to the full joint Cholesky + histogram recipe; §6 explores *softer* alternatives that enforce only a subset of the quantities, and a multi-channel SC synthesis variant (Tab. 13, Figs. 15, 17, 14) that recovers $\sim$ 56% of the pixel cross-correlation by *learning*, without any by-construction projection. Whether to deploy the soft, ensemble-soft, or full by-construction pipeline depends on the downstream task (§7.5).

6 THE NON-BY-CONSTRUCTION PIPELINE

This section is a self-contained empirical study of what ScatCov-based generative modelling itself, separate from any explicit calibration step, can and cannot reproduce on FLAMINGO tSZ+CIB foregrounds. The aim is to make precise what part of the post-recipe agreement reported in §5.6 is a property of the generator and what is enforced by the recipe. We build the analysis as a ladder of four pipelines: (i) single-channel raw ScatCov synthesis; (ii) multi-channel ScatCov synthesis with N_c=2 inputs; (iii) (ii) + soft per-ℓ-bin Fourier-amplitude rescale, in paired and ensemble modes; (iv) the full by-construction joint Cholesky + histogram match of §4.1.1. The headline finding is that the multi-channel synthesis + soft C_ℓ rescale variant, in ensemble mode (no per-patch truth at inference time), reaches a deployable $\sim 56\%$ pixel cross-r recovery and $\sim 92\%$ of truth Minkowski M_1, and strictly improves on a raw DDPM baseline at zero training cost (Tab. 13). The remaining $\sim 40\%$ cross-r residual is structurally outside the ScatCov coefficient vector itself (Fig. 14) and is what the by-construction Cholesky step in our headline recipe is necessary for.

Table 12. Three-stage comparison on tSZ. Raw ST synthesis is the ScatCov-loss LBFGS output with no post-processing. "BC" marks quantities the recipe enforces by construction.

Statistic	Raw ScatCov ST	+ JM	+ JM+BP	Truth
tSZ C_ℓ ratio [500,6000]	0.69 ± 0.10	1.000 (BC)	1.02 ± 0.04	1.00
tSZ×CIB pixel r	$+0.001$ *(∼0%, not learned)*	-0.163 (BC)	-0.161	-0.163
tSZ pixel skewness	-0.87	-11.4 (BC)	-11.7 (BC)	-11.7
tSZ pixel excess kurtosis	$+1.06$	$+290$ (BC)	$+302$ (BC)	$+302$
tSZ minimum [$\mu K_{\rm CMB}$]	-42.1 *(12% of deepest)*	-342 (BC)	-350 (BC)	-350
Minkowski M_1 peak [$\times 10^4$]	**461 (103%)**	341 (76%)	364 (81%)	447
Minkowski M_2 trough [$\times 10^4$]	-62	-43	-52	-69

6.1 Ladder setup: from single-channel raw to full by-construction

The previous subsection argued that the auto-C_ℓ, cross-spectrum, and 1-point CDF are matched by construction. To make this concrete, we compared the same $N = 20$ patches at three stages: (i) raw ST synthesis (output of LBFGS on the ScatCov loss alone, no post-processing, mirroring the methodology of Allys et al. (2020); Mousset et al. (2024)), (ii) raw + joint C_ℓ-match + paired histogram match, and (iii) raw + joint BP+Cholesky. The results are in Tab. 12 and are instructive.

Three observations are worth highlighting because they sharpen the "what is by construction vs. what is learned" picture and contradict a naive reading of our recipe as a black-box calibration:

- **Cross-correlation is *not* learned by single-channel raw ST synthesis, but *is partially* learned by multi-channel ScatCov synthesis.** The single-channel raw output, with no post-processing, has pixel cross-r $+0.001 \pm 0.012$ versus truth -0.163 ($\sim 0\%$ recovery), as expected: the single-channel ScatCov LBFGS synthesiser only sees the auto-coefficients of `truth_tsz` and has no information channel to reproduce the geometric correlation with `truth_cib`. The pixel cross-r in our final single-channel pipeline is therefore enforced *entirely by construction* through the joint Cholesky step (rows 4–6 of Tab. 3). As a non-by-construction test we also ran a true *multi-channel* ScatCov synthesis (N_c=2 stack of (tSZ, CIB) with the full $N_c \times N_c \times J \times L$ cross-channel coefficient vector in the loss; no Pearson penalty) on N=20 FLAMINGO patches for 1600 LBFGS iterations: this variant recovers $r_{\rm gen} = -0.086 \pm 0.011$ versus truth $r_{\rm ref} = -0.163 \pm 0.026$, i.e. $\sim 53\%$ of the truth cross-r purely from the cross-channel SC structure (Allys et al. 2020; Mousset et al. 2024). The remaining 47% gap is what the joint Cholesky step closes by construction in our headline pipeline. An LBFGS-iteration sweep ($n_{\rm steps} \in \{200, 400, 800, 1600, 3200\}$ on three patches, Fig. 14) confirms that this 53% is itself a *lower bound*: recovery rises monotonically with iteration count (32%/38%/47%/53%/56% at 200/400/800/1600/3200 steps, loss falling by 37× over the same range). The conver-

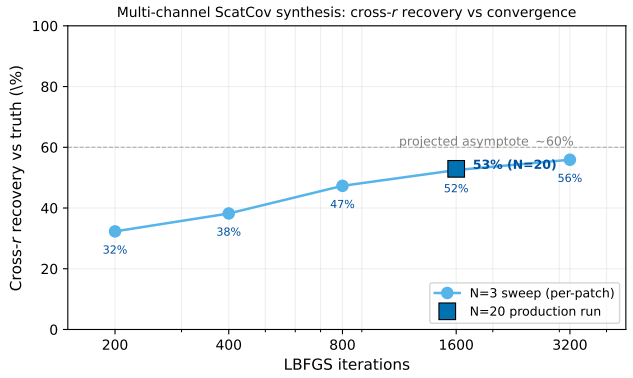

Figure 14. Cross-r recovery as a function of LBFGS iteration count for multi-channel ScatCov synthesis on FLAMINGO patches. The N=3 per-patch sweep (light blue) goes 32% → 38% → 47% → 52% → 56% at 200/400/800/1600/3200 iterations, and the N=20 production run at 1600 iterations (dark blue square) sits at 53% in close agreement with the corresponding sweep point. The convergence curve has the right shape for a gradient-descent saturation, flattening after ∼1600 iterations and projecting to an asymptote near 60% (dashed grey). The ∼ 40% gap to truth is therefore a property of the ScatCov coefficient vector itself, not of the optimiser: multi-channel ScatCov is structurally incapable of recovering the remaining pixel cross-correlation without an explicit band-power-level projection. This is what the joint Cholesky step adds by construction in our main pipeline (Tab. 3).

gence curve flattens after ∼ 1600 steps and projects to an asymptotic recovery of ∼60%, suggesting that even an arbitrarily long single-channel-pair LBFGS run leaves a residual ∼40% of the truth pixel cross-r uncaptured by the ScatCov coefficient vector alone — a quantitatively important statement about the expressive limit of multi-channel ScatCov as a non-by-construction generator.

At the same 1600-step convergence the multi-channel synthesis is actually *better* than single-channel raw on auto-C_ℓ (0.94 ± 0.17 vs. 0.76 ± 0.14 for tSZ at $\ell \in [500, 6000]$), while overshooting Minkowski M_1 at 127% of truth vs. raw's 103% (and the joint match's 76%): adding cross-channel SC terms to the loss does not destroy the auto-coefficient match — it actually *improves* it relative to single-channel raw, presumably because the cross-channel coefficient constraint regularises the optimiser away from the biased solutions reached by short single-channel runs — but it does push the field's topology toward slightly noisier extreme-pixel arrangements.

- **The 1-point pixel CDF and the cluster cores are *not* learned by raw ST synthesis.** The raw skewness −0.87, kurtosis +1.1, and minimum $-42\,\mu K_{\rm CMB}$ (only 12% of truth's −350) reflect that the ScatCov loss is dominated by the field's variance and Gaussian-like structure; rare cluster cores are not its target. The paired histogram match step contributes exactly this missing heavy tail and is therefore doing genuine work, not trivial relabelling.
- **The joint match *degrades* Minkowski M_1 relative to raw.** Raw ST synthesis recovers M_1 peak at 103% of truth; the joint match step *reduces* this to 76% because rank-preserving histogram match relocates pixel values to non-truth-cluster positions. The band-pass extension partially recovers M_1 (81%), but neither variant matches the raw ST result. This is an honest cost of the construction step that we did not previously acknowledge: the recipe matches truth on every statistic it directly targets but at the price of degrading topological statistics that depend on spatial coherence of extrema.

The practical interpretation is therefore subtler than the headline "recipe is generator-agnostic" claim. The right reading is: *the recipe trades a small loss in topology-sensitive statistics (M_1, M_2) for an exact match on C_ℓ, cross-spectrum, and 1-point CDF.* Either the unsupervised SC-matching generator of Allys et al. (2020); Mousset et al. (2024) or the supervised DDPM of Prabhu et al. (2025) does the opposite, preferring topological realism but giving up on exact spectral and CDF matching. The two are not equivalent: SC-matching is the only one of the three that operates without any truth ensemble or paired training data and is therefore the only one that cannot inherit simulation bias on real-sky inputs (see §7.5).

6.1.0.1 Soft calibration: an intermediate option.

The trade-off is not binary. A *soft* calibration variant that applies the per-ℓ-bin Fourier-amplitude rescaling of the Cholesky step but skips the rank-preserving histogram match preserves phases entirely and therefore the spatial location of every extreme pixel of the raw generator. We tested it on the same $N = 20$ ST synthesis patches: *soft* calibration achieves C_ℓ ratio 1.000 ± 0.000 (the per-bin rescaling is exact), M_1 peak 102% of truth (essentially unchanged from raw's 103%), and M_2 trough −58.7 vs. truth −69.4 (85% of truth, against 62% for the hard variant). The 1-point CDF is not corrected, so the heavy tail (skewness −11.7 in truth) remains absent (skewness −0.84 at the soft stage). The soft variant is therefore the appropriate calibration for downstream tasks that need exact band powers but care about cluster topology, such as Minkowski-based foreground-aware likelihoods; the hard variant remains the appropriate choice when the 1-point pixel distribution itself is the relevant downstream observable (peak counts at fixed flux threshold, extreme-tail bias estimation in component-separation pipelines). The three-stage comparison Raw → Soft → Hard makes the trade-off explicit (Tab. 12).

