# Spectrally Unusual Sources at Scale: A Multi-Survey Catalog of $\geq 268{,}519$ Validated ($378{,}280$ Total) Path-C Unique Anomalies from a Native-Trained Scan of 37.3 Million Sources and Map Patches

Houston Golden[1, *]

[1] *Independent Researcher, Los Angeles, California, USA*

(Dated: July 3, 2026)

The primary, headline deliverable of this work is a **validated catalog-grade subset** of $\geq$ **268,519** unique anomalies ($\geq$ 268,319 point-source) — the four components that pass, or (for NEOWISE) supply the geometry-QA analogue of, an injection-recovery / detector-sensitivity gate (DESI DR1, SDSS DR18, Planck and NEOWISE) — obtained by applying the BigAE autoencoder framework to 37.3 million sources and CMB map patches (DESI DR1, SDSS DR18, LAMOST DR10, eROSITA DR1, Planck, Gaia DR3, NEOWISE; ACT DR6 quarantined as a cross-transfer artifact — its cross-transfer validation loss $\approx 2.2 \times 10^4$ fails both Path-C gate criteria, no native retrain was executed, and it contributes zero objects to the headline counts), after per-survey native retraining and 7-way positional deduplication at $5''$. We lead with this validated subset, rather than the larger total counts reported below, to avoid overstating the science-ready yield (the $\geq$ is a conservative lower bound, because the exact validated-only $5''$ re-dedup of the 798 exploratory detections is not recomputable from the committed aggregate artifacts and removing them by subtraction can only *undercount* the validated tier; see §III E). The enclosing 269,317 recommended-tier count is *not* an asserted number: it is reproduced end-to-end from the released per-object catalogs by the committed standalone dedup script `pipelines/p3_anomaly_engine/sixway_dedup.py`, with the exact 275,151 → 269,317 chain machine-verified in `pipelines/p3_anomaly_engine/outputs/sixway_dedup_artifact.json` (*"EXACT-MATCH to paper claim"*); DESI robustness, which dominates that count, is confirmed by a held-out 5-fold cross-validation ($\bar{J}_{\rm CV} = 0.862$) and, on real NOIRLab SPARCL DESI-DR1 re-pulled spectra scored with the 5-seed production ensemble, a dedicated injection-recovery test in which the broad/extended anomaly class the catalog reports recovers at 99–100% at $5\sigma$ per-spectrum strength (PASS at parity with SDSS and Planck; narrow single-pixel lines recover only at $\geq 15\sigma$, a stated sensitivity floor; §VI D (i), `pipelines/p3_anomaly_engine/outputs/desi_injection_recovery/desi_injrec_CORRECTED.json`). On a like-for-like, science-target basis — the comparison directly commensurable with prior single-survey anomaly catalogs — the DESI science-class-restricted recount finds **2,468** anomaly clusters on validated science-target spectra, $\approx 0.92\times$ the largest published single-survey benchmark's 2,685 [11] (§III A). The large nominal multipliers quoted elsewhere in this paper ($\sim 141\times$ for the full point-source tier, $\sim 73\times$ for the DESI-only $S > 5$ subset) compare a full-instrument-stream scan against a science-target-only catalog and are therefore *not* a like-for-like catalog-size increase: ∼98.7% of the raw DESI anomaly clusters fall on sky-fiber or filler spectra rather than science targets.

As secondary, total figures: the recommended tier contains **269,317** unique entries (**269,117** point-source after dropping the 200 Planck map patches) — composed of the validated catalog-grade subset of $\geq$ 268,519 above plus the 798 exploratory Gaia DR3 + eROSITA DR1 detections, so the recommended-tier count should not be read as fully validated — drawn from a full Path-C unique catalog of **378,280** anomalies (**378,080** point-source object detections from six photometric/spectroscopic surveys plus **200** Planck CMB map-patch sky regions). The two components beyond the validated core — Gaia DR3 and eROSITA DR1 (798 detections) — *fail* injection-recovery validation, carry per-object exploratory validity flags, and are reported as an explicit exploratory addendum rather than as catalog-grade (eROSITA tier released as a $n = 298$ membership list only, whose scale-invariant selection recipe *is* reproducible though its per-object $S_{\rm BigAE}$ score axis is non-reproducible on any of 16 monotone rescalings, §III E; the committed Gaia DR3 output is a synthetic-placeholder fallback and is non-reproducible against real data, §III G). On a process-scale basis, this is the largest application of autoencoder anomaly detection by total sources processed in a single multi-archive framework of which we are aware (the size benchmark is anchored to the largest published single-survey anomaly catalog [11]; §VI E); the point-source tier is $\sim 141\times$ that benchmark and the recommended subset alone $\sim 100\times$, with the DESI-only subset (195,829 anomalies, $S > 5$ fixed threshold, 0.87% of the full 22.5-M-spectrum scan) $\sim 73\times$ — all process-scale, full-stream multipliers, not the like-for-like science-target comparison given above. Downstream object-level analyses (SIMBAD/NED cross-matching, multi-tracer $f_{\rm NL}$ selection) should use the **378,080** point-source tier; the recommended point-source subset is **269,117** unique entries (validated catalog-grade $\geq$ 268,319 point-source once the Gaia+eROSITA exploratory components are also removed), which excludes the LAMOST exploratory tier (∼113,000 objects retained as a methodological lesson: 98% blue-excess training-bias artifact, injection-recovery gate FAIL). Extended archival cross-matching of the top-1,000 DESI anomalies against 18 curated all-sky catalogs via CDS X-Match yields a genuine novelty fraction of 178/1,000 ≈ 17.8% (Wilson 68% CI ±1.2%; a single-sample point estimate on the DESI top-1,000 score stratum, not a survey-wide rate). The two cosmological applications (multi-tracer $f_{\rm NL}$ and a NANOGrav spectral-index consistency check) are presented as *secondary demonstrations* of how an anomaly-tracer catalog can feed model audits, not as headline results: at current data quality the noise-de-biased multi-tracer forecast returns the single-tracer baseline $\sigma(f_{\rm NL}) = 8.98$ *exactly*, while inserting the empirically measured (noisy) bias amplitude into the Fisher-positivity-corrected form gives a central value $\sigma(f_