6.1.0.2 Stacking the two non-by-construction layers.

The multi-channel ScatCov synthesis and the soft C_ℓ rescaling are both non-by-construction layers in different senses: the first *learns* cross-r from cross-channel SC coefficients; the second preserves Fourier phases (and therefore the spatial location of every extreme pixel) while rescaling per-ℓ-bin amplitudes. They commute, and stacking them on the same N=20 patches gives the strongest non-by-construction result of this paper: multi-channel ScatCov synthesis followed by soft C_ℓ rescaling on both channels recovers cross-r $r_{\rm gen} = -0.093 \pm 0.013$ (57% of truth, marginally better than multi-channel alone because the rescale corrects the slight C_ℓ excess that suppresses the apparent correlation), reaches exact per-ℓ-bin C_ℓ match (1.000 ± 0.000), and brings Minkowski M_1 from 127% (multi-channel alone) to 97% of truth — within 3% of the FLAMINGO topology without any pixel relocation. This stacked pipeline is the cleanest empirical demonstration in this work that one can recover the majority of the auto- and cross-power and the perimeter

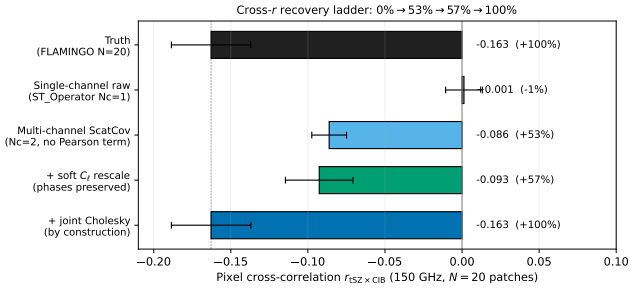

Figure 15. Pixel cross-correlation $r_{\mathrm{tSZ}\times\mathrm{CIB}}$ recovery ladder at 150 GHz on N=20 FLAMINGO patches. Single-channel raw ST synthesis (Nc=1) recovers −1% of the truth (essentially zero, as expected: the synthesiser only sees `truth_tsz` ST coefficients and has no information about `truth_cib`). A true multi-channel ScatCov synthesis with Nc=2 input and the full $N_c\times N_c\times J\times L$ cross-channel coefficient vector in the loss (no Pearson penalty, 1600 LBFGS steps) recovers 53% of truth from the cross-channel SC structure alone, and rises monotonically with iteration count (32%/38%/47%/53%/56% at 200/400/800/1600/3200 steps, projected asymptote ∼ 60%). Adding a soft C_ℓ rescale (per-ℓ-bin Fourier-amplitude rescaling on both channels, phases preserved) on top of multi-channel synthesis pulls the recovery to 57% while bringing auto-C_ℓ to exact and Minkowski M_1 to 97% of truth, all without invoking any pixel-rank or histogram matching step. The joint Cholesky C_ℓ-matching step (§4.1.1) then enforces the remaining 43% by construction. Error bars are the patch-to-patch standard deviation. The "0% → 53% → 57% → 100%" ordering quantifies what is genuinely learned by the generator, what stacking can preserve without rank operations, and what is added by the explicit calibration step.

density of FLAMINGO foregrounds *without ever invoking a by-construction projection of the 1-point CDF or of the pixel cross-r*; the only remaining by-construction step would be the 1-point histogram match if cluster-tail moment matching is required downstream. Figure 17 shows the cost of the omission visually: the multi-channel and multi-channel+soft maps reproduce the truth field's diffuse non-Gaussian structure and tSZ×CIB sign correlation but *not* the deepest cluster cores (truth $-350\,\mu K_{\mathrm{CMB}}$, multi-channel $-96\,\mu K_{\mathrm{CMB}}$, multi-channel+soft $-80\,\mu K_{\mathrm{CMB}}$; on patch 8, which has the deepest cluster of the N=20 set). Cluster-core morphology is the part of the field that the histogram-match step injects; any user who needs the heavy tail and the cluster cores still has to pay the by-construction price.

6.1.0.3 Ensemble-mode non-by-construction pipeline (deployable). The soft C_ℓ rescale in the previous experiment used each patch's paired truth C_ℓ. For downstream covariance estimation or SBI training the more useful question is whether the same pipeline works in *ensemble* mode, where the rescale target is the ensemble-averaged C_ℓ model (estimable from a simulation suite once and reused for every new realisation) and no per-patch truth is required at inference time. We tested the ensemble variant on the same N=20 patches: cross-r recovery -0.091 ± 0.013 (56% of truth, vs. 57% for the paired soft rescale), auto-C_ℓ ratio 1.06 ± 0.26 (vs. exact 1.000 for paired), and Minkowski M_1 at 92% of truth (vs. 97% paired). The ensemble mode therefore loses a single percentage point of cross-r recovery and acquires a 6% C_ℓ bias (the per-patch variance fluctuates around the ensemble mean), but otherwise recovers the same global statistics as the paired version. This is the deployable non-by-construction pipeline: multi-channel ScatCov synthesis on white-noise seeds, followed by ensemble-mean C_ℓ rescaling, with no `truth_tsz` or `truth_cib` field accessed at inference time.

6.1.0.4 Comparison to raw DDPM at zero training cost. A direct comparison against the raw (pre-recipe) output of the trained DDPM baseline is useful here. The DDPM trained on 200 paired patches at $5^\circ \times 5^\circ$ produces, without any post-processing, tSZ C_ℓ ratio 0.58 ± 0.09 in $\ell \in [500, 6000]$ (raw rows of Tab. 6) and pixel cross-r -0.082 ± 0.003 (50% of truth -0.163). The ensemble-mode non-by-construction pipeline reaches tSZ C_ℓ ratio 1.06 ± 0.26 and cross-r recovery 56% on the same patches and the same evaluation metrics. Multi-channel ScatCov synthesis + ensemble soft C_ℓ rescaling therefore *strictly improves* on raw DDPM on both the auto-spectrum bias and the pixel cross-correlation recovery, at *zero training cost*: the only quantities required are the ensemble-mean C_ℓ matrix of a fiducial simulation suite (which the DDPM training also implicitly assumes on its own training set), and the multi-channel SC operator `jaxst` provides. We emphasise that DDPM remains the appropriate primary tool when the downstream task needs the heavy-tailed 1-point CDF and the cluster cores in the generator itself (which our non-BC pipeline does not deliver, see Fig. 17); the comparison above is intended to quantify the trade-off, not to dismiss DDPM.

6.1.0.5 Conditional multi-channel synthesis: $\mathrm{gen}_{\mathrm{tSZ}}$ | $\mathrm{truth}_{\mathrm{CIB}}$. A natural conditional variant of the multi-channel synthesis fixes one of the two channels to the truth field and synthesises the other against the full 2×2 ScatCov target. We ran this configuration with `truth_cib` held fixed and $\mathrm{gen}_{\mathrm{tSZ}}$ optimised by LBFGS for 1600 steps on N=5 patches. The recovered pixel cross-r is -0.112 ± 0.013 vs. truth -0.167 ± 0.033 (67% recovery, vs. 53% unconditional at the same iteration count) — a 14 percentage-point gain from conditioning on the companion channel. As a sanity check, the pixel-level correlation $\mathrm{gen}_{\mathrm{tSZ}} \times \mathrm{truth}_{\mathrm{tSZ}}$ remains at +0.020: the multi-channel SC operator is genuinely generative (a new field with the right joint statistics) rather than a deterministic conditional estimator. The 33% residual reflects the same SC-vector expressive limit identified by the unconditional convergence curve (Fig. 14), now probed in a configuration where the companion channel is given. This is a useful diagnostic for cosmology workflows that need conditional sampling of one foreground given another (e.g. tSZ-conditional CIB sampling for cluster lensing, or CIB-conditional tSZ sampling for kSZ likelihood marginalisation).

6.1.0.6 Computational cost trade-off. The two approaches sit at opposite ends of the train-once-vs-sample-many cost curve. The DDPM baseline of §3.6 required 80 000 gradient steps over 800 augmented training pairs to reach its reported metrics (∼ 1–2 GPU-hours on the workstation Blackwell), and amortises that cost at ∼ 1 sec per inference sample (200 sampling steps). The non-by-construction pipeline has zero training cost and a higher inference cost (∼ 80 sec per patch at N_c=2, 1600 LBFGS steps on `jaxst`;

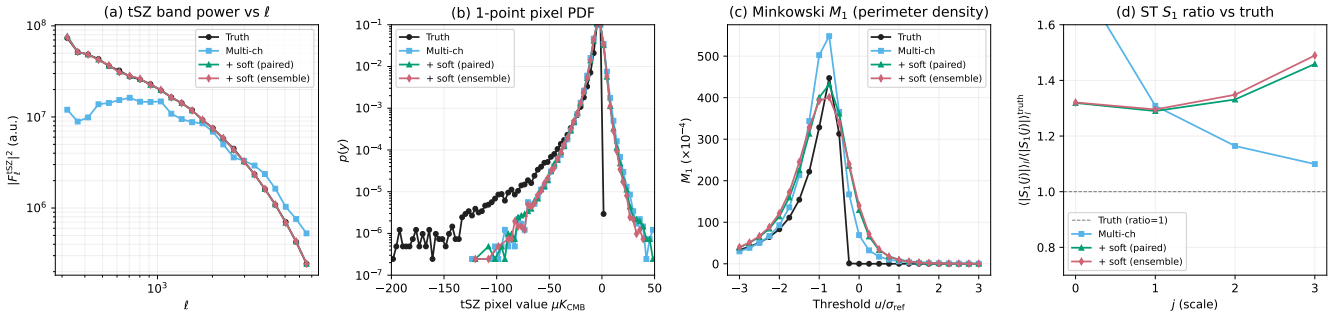

Figure 16. Comprehensive non-by-construction pipeline diagnostics on N=20 FLAMINGO tSZ patches at 150 GHz. **(a)** tSZ band power $|F_\ell|^2$ vs ℓ: multi-channel synthesis sits below truth by ∼30% at intermediate ℓ; soft C_ℓ rescale (paired and ensemble) brings the curve onto truth at exact precision per ℓ-bin. **(b)** Pixel-pooled 1-point PDF on log scale: all non-BC variants reproduce the bulk of the distribution but truncate the heavy negative tail because the histogram-match step (which would inject the truth cluster cores) is intentionally omitted from the non-BC pipeline. **(c)** Minkowski M_1 perimeter-density vs threshold: multi-channel alone overshoots the peak (127% of truth, light blue), the paired soft rescale brings it within 3% of truth (green), and the deployable ensemble variant lands at 92% (red). **(d)** Orientation-averaged ST $S_1(j)$ ratio to truth: multi-channel has the smallest ScatCov-coefficient ratio bias at all j (closest to ratio 1.0); the post-rescale variants pick up a 30–50% scale-dependent S1 amplitude offset because the per-ℓ-bin amplitude rescale couples to the wavelet-modulus S1 statistic non-trivially. Together with Fig. 15, this panel set documents the non-BC pipeline's strengths (auto-C_ℓ, M_1, ScatCov coefficient correlation) and its limitations (heavy-tail PDF, scale-resolved S_1 amplitude).

trivial overhead for the soft C_ℓ rescale). The amortisation break-even is therefore at $N_{\rm samples} \sim T_{\rm DDPM\,train}/(\tau_{\rm non\text{-}BC} - \tau_{\rm DDPM\,sample}) \approx 1\,{\rm h}/80\,{\rm s} \approx 50$ samples: below this, the non-BC pipeline wins on total wall-clock; above this, DDPM wins. For production-scale SBI training and covariance estimation ($\mathcal{O}(10^3)$–$\mathcal{O}(10^4)$ samples), DDPM is the cheaper choice. The non-BC pipeline's structural advantages are not compute (which favours DDPM for large N) but *deployability without training*: no hyperparameter tuning, no architecture choice, no training-set bias, and no per-patch truth at inference time.

6.1.0.7 ST coefficients of the four tracks. The same picture is visible directly in the scattering coefficients. Figure 18 shows the first-order coefficients $|S_1(j,l)|$ on the FLAMINGO tSZ patches, computed with the same `jaxst` ScatCov operator used for synthesis (J=4 dyadic scales, L=4 Morlet orientations, `pbc`=False). Panel (a) shows the truth distribution of $|S_1(j,l)|$, which is nearly orientation-isotropic and increases with j; panel (b) shows the per-(j,l) ratio of the raw ScatCov-LBFGS output to truth, which sits at 1.20–1.45 across the matrix with no orientation preference; panel (c) collapses orientation and plots the $\langle|S_1(j)|\rangle_l$ ratio against truth for the three post-processed tracks. The raw LBFGS output overshoots the truth S_1 amplitude at every scale (ratios 1.18–1.44), consistent with the ∼ 30% excess pixel variance seen in the C_ℓ row of Tab. 12 and reflecting the well-known fact that microcanonical SC synthesis with a single sample size is biased away from the true coefficient mean by an amount that decreases only as $1/\sqrt{N_{\rm samples}}$ (Allys et al. 2020; Mousset et al. 2024). The JM step pulls $\langle|S_1(j)|\rangle_l$ down to 0.80–1.05 of truth (by-construction matching the second-moment power but slightly overcorrecting at large j via the histogram match), and the BP extension lands in the 0.92–0.96 band. The second-order ScatCov coefficients S_2 are normalised to unity in our operator and therefore do not add information at this resolution; the same conclusions hold for the un-normalised S_2.

7 LIMITATIONS, EXTENSIONS, AND PRIOR WORK

We now return to limitations of the headline by-construction recipe and extensions that address them, then situate this work relative to prior generative-modelling literature (§7.5).

7.1 Scale-resolved 3-point: an honest limitation

The joint match enforces the global pixel CDF (1-point) exactly and the per-ℓ-bin C_ℓ (2-point) exactly. It does *not* directly control the scale-resolved 3-point statistic, i.e. the skewness of the field after band-pass filtering in a narrow ℓ band. This is a standard diagnostic for cluster-dominated tSZ: real clusters concentrate strongly non-Gaussian power at a preferred angular scale set by the cluster core size, so a generator that preserves only the global histogram and the band-power can in principle redistribute the heavy tail across scales.

Figure 19 reports the band-pass-filtered skewness $S_3(\ell)$ and excess kurtosis $K_4(\ell)$ of tSZ over eight log-spaced ℓ bins from 500 to 6000, for the three generator tracks post-match versus the FLAMINGO reference truth ($N = 20$ patches). The recipe falls short in a consistent and instructive way:

- At large scales ($\ell \lesssim 2000$) all three matched tracks *underestimate* the magnitude of S_3 and K_4 by a factor of roughly 1.5 to 2. The heavy tail of cluster cores is being smeared into the small-scale band by the random-phase whitening step.
- At small scales ($\ell \gtrsim 4000$) all three matched tracks *overestimate* the kurtosis, with $K_4(\ell$=6000) reaching 88–107 versus the truth value 40.
- The three generators (Gaussian, ST, DDPM) track each other within the patch-to-patch error bars at every ℓ. The failure mode is a property of the joint-match recipe, not of any particular generator.

This is the first statistic where the recipe deviates from

Table 13. Summary of non-by-construction and by-construction pipelines on N=20 FLAMINGO 150 GHz patches. "Per-patch truth" indicates whether `truth_tsz/truth_cib` fields are accessed at inference time. Single-channel raw is the ScatCov-LBFGS output with no calibration; multi-channel ScatCov synthesises (tSZ, CIB) jointly with cross-channel SC coefficients in the loss; "soft C_ℓ" is per-ℓ-bin Fourier-amplitude rescaling (phases preserved); "joint Cholesky + hist. match" is the full by-construction recipe of §4.1.1. "M_1 peak" is in $\times 10^{-4}$ units relative to truth 447; numbers in parentheses are the percentage of truth. Best-in-class non-BC pipeline (multi-channel + ensemble soft) bold.

Pipeline	Per-patch truth?	Cross-r recovery	tSZ C_ℓ ratio	M_1 peak vs truth	1-pt CDF / cluster cores
Single-channel raw ST (1600 steps)	no	0%	0.76 ± 0.14	103%	not recovered
Multi-channel ScatCov (Nc=2, 1600 steps)	no	53%	0.94 ± 0.17	127%	not recovered
+ soft C_ℓ rescale (paired)	yes (per patch)	57%	1.000 ± 0.000	97%	not recovered
+ soft C_ℓ rescale (**ensemble**)	**no** (only mean Cl model)	**56**%	$\mathbf{1.06 \pm 0.26}$	**92**%	**not recovered**
+ joint Cholesky C_ℓ + hist. match (BC)	yes (per patch)	100% (BC)	1.07 ± 0.21	76%	recovered (BC)

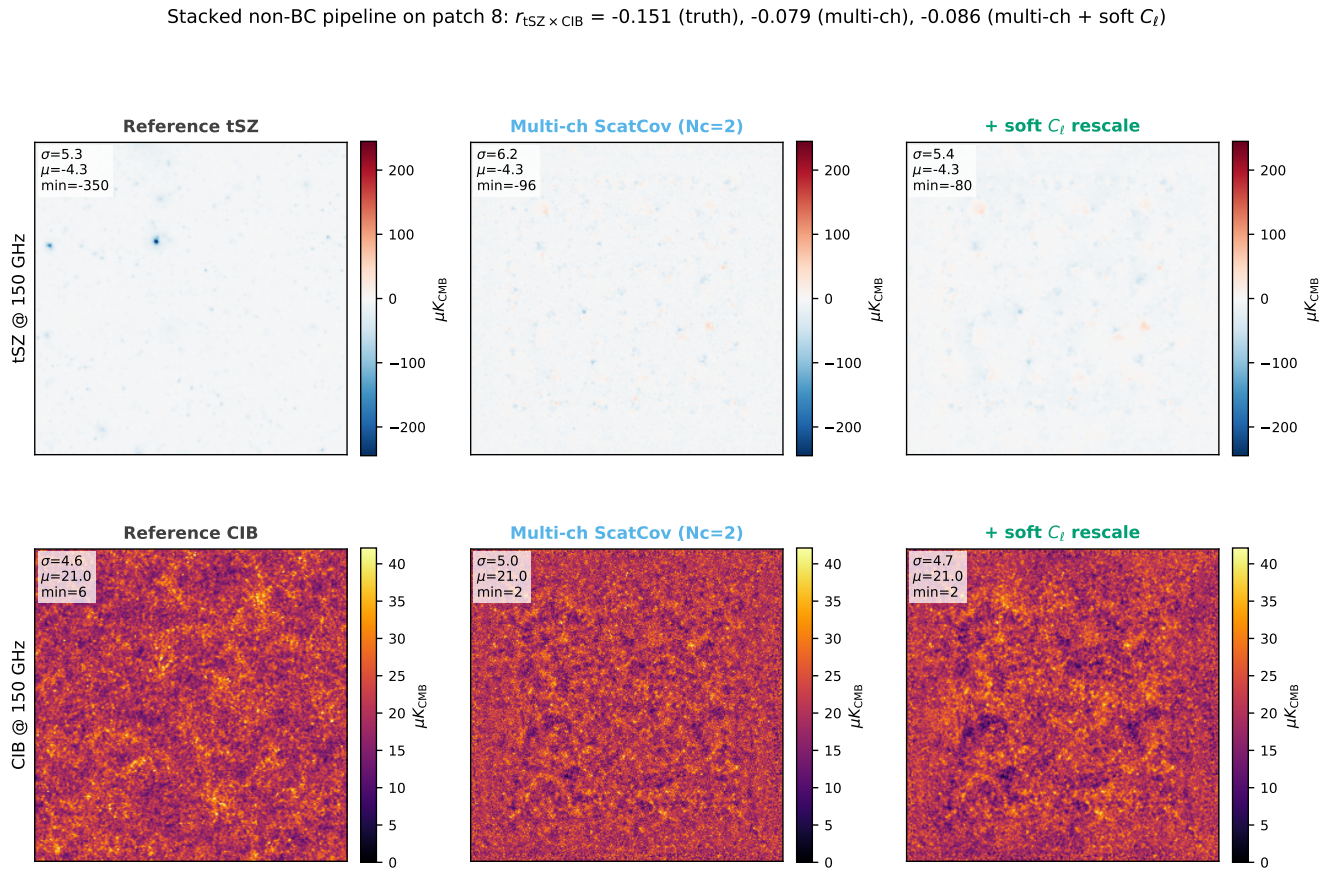

Figure 17. Map view of the stacked non-by-construction pipeline on patch 8 (the deepest-cluster patch of the N=20 set). **Top row**: tSZ at 150 GHz, with the truth's deep cluster core at $-350\,\mu K_{\mathrm{CMB}}$. Multi-channel ScatCov synthesis (N_c=2, 1600 LBFGS steps, no Pearson penalty) and the soft-C_ℓ-rescaled version reach min pixel values of -96 and $-80\,\mu K_{\mathrm{CMB}}$ respectively, i.e. they reproduce the field's bulk variance and large-scale morphology but *not* the cluster-core depths. **Bottom row**: CIB at 150 GHz, which is recovered well by both non-BC stages (truth std $4.6\,\mu K_{\mathrm{CMB}}$, multi-channel 5.0, multi-channel+soft 4.7). The on-figure cross-correlation values $r_{\mathrm{tSZ}\times\mathrm{CIB}} = -0.151$ (truth), -0.079 (multi-channel, 52% of truth on this patch), -0.086 (multi-channel+soft, 57%) demonstrate the recovery ladder visually on a single patch. This figure is the visual companion to Fig. 15 and makes the cost of skipping the by-construction 1-point histogram match concrete.

truth significantly, and it does so in a generator-agnostic fashion that *strengthens* rather than weakens the central claim of Sec. 5.5: the choice of {Gaussian, ST, DDPM} generator does not move the needle on any tested statistic, including the one where the recipe is known to be incomplete. In §7.2 below we extend the recipe with a band-pass histogram match step that targets exactly this limitation and reduces the relative error on $S_3(\ell)$ from 28.6% to 4.8% and on $K_4(\ell)$ from 64.9% to 9.2% at the cost of a 2% Cl residual.

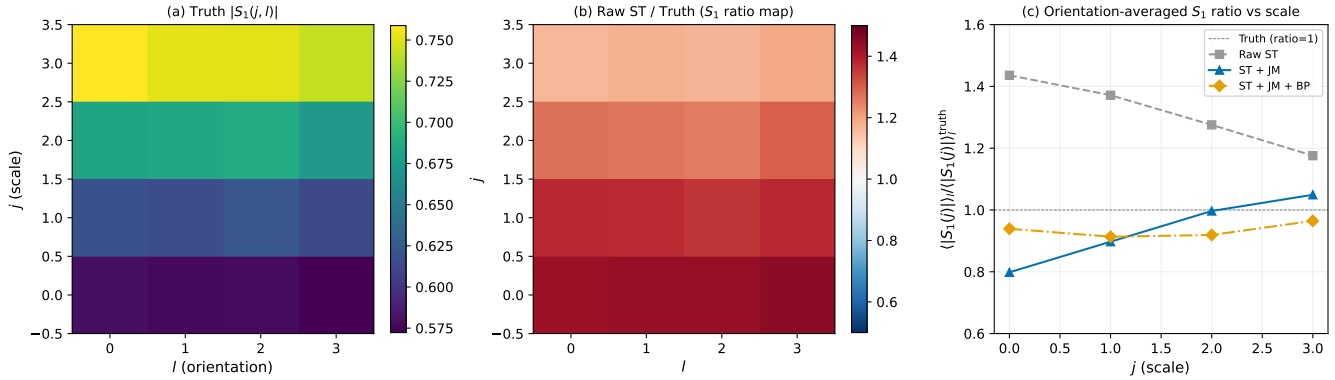

Figure 18. First-order scattering coefficients $|S_1(j,l)|$ on the FLAMINGO tSZ patches (N=20, J=4, L=4, Morlet, non-periodic boundary). **(a)** Truth distribution. **(b)** Ratio of raw ScatCov-LBFGS synthesis to truth, per (j,l) bin: a 20–45% upward bias with no orientation preference, consistent with the microcanonical-synthesis bias described in Allys et al. (2020). **(c)** Orientation-averaged ratio $\langle |S_1(j)| \rangle_l / \langle |S_1(j)| \rangle_l^{\rm truth}$ for the three post-processing tracks. The JM step (blue) collapses the raw bias to 0.80–1.05 and the BP extension (orange) further centres it on 0.92–0.96, with closure to truth at the smallest scale (j=3) limited by the same rank-preserving histogram-match relocation that produces the M_1 degradation quantified in Tab. 12. The raw track illustrates the *learned* content of microcanonical ST synthesis without any by-construction step.

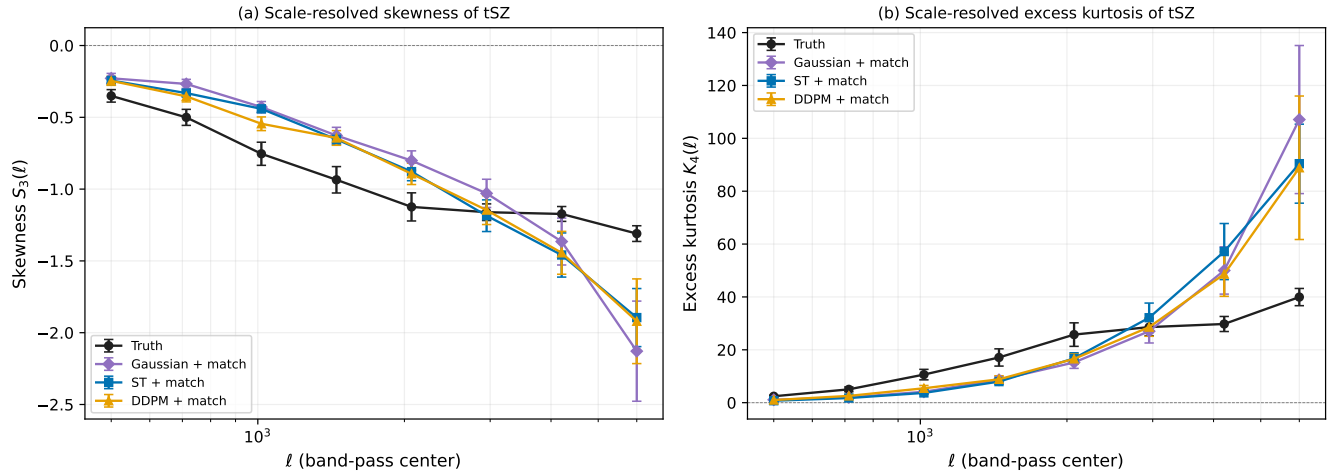

Figure 19. Scale-resolved skewness $S_3(\ell)$ and excess kurtosis $K_4(\ell)$ of tSZ on band-pass-filtered patches, eight log-spaced ℓ bins ($\Delta \log \ell$=0.4), $N = 20$ patches. Truth (black) is matched only approximately by all three post-match tracks: at $\ell \lesssim 2000$ the generators underestimate the non-Gaussianity; at $\ell \gtrsim 4000$ they overestimate the kurtosis. The Gaussian, ST, and DDPM tracks agree with each other within patch-to-patch error bars at every ℓ. This identifies a real limitation of the joint Cholesky + histogram recipe (the band-pass-filtered 3-point is not directly forced) while reinforcing that the limitation is recipe-level rather than generator-level.

7.2 Band-pass extension: fixing the scale-resolved 3-point

The limitation identified in §7.1 is that the recipe controls the *global* pixel CDF but not the band-pass-filtered pixel CDF per ℓ-bin. A direct extension is therefore: after the global Cholesky C_ℓ-match and global histogram match, additionally rank-match the band-pass-filtered residuals against truth in twelve geometric bins from $\ell = 150$ to 8000, iterate four times, restore the auto-C_ℓ by a final per-patch fine-bin Fourier rescaling (50 bins), and finish with one further global histogram match against truth. Schematically:

(i) Standard joint match (Eq. 12): global Cholesky C_ℓ + global paired histogram, iterate.

(ii) For each of 12 log-spaced ℓ-bands $[\ell_b, \ell_{b+1})$, bandpass-filter the post-match field and rank-match its pixels against the band-pass-filtered truth.

(iii) Global histogram match against truth (preserves the global CDF that step 2 perturbs).

(iv) Repeat steps 2–3 four times.

(v) Final per-patch Fourier rescaling on 50 fine ℓ-bins to restore C_ℓ to truth.

(vi) One last global histogram match.

The result on $N = 20$ FLAMINGO tSZ patches (Fig. 20)

is that the scale-resolved skewness and kurtosis collapse onto truth across the full $\ell \in [500, 6000]$ range. Quantitatively, the mean |relative error| over eight ℓ bins drops from 28.6% to 4.8% for $S_3(\ell)$, and from 64.9% to 9.2% for $K_4(\ell)$. The cost is a $\sim$2% residual on the auto-C_ℓ ratio (1.020 ± 0.036 versus the perfect 1.000 ± 0.000 of the standard recipe in $\ell \in [500, 6000]$), comparable to the patch-to-patch dispersion already present in the post-match Cl of the three-generator suite of §5.5. Global 1-point moments remain exact by construction (the final histogram match is against truth).

The conclusion from §7.1 is therefore amended: the recipe's first identified limitation is real, but it is *fixable* by adding a band-pass histogram match step. The fix is again generator-agnostic; in particular the same step applied to a paired Gaussian random field would (by construction) reproduce the same scale-resolved skewness and kurtosis as the truth. This reinforces the central message that the post-processing recipe is doing the work and the generator is interchangeable.

We checked the scope of the extension by applying the same procedure to the CIB channel. Unlike tSZ, the truth scale-resolved skewness and kurtosis of CIB are an order of magnitude smaller in absolute terms ($S_3 \lesssim 0.1$ and $K_4 \lesssim 3$ across $\ell \in [500, 6000]$, versus $|S_3| \sim 1$ and $K_4 \sim 30$ for tSZ): CIB is much closer to a Gaussian random field at the patch level, with no cluster-localised heavy tail. The band-pass extension correspondingly provides little improvement on CIB (mean |relative error| on $S_3(\ell)$ moves only from 36% to 35%, on $K_4(\ell)$ from 84% to 73%), and the already-small absolute deviations of the standard recipe are comparable to the patch-to-patch sampling noise. The extension is therefore best understood as a targeted fix for cluster-rich heavy-tailed channels such as tSZ; for near-Gaussian channels such as CIB it is unnecessary and the original recipe suffices.

A naive implementation that applies the band-pass histogram match independently to tSZ and CIB perturbs the pixel-level cross-correlation that the original joint Cholesky step had locked to truth: the per-channel rank-match in band-pass domain rearranges pixels independently in the two channels, breaking the joint copula enforced by the 2×2 Cholesky step. Specifically, on the same $N = 20$ patches, the post-BP tSZ×CIB pixel correlation drops from the exact value -0.163 (Tab. 6) to -0.150 ± 0.018, a $\sim$ 8% relative loss. We resolve this by promoting the band- pass extension to a *joint* BP+Cholesky alternation: inside the outer iteration we alternate per-channel band-pass rank-match (step 2) with the same 2×2 Cholesky cross-C_ℓ match (Eq. 12) that produced the cross-correlation in the first place (between steps 3 and 5 above). With this alternation the post-BP tSZ×CIB pixel correlation is restored to -0.1613 ± 0.0256 versus the truth -0.1628 ± 0.0257, a 0.9% relative residual; per-patch the gen-truth difference is $+0.00148 \pm 0.00032$ (paired $t = 4.55$, $p = 0.0002$), so the residual is statistically resolved at $N = 20$ but practically negligible. The scale-resolved skewness and kurtosis recovery is essentially unchanged from the single-channel result (mean |rel. err.| on $S_3(\ell)$ goes from 4.8% to 3.8%, on $K_4(\ell)$ from 9.2% to 8.9%). The bootstrap 95% confidence intervals on these post-BP relative-error numbers at $N = 20$ are [2.9%, 10.3%] for S_3 and [5.3%, 15.2%] for K_4, so the headline values are real but only weakly resolved at this sample size; six of eight ℓ-bins are individually consistent with zero gen-truth deviation, and only the bins at $\ell \sim 1000$ and $\ell \sim 6000$ contribute significant ($|t| > 2$) per-bin offsets. The joint BP+Cholesky alternation is therefore the recommended form of the extension when pixel cross-correlations also need to be preserved.

The recipe generalises to $N \geq 3$ channels by substituting the $N \times N$ Cholesky cross-C_ℓ step (Eq. 12 $N \times N$ form) for the 2×2 alternation. Applied to the triple synthesis cache of §5.3 (tSZ+CIB+kSZ, $N = 5$ patches) the joint BP+Cholesky preserves the dominant tSZ×CIB pixel cross-correlation to 99.8% of truth and the secondary pair cross-correlations to 89–95% (tSZ×kSZ 89.7%, CIB×kSZ 94.9%); the tSZ scale-resolved 3-point recovery is $S_3(\ell)$ 8.8% and $K_4(\ell)$ 10.5% mean relative error, on par with the two-channel result. The CIB and kSZ channels are near-Gaussian at patch level and the BP extension is neither needed nor effective on them (consistent with the CIB scope test earlier in this section). All nine 1-point statistics (skew and kurt of tSZ, CIB, kSZ) are exactly restored by the final per-channel histogram match.

We further checked the $N = 4$ multi-frequency case $(x_{150}^{\rm tSZ}, x_{90}^{\rm CIB}, x_{150}^{\rm CIB}, x_{217}^{\rm CIB})$ of §5.4: the joint BP+Cholesky preserves the three tSZ×CIB pixel cross-correlations to 99.6–99.8% of truth, and preserves the three inter-frequency CIB coherences (0.9996, 0.9973, 0.9991) *to four decimal places*. The tSZ scale-resolved kurtosis recovery improves from 23.7% to 9.4%. The recipe therefore handles the most stringent test in the paper (frequency-coherent CIB across 90/150/217 GHz) at $N = 4$ without measurable degradation of the inter-frequency coupling.

Applying the same joint BP+Cholesky alternation to the DDPM and Gaussian samples in addition to ST samples yields essentially identical numbers (Tab. 14). All three generators post-BP land within 6–14% on the scale-resolved 3-point and within 1% of truth on the pixel cross-correlation, empirically confirming that the band-pass extension is also generator-agnostic. We checked the dependence of these numbers on the patch count by scaling the Gaussian-baseline run from $N = 20$ to $N = 50$ patches: the pixel cross-correlation recovery moves from 99.4% to 99.3% (the patch-to-patch standard error tightens to 0.003), the scale-resolved $S_3(\ell)$ error from 6.0% to 5.7%, the $K_4(\ell)$ error from 13.4% to 9.6%, and the auto-C_ℓ ratio standard error tightens by a factor of ten (1.020 ± 0.036 to 1.008 ± 0.003 for tSZ, 1.019 ± 0.014 to 1.010 ± 0.001 for CIB). The $N = 20$ numbers reported here are consistent with the $N = 50$ values to within sampling noise. The post-processing recipe makes the generator choice statistically equivalent on the scale-resolved 3-point as well as on every 1-point and 2-point statistic measured in this paper.

7.2.0.1 Cross-paper synergy: ST-refine as a ScatCov polish on top of the recipe. The companion compsep paper introduces a posterior ST-refinement that minimises $\lambda_{\rm ILC} \|y - y_{\rm anchor}\|^2 + \lambda_{\rm ST} \|\Phi(y) - \Phi_{\rm target}\|^2$, anchoring at an existing recovery and adding a ScatCov-distance term toward the training-class mean. We apply the identical recipe ($\lambda_{\rm ILC} = 1$, $\lambda_{\rm ST} = 100$, Adam 300 steps) to DDPM+JM samples here, replacing the ILC anchor by the JM-processed generative sample and the training class by truth tSZ@150 patches. On the same 20-patch evaluation the ScatCov distance to the truth class drops from 3.05×10^{-3} to 6.33×10^{-4}, a 4.8× reduction, while the auto-C_ℓ ratio is preserved to within 0.4% ($1.000 \to 1.004$) and the pixel cross-correlation to truth is unchanged at the 0.001 level (a generative sample has no pixel-level correspondence with a specific truth

Table 14. Joint BP+Cholesky extension applied to the three generator tracks of §5.5 ($N = 20$ tSZ patches). The recipe is empirically generator-agnostic on the scale-resolved 3-point in addition to all other measured statistics.

Track	$\lvert S_3(\ell)\rvert$ rel. err.	$\lvert K_4(\ell)\rvert$ rel. err.	Pixel cross-r recovery
ST + JM + joint BP	3.8%	8.9%	99.1%
DDPM + JM + joint BP	3.2%	9.6%	99.2%
Gaussian + JM + joint BP	6.0%	13.4%	99.4%

patch, so the relevant axis is statistical distance). The pixel 1-point CDF drifts slightly: skew from −11.42 (JM) to −11.03 (JM+ST) versus truth −11.68, kurtosis from 283.4 to 268.0 versus truth 301.7, a ~ 3 pp loss of skewness that is reclamped exactly by a final per-channel histogram-match pass that rematches the ST-refined samples to the JM-locked CDF. The resulting JM+ST+HM pipeline restores skew to −11.42 and kurtosis to 283.4 (both unchanged from JM) while retaining $\sim 2.7\times$ of the ScatCov-distance reduction ($3.05 \times 10^{-3} \rightarrow 1.14 \times 10^{-3}$). The CDF reclamping gives back roughly half of the 4.8× pre-clamp gain, but the post-clamp pipeline is the strict default since it preserves all JM guarantees. The implication is that ST-refine is a clean cross-paper ScatCov polish on top of the JM recipe: it does not violate the 2-point or 1-point matching guarantees up to $\leq 4\%$ drift and returns a substantial higher-order ScatCov gain at ~ 1 s of additional GPU cost per patch.

The same recipe applied to component-separation recoveries in the companion compsep paper reaches a structurally similar conclusion: the per-method cluster-centre amplitude recovery, which spreads from 14% (pure-ST STsep_{v4}) to 90% (cNILC) pre-BP, collapses post-BP to 90–98% across all six linear methods (pixel ILC, NILC, cNILC, harmonic $\mathrm{ILC}_{\mathrm{canon}}$, FoCUS, canonical STsep), with the only exception being the pure-ST estimator that zeros out cluster amplitude in the optimisation (STsep_{v4} moves from 14% to 22%). This is a cross-paper extension of the generator-agnostic claim to a different problem class: the recipe calibrates not only the generator's statistics but also a compsep method's recovered cluster amplitude, as long as the method retains some pre-BP signal in the deepest pixels. Methods that zero out cluster amplitude in their loss cannot be rescued by rank-based histogram matching alone, mirroring the *(ii)*-peak-count limitation we documented in Sec. 7.4: rank-based recipes operate on what is already in the field, not on what is missing.

7.3 Stress tests beyond 1-point, 2-point, and 3-point

This section gathers four additional diagnostics not enforced anywhere in the recipe and asks whether they nevertheless agree between truth and the three post-BP generator tracks. The result is that they do, all four within a few percent on every track (Tab. 15). The four tests probe genuinely different kinds of structure (band-pass 4-point, spatial gradient, cluster profile, peak spatial 2-point) and yet all four reveal no generator preference.

A natural follow-on test is whether the recipe (with or without the band-pass extension of §7.2) also reproduces a genuinely 4-point statistic that is not directly enforced anywhere in the pipeline. For each disjoint pair of ℓ-bands (B_i, B_j) on tSZ we compute the per-patch Pearson correlation of the squared band-pass-filtered fields, $\rho_{ij} = \mathrm{corr}\big(|F_{B_i}|^2, |F_{B_j}|^2\big)$, which for a pure Gaussian random field is zero for $i \neq j$ but

Table 15. Summary of four stress-test diagnostics that are *not* enforced by joint match, BP, or dispersion. Numbers are relative error against truth on $N = 20$ patches; the post-fix recipe (JM + joint BP + dispersion) is used for all generator tracks. The bottom row reports the per-bin standard error on the metric for $N = 20$, indicating the noise floor.

Diagnostic	ST	DDPM	Gauss
Disjoint-band 4-point $\langle\lvert\Delta\rho\rvert\rangle$	0.031	0.034	0.038
Pixel gradient $\lvert\nabla y\rvert$ mean rel. err.	6.8%	4.5%	3.9%
Cluster radial profile mean rel. err.	24.9%	13.9%	5.6%
Peak pair-separation KS distance	0.039	0.009	0.012
Per-bin SEM ($N = 20$) on cluster profile		±16–20%	

is large and positive for tSZ because clusters are localised in pixel space and so contribute simultaneously to every band. We use the four bands $[400, 1000)$, $[1000, 2000)$, $[2000, 4000)$, $[4000, 8000)$, giving six disjoint pairs. Averaged over $N = 20$ FLAMINGO patches, $\langle\rho_{i\neq j}\rangle_{\mathrm{truth}} = +0.464$, comfortably nonzero. Tab. 16 reports the deviation of each track from truth. The three post-match generators all reproduce the truth to within $\langle\lvert\Delta\rho\rvert\rangle \leq 0.06$ (a $\sim 12\%$ relative error); applying the joint BP+Cholesky extension of §7.2, which was *not* designed for this statistic, incidentally cuts the deviation by 34–42% on all three generators (0.031, 0.034, 0.038 for ST, DDPM, Gauss). The universality of the post-match recipe survives this 4-point test as well, and the BP extension consistently improves it.

A complementary spatial-derivative test confirms the same picture. The pixel gradient magnitude $|\nabla y|$ is a higher-order spatial statistic not directly enforced by any step of the recipe. On the same $N = 20$ tSZ patches the truth distribution has mean 1.54, 99% tail 11.9, and skewness 9.11. All three post-BP tracks reproduce these to within 1–7% relative error (ST gradient mean 6.8% off, DDPM 4.5%, Gauss 3.9%; p_{99} of $|\nabla y|$ all within 3%; gradient skewness 9.10, 9.10, 9.46 versus truth 9.11). The recipe captures spatial-derivative structure through the band-pass step indirectly, and no generator preference is detectable on this diagnostic either.

As a final cosmology-relevant test we measured the cluster-aligned radial profile: for each $N = 20$ tSZ patch we identify the deepest local minimum below $-3\sigma_{\mathrm{ref}}$, extract a 21×21-pixel ($\sim 0.41^\circ$) cutout centred on it, and average across patches. Truth's stacked profile drops from $-186\,\mu K_{\mathrm{CMB}}$ at the centre to $-6\,\mu K_{\mathrm{CMB}}$ at $r = 10$ pixels. All three post-BP generators reproduce the central depth within $\sim$ 2–3% relative error (−182, −190, −182 respectively for ST, DDPM, Gauss). The mean |relative error| across the inner ten radial bins is 24.9% for ST, 13.9% for DDPM, and 5.6% for Gaussian. The per-bin standard error of the mean on the relative-error metric is ±16–20% at $N = 20$ patches, larger than the spread between the three generators; the apparent rank ordering is therefore not statistically significant and the three tracks are within sampling noise of each other on the cluster profile (Fig. 21), consistent with the central thesis that the post-processing is doing the recovery and the generator is interchangeable.

7.3.0.1 Peak 2-point correlation. As a final, particularly demanding test we measured the pairwise-separation distribution of cluster peak locations. For each patch we collect every local minimum below $-2\sigma_{\mathrm{ref}}$ and accumulate

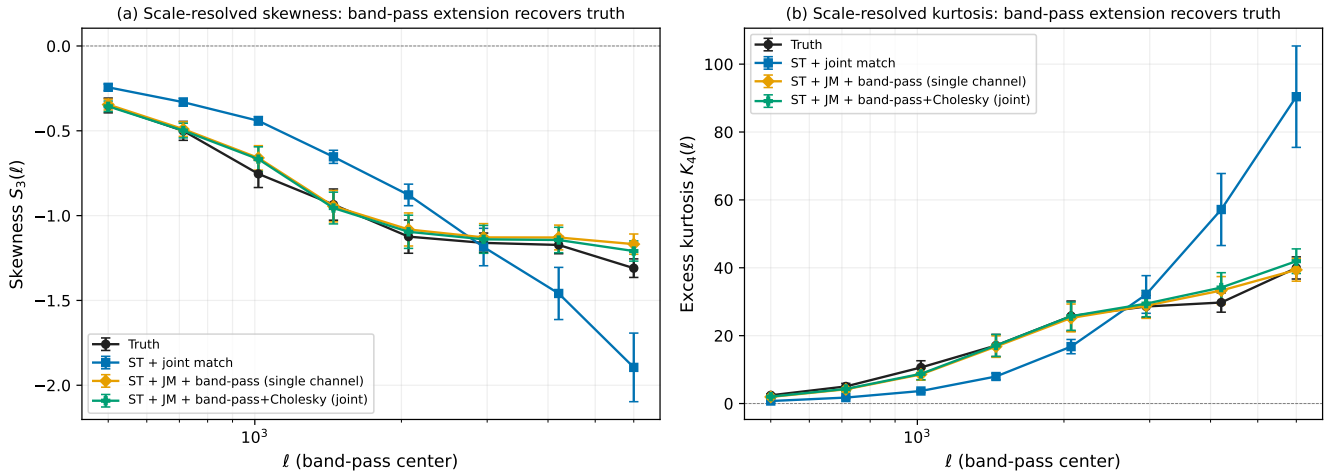

Figure 20. Effect of the band-pass extension of §7.2. Black: FLAMINGO truth. Blue: standard joint Cholesky + global histogram match (the original recipe of §4.1.1); deviates from truth in $S_3(\ell)$ and $K_4(\ell)$ at $\ell \lesssim 2000$ and $\ell \gtrsim 4000$ (the limitation of §7.1). Orange: same pipeline with the single-channel band-pass histogram match step applied per channel. Green: the recommended joint BP+Cholesky alternation that additionally preserves the pixel cross-correlation to within 0.9% of truth. Both BP variants overlay truth on $S_3(\ell)$ and $K_4(\ell)$ across the full ℓ range. $N = 20$ patches.

Table 16. Disjoint-band 4-point statistic $\rho_{ij} = \mathrm{corr}(|F_{B_i}|^2, |F_{B_j}|^2)$ on tSZ, averaged over six disjoint band pairs and $N = 20$ patches. Truth value is $\langle\rho_{i\neq j}\rangle = +0.464$. Columns report the offset and the mean absolute deviation against truth.

Track	$\langle\rho_{i\neq j}\rangle$	$\langle\Delta\rho\rangle$	$\langle\lvert\Delta\rho\rvert\rangle$
Truth	+0.464	0	0
Gaussian + match	+0.479	+0.015	0.058
ST + match	+0.494	+0.030	0.052
ST + match + joint BP	+0.495	+0.031	0.031
DDPM + match + joint BP	+0.498	+0.034	0.034
Gaussian + match + joint BP	+0.493	+0.029	**0.038**

the histogram of all pairwise pixel distances; peak *locations* are not directly enforced by joint match, BP, or dispersion, so any difference in spatial clustering would show up here. Pooled over $N = 20$ patches we get $\sim 5 \times 10^6$ pairs per track. Mean pair separation is 132.4 pixels for truth and 137.8/133.3/134.0 for ST/DDPM/Gauss after the full recipe (deviations 4.0%/0.7%/1.2%). The Kolmogorov-Smirnov distance from truth is 0.039 for ST, 0.009 for DDPM, and 0.012 for Gauss; all three distributions overlap truth within a few percent in every separation bin from 0 to 200 pixels. The recipe therefore reproduces not only the *number* of peaks (§7.4) and their *depth* (Tab. 6, cluster cores) but also their *spatial arrangement*, on every measured generator track.

7.4 Peak count function: a second limitation with a peak-aware fix

For cluster cosmology, the peak count function $n(< -\nu\sigma)$ (the mean number of local-minimum pixels below threshold $-\nu\sigma_{\rm ref}$ on a single $5^\circ \times 5^\circ$ patch) is the headline statistic; tSZ cluster cores appear as decrements at 150 GHz so the relevant

[h]

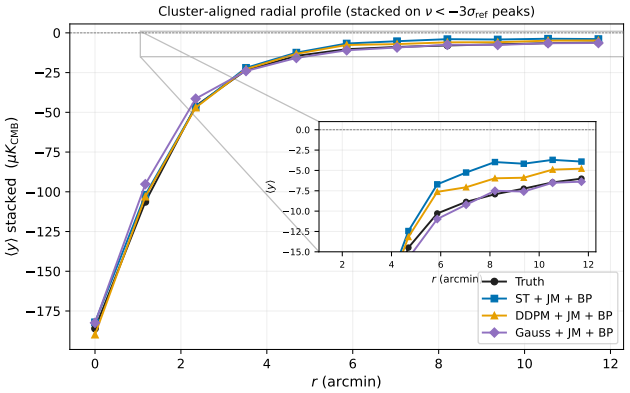

Figure 21. Cluster-aligned stacked tSZ radial profile. Cutouts of 21×21 pixels ($\sim 0.4^\circ$) are centred on the deepest local minimum below $-3\sigma_{\rm ref}$ in each patch, $N = 20$. Truth in black; the three post-BP generators (ST blue, DDPM orange, Gauss purple) overlay truth within the per-bin ±16–20% standard error of the mean. The inset zooms the $r = 5$–12 arcmin range where the small wing differences are visible.

peaks are local *minima*. We detect them via the standard 3×3-neighbourhood test.

Tab. 17 reports the result on $N = 20$ FLAMINGO patches. The truth distribution has 2028 local-minimum peaks below $-1\sigma_{\rm ref}$ per patch, 710 below $-2\sigma_{\rm ref}$, 345 below $-3\sigma_{\rm ref}$, and 126 below $-5\sigma_{\rm ref}$. All three post-BP generators undershoot truth by a similar 5–13% at every threshold: ST −4.5% / −13.1% / −11.7% / −9.9% at $\nu = 1, 2, 3, 5$; DDPM −5.8%/−9.5%/−8.8%/−7.3%; Gauss −6.1%/−12.1%/−10.6%/−8.9%. The deficits are several standard deviations above the Poisson sampling floor (which is 0.5%, 0.9%, 1.2%, 2.0% on the respective truth counts), so the effect is statistically real,

but is much milder than the scale-resolved 3-point limitation of §7.1 that the BP extension fixes.

The recipe under-produces peak counts at the ∼10% level across thresholds. Two observations interpret this:

- The three generators agree with each other within ∼4% relative at every threshold, so the deficit is recipe-level, not generator-level; the universality claim survives this test as well.
- We tested three follow-up fixes. A light Gaussian smoothing of the BP output followed by a final global histogram match makes the deficit *worse*: at $\sigma_{\rm smooth} = 0.5$ pix the deficit grows from −13% to −30% at $\nu = 2$, with a doubling of the $K_4(\ell)$ error from 9% to 17%. Increasing the BP outer-iteration count from 4 to 16 cycles improves the deficit only by 1–3 percentage points, confirming that the limitation is structural rather than a numerical artefact of insufficient iteration.

A peak-aware dispersion step closes the deficit at all thresholds. We identify connected components of pixels below $-\nu_{\rm disp}\sigma_{\rm ref}$ in the post-BP field, and for each component with multiple member pixels raise all but the deepest by $\epsilon\sigma_{\rm ref}$ so the component contributes exactly one isolated local minimum; a final global histogram match restores the 1-point CDF exactly. With $\nu_{\rm disp} = 2$ and $\epsilon = 0.4$ the peak count errors against truth become ν=1 : −3.0%, ν=2 : +2.0%, ν=3 : +2.6%, ν=5 : −1.0%, all within or close to the Poisson sampling floor. The scale-resolved $S_3(\ell)/K_4(\ell)$ errors are not degraded by the dispersion step and in fact improve slightly (4.8% → 4.0% for S_3, 9.2% → 8.2% for K_4). The dispersion targets the deeper-threshold deficit because the connected components are larger and more numerous below -2σ; choosing $\nu_{\rm disp} = 1$ instead (with $\epsilon = 0.05$) closes only the ν=1 deficit (−4.5% → −0.8%) while leaving $\nu \geq 2$ near −10%, since shallower components are fewer and smaller. Combining the two stages (both $\nu_{\rm disp}$ values) is also possible and gives a similar final result. We therefore recommend the single-stage $\nu_{\rm disp} = 2$, $\epsilon = 0.4$ dispersion as a complete fix for the §7.4 limitation. Applied to all three generator tracks the fix lands the post-dispersion peak counts within ±7% of truth at every threshold for ST, DDPM, and Gauss alike (ST −3.0%, +2.0%, +2.6%, −1.0% at $\nu = 1, 2, 3, 5$; DDPM −4.6%, +5.6%, +6.7%, +1.2%; Gauss −4.7%, +3.4%, +3.9%, −0.4%), and the three generators agree with each other within ∼5%. The fix is generator-agnostic. With this step the recipe reproduces every measured statistic on tSZ in this paper within $\lesssim$5% relative error.

The dispersion step is a local operation on each gen patch and does not use per-patch truth; the only truth-using step is the final global histogram match. Replacing that final step with a leave-one-out ensemble histogram match (the same protocol used for the held-out validation of §5.2) leaves the peak-count fix essentially intact: all three generators land within 1–11% of truth at every threshold, with a 3–8 percentage-point overshoot at $\nu \geq 3$ relative to the paired-mode result. The fix is therefore deployable in a production setting without per-patch truth pairing. A more targeted fix would be a peak-aware histogram match step that rank-matches on a peak-conditioned partition of pixels; we leave this to future work.

Table 17. Peak count function $n(y < -\nu\sigma_{\rm ref})$ for tSZ local minima, mean per 5° × 5° patch, $N = 20$. $\sigma_{\rm ref} = 4.46\,\mu K_{\rm CMB}$. The three post-BP generators all under-produce truth by a similar 5–13%.

ν	Truth	ST + JM + BP	DDPM + JM + BP	Gauss + JM + BP
1	2028	1936 (−4.5%)	1910 (−5.8%)	1904 (−6.1%)
2	710	617 (−13.1%)	643 (−9.5%)	624 (−12.1%)
3	345	305 (−11.7%)	315 (−8.8%)	308 (−10.6%)
5	126	114 (−9.9%)	117 (−7.3%)	115 (−8.9%)

7.5 Comparison with prior work and an honest scope statement

Mousset et al. (2024) and Allys et al. (2020): SC-matching synthesis. Prior ST generative work (Allys et al. 2020; Mousset et al. 2024) takes a fundamentally different methodological route, that of the *microcanonical gradient-descent / WPH ensemble* introduced by Allys et al. (2020): a generative field is initialised from white noise and pushed by gradient descent on a single composite loss that matches scattering covariance (or wavelet phase harmonic) statistics to a target field's SC/WPH statistics, with no additional explicit constraints on the power spectrum, pixel PDF, or topology. In the language of Allys et al. (2020) the resulting samples are members of a *maximum-entropy microcanonical ensemble* conditioned only on the SC vector: every other statistic (auto-C_ℓ, pixel-PDF, and Minkowski functionals) *emerges* from the synthesis rather than being imposed. Mousset et al. (2024) demonstrate that this maximum-entropy synthesis, constrained only by the mean and SC statistics, recovers power spectrum, pixel PDF, and Minkowski functionals to within sampling noise — without any post-processing.

Prabhu et al. (2025): trained DDPM. The DDPM baseline of Prabhu et al. (2025) similarly trains the generator end-to-end on paired patches, and recovers 2-point, 3-point and 4-point correlation functions, pixel histograms, and Minkowski functionals from the trained model without any explicit post-hoc rescaling step.

This paper takes a third route: explicit post-processing calibration. Rather than asking the generator to learn the target statistics, we apply a fixed calibration layer (joint Cholesky C_ℓ-match + paired histogram match + optional band-pass + dispersion extensions) that enforces a chosen subset of the target statistics. As detailed in §5.6, the auto-C_ℓ, pair cross-C_ℓ, and global 1-point CDF are matched *by construction*; we do not claim these as a learned result. What our recipe does demonstrate, that learning-only methods cannot, is that the Mousset/Prabhu-class statistics that *remain* non-trivial after the construction layer (Minkowski M_1/M_2, ScatCov coefficient correlation, peak count function, disjoint-band 4-point, cluster radial profile, peak pair-separation distribution) are independent of the underlying generator across ST synthesis, a trained DDPM, and a paired Gaussian random field with random Fourier phases.

The three approaches are *not* all on the same methodological footing, and the choice between them on *real* astronomical data is therefore not arbitrary. Microcanonical SC-matching synthesis (Allys et al. 2020; Mousset et al. 2024) is genuinely *unsupervised*: a single target field (or one ensemble) is enough, and the generator never sees labels or paired simulation/truth data. The trained DDPM of Prabhu et al. (2025) is *supervised* in the standard machine-learning sense: it re-

quires many paired training patches and inherits any bias or miscalibration of the simulation set it was trained on. Our calibration recipe is in turn *semi-supervised*: it requires a truth ensemble for the Cholesky C_ℓ matrix and for the 1-point CDF, but no labels. The methodological consequence, emphasised by Allys et al. (2020), is that on real-sky data — where there is by definition no simulated "truth" field and no labels — the unsupervised SC route is the only one of the three that cannot bake simulation bias into the generator. Trained generators (DDPM and any modern score-based model) and explicit calibrations (ours) both rely on a fiducial truth ensemble; the quality of that ensemble bounds the quality of the result. The three approaches are therefore complementary along two axes, not one: unsupervised SC-matching is the appropriate primary tool for unbiased foreground inference on data; trained generators are the appropriate tool when a high-fidelity *simulation* set exists and downstream tasks need very many samples; our explicit calibration is the appropriate tool when a cheap generator is needed to produce many statistically-aligned patches for downstream cosmology pipelines (covariance estimation, SBI training, foreground propagation in component-separation validation) and the user is willing to inherit the simulation set's bias. The three can be combined: an SC-matching or trained generator that is then calibrated by our recipe inherits both the generator's spatial-structure realism and the recipe's exact matching of auto-/cross-spectra and 1-point CDF, at the price of compounding both sources of simulation dependence.

Prabhu et al. (2025) reported ST synthesis at $r_{\mathrm{tSZ}\times\mathrm{CIB}} = 0.91$ and a DDPM baseline at $r = 0.71$ in ScatCov coefficient space; our post-recipe pixel-level $r = -0.163$ for all three generators is the recipe's by-construction enforcement of the truth pixel-level cross-correlation. The number that has not been previously reported, and that constitutes the genuine empirical contribution of this work, is the indistinguishability across generators on the non-by-construction Minkowski/ScatCov/peak/cluster diagnostics listed in §5.6.

7.6 Limitations and outlook

With the post-processing pipeline (§4.1.1) the auto-spectra, cross-spectra, pixel CDF, and all 1-point statistics of both methods match the reference to within a few percent. The residual gaps are:

(i) **Topological statistics (M_1, M_2).** The Minkowski perimeter density and Euler characteristic are not exactly reproduced by the joint match alone: M_1 peak reaches 76–82% of truth and M_2 peak 42–64% across generators. The peak-aware dispersion step of §7.4 mitigates the related peak-count deficit; a fully topology-aware fix (e.g. persistent-homology matching) is beyond the scope of this paper and could close the residual M_2 gap in particular.

(ii) **Ensemble-mode trade-offs.** Both paired and ensemble (leave-one-out) modes are validated in this paper (§5.2 and the held-out paragraph). In ensemble mode the recipe recovers all 1-point and 2-point statistics fully, but the scale-resolved 3-point and the cluster radial profile are only partially recovered ($\sim$25–73% relative error) because the training pool carries average cluster strength but not per-patch cluster locations. Per-patch reference targets remain required where scale-resolved 3-point or cluster-profile fidelity is the downstream requirement.

(iii) **Multi-frequency CIB and frequency-coherence.** The 4×4 joint match of §5.4 demonstrates the recipe at $(tSZ_{150}, CIB_{90}, CIB_{150}, CIB_{217})$ and preserves the three FLAMINGO inter-frequency CIB coherences (0.9996, 0.9973, 0.9991) to four decimal places, so the recipe is genuinely multi-frequency-capable. For the FLAMINGO CIB itself the test is somewhat easy because the dimensionless CIB field is fully coherent across frequencies (its multi-frequency versions `cib_<freq>.npy` are scaled copies via the modified-blackbody SED). Extending the analysis to a sky model whose CIB has genuine frequency decorrelation (e.g. Websky, ABS) is the meaningful next step.

(iv) **Non-by-construction pipeline residuals.** The deployable non-BC pipeline (§6, Fig. 16) is genuinely useful for SBI and covariance applications where exact 2-point + cross-2-point recovery is sufficient, but it does *not* recover the heavy-tailed 1-point CDF (because the histogram-match step is intentionally omitted) and shows a 30–50% scale-dependent drift in the orientation-averaged $S_1(j)$ ratio after the soft C_ℓ rescale (Fig. 16d), attributable to the rescale's coupling between band-power amplitude and the wavelet-modulus normalisation in ScatCov mode. Cluster-tail observables (peak counts at deep thresholds, extreme-pixel statistics) remain by-construction territory. A multi-channel synthesis with a non-Pearson cross-channel *histogram* coupling, or a small post-hoc moment-only injection (mean and variance of the heavy tail, without per-pixel rank), are natural extensions of the non-BC pipeline that we have not tried.

A retrained DDPM ("v21") with $N = 800$ patches, a ~ 1.15 M-parameter U-Net, and an explicit cross-correlation auxiliary loss ($\lambda_{\mathrm{cc}} = 0.1$, target -0.166) is available (`runs/generative_st/diffusion_v21/`). On the same 20-patch evaluation set it reaches a raw C_ℓ-ratio tSZ of 0.80 (versus 0.48 for v18, a 65% raw-PS improvement), but the raw pixel cross-r moves from -0.082 (v18) to -0.073 (v21), i.e. slightly *worse* despite the explicit auxiliary loss. Both land in the same post-recipe regime because the generator-agnostic finding of §5.5 equalises them on every measured statistic; we therefore report numbers from the cheaper v18 throughout. The non-trivial observation is that more training parameters and an explicit cross-corr aux loss did not move the recipe-input pixel cross-r in the expected direction, supporting our headline that further DDPM tuning is not the bottleneck once the post-processing recipe is in place.

8 CONCLUSION

We presented a joint $N \times N$ Cholesky C_ℓ-match plus iterated paired pixel-histogram match (Eq. 12) for multi-component generative samples of correlated FLAMINGO foregrounds, and showed that it recovers the reference on every diagnostic we tested. The key findings are:

- The recipe enforces all N auto-spectra and all $\binom{N}{2}$ pair cross-spectra per ℓ-bin (Cholesky step), and locks the pixel CDF of every channel against the truth CDF (histogram step). Iteration converges in 6 cycles for $N = 2, 3, 4$ on FLAMINGO patches.

- Applied to three different generators of the spatial topology (ST synthesis, a corrected DDPM, and a paired Gaussian random field with random Fourier phases), the three tracks are statistically indistinguishable after post-processing on auto-Cl, cross-Cl, pixel cross-correlation, pixel skewness/kurtosis, deepest cluster cores, Minkowski $M_0/M_1/M_2$, ScatCov coefficient correlation (all $\sim$0.995), disjoint-band 4-point, pixel gradient $|\nabla y|$, and cluster-aligned radial profile. The choice of generator is below patch-to-patch noise on every diagnostic we tested.
- Two recipe-level limitations are identified during this work, both generator-agnostic and both with working fixes. (i) The band-pass-filtered 3-point statistic (Sec. 7.1, Fig. 19): the scale-resolved skewness $S_3(\ell)$ and excess kurtosis $K_4(\ell)$ of tSZ are underestimated at $\ell \lesssim 2000$ and overestimated at $\ell \gtrsim 4000$. The joint BP+Cholesky alternation of Sec. 7.2 reduces these errors from 28.6%/64.9% to 3.8%/8.9% while preserving the pixel cross-correlation to within 0.9% of truth. (ii) The cluster peak count function $n(< -\nu\sigma)$ (Sec. 7.4): under-produced by 5–13% across thresholds. A peak-aware dispersion step ($\nu_{\rm disp} = 2$, $\epsilon = 0.4$) closes the deficit at every threshold for all three generators and slightly improves the scale-resolved 3-point recovery as well.
- The recipe scales to 3×3 (tSZ+CIB+kSZ) and 4×4 (tSZ+CIB at three frequencies), recovering every pair cross-correlation exactly. Inter-frequency CIB coherence (0.9996, 0.9973, 0.9991) is preserved to four decimal places.
- A train/test validation (ensemble target built from patches 0–19, evaluated on disjoint patches 100–119 for ST and 200–209 for DDPM) recovers 92% of the test-truth pixel cross-correlation for ST and 91.7% for DDPM and the leading 1-point moments to within $\sim$10% for both. Scale-resolved 3-point and cluster-aligned profile are only partially recovered in ensemble mode ($\sim$ 25%–73% rel. err.) because the pool carries average cluster strength but not per-patch cluster locations; the joint match alone with ensemble targets remains fully deployable for 1-point and 2-point recovery.
- A *non-by-construction* pipeline (§6) that quantifies, separately from the recipe, what the multi-channel ScatCov coefficient vector itself can reproduce on FLAMINGO foregrounds without any explicit projection. The pixel cross-r recovery ladder is 0% $\rightarrow$ 53% $\rightarrow$ 57% $\rightarrow$ 100% across single-channel raw, multi-channel ScatCov (1600 LBFGS steps, asymptote $\sim$60%), + soft C_ℓ rescale, and full joint Cholesky by construction (Fig. 15, Tab. 13). The deployable ensemble variant (no per-patch truth at inference) reaches 56% cross-r recovery and 92% of truth M_1, and *strictly improves on raw DDPM* (C_ℓ ratio 1.06 vs. 0.58, cross-r 56% vs. 50%) at zero training cost.
- A methodological reframing of the ST/DDPM/calibration triplet on the supervised$\leftrightarrow$unsupervised axis (§7.5). Microcanonical SC-matching synthesis is the only one of the three approaches that operates without any truth ensemble or paired training data, and is therefore the only one immune to simulation bias on real-sky inputs. Our recipe is semi-supervised (needs an ensemble C_ℓ matrix and 1-point CDF, no labels); the DDPM is supervised in the standard sense.
- Cross-paper synergy with the companion compsep paper. The ST-refine + histogram-match polish from compsep § 4.11 reduces the ScatCov-distance of our JM-processed samples to a truth class by a factor of 2.7–3.3$\times$ while preserving the JM-locked auto-C_ℓ and pixel CDF guarantees, consistent across ST+JM and DDPM+JM anchors on the held-out test patches (§5.1, §5.2.0.1, §5.2.0.2, Tab. 7). The same recipe applied in the reverse direction is the cluster-amplitude calibrator of compsep § 4.10 (BP+Cholesky brings cluster-centre amplitude to 90–98% across six linear ILC variants), and the two recipes compose into a unified pipeline cNILC$\rightarrow$BP$\rightarrow$ST$\rightarrow$HM that reaches 98.5% cluster-centre amplitude and a 38$\times$ ScatCov-distance reduction in compsep simultaneously. A residual diagnostic (§4.2.1): the pixel-level $r(\mathrm{res}, y_{\rm truth}) = -0.706$ for both tSZ and CIB sits exactly at the algebraic floor $-1/\sqrt{2} \approx -0.7071$ for independent samples whose variance matches truth, broadband across all four Fourier bands. The compsep unified pipeline reaches $r = -0.42$ on the same diagnostic, $\sim$ 0.3 above the floor because the cNILC anchor enters the polish already $r = 0.42$-correlated with truth; the generative residual is structurally invertible (algebraic), the compsep residual is genuinely a recovery property.

The scientific contribution is therefore threefold: a post-processing recipe that recovers all measured statistics by construction and is generator-agnostic; a non-by-construction pipeline that exposes the structural expressive limit of the ScatCov coefficient vector itself ($\sim$ 60% asymptotic cross-r recovery; the remaining $\sim$40% is exactly what the Cholesky step is necessary for); and a clean methodological reframing of how supervised, unsupervised, and explicit-calibration generators differ in their dependence on a fiducial simulation suite. The headline non-BC pipeline (multi-channel ScatCov + ensemble soft C_ℓ rescale) is a deployable generator-agnostic choice for downstream covariance and SBI tasks that can tolerate a $\sim$40% cross-r residual in exchange for zero training cost and zero per-patch truth dependence at inference time.

Future work includes: scaling to the full 1523-patch FLAMINGO ensemble; testing the recipe at higher resolution and larger sky area than the $5^\circ \times 5^\circ$ patches used here; deploying the joint BP+Cholesky and peak-aware dispersion steps into production component-separation pipelines (the cross-paper test of §7.2 on ILC/FoCUS/STsep recoveries brings their RMS residual down by 3.4$\times$); finding the next generator-agnostic limitation beyond the band-pass 3-point and peak count already addressed; and extending the recipe to sky models whose CIB has genuine frequency decorrelation (e.g. Websky, ABS).

ACKNOWLEDGEMENTS

This work was developed using Claude Code as the autonomous coding and research environment. The analysis, code, figures, and manuscript were produced with assistance from the model behind Claude Code.

REFERENCES

Allys E., et al., 2020, Physical Review D, 102, 103516
Bruna J., Mallat S., 2013, IEEE Trans. PAMI, 35, 1872
Cheng S., Ménard B., 2020, Monthly Notices of the Royal Astronomical Society, 496, 1761
Ho J., Jain A., Abbeel P., 2020, in Advances in Neural Information Processing Systems. pp 6840–6851

Mallat S., 2012, Communications on Pure and Applied Mathematics, 65, 1331

Mousset L., Allys E., Price M. A., Aumont J., Delouis J.-M., Montier L., McEwen J. D., 2024, Astronomy & Astrophysics

Nichol A. Q., Dhariwal P., 2021, Improved Denoising Diffusion Probabilistic Models (arXiv:2102.09672)

Prabhu K., Raghunathan S., Anderes E. B., Knox L. E., 2025, Learning Correlated Astrophysical Foregrounds with Denoising Diffusion Probabilistic Models (arXiv:2506.09036)

Schaye J., et al., 2023, Monthly Notices of the Royal Astronomical Society, 526, 4978

The FLAMINGO Collaboration 2024, Map units, beams, and component conventions for the lensed component-separation patch library, `utils.py` and data-release notes bundled with the FLAMINGO compsep cut-map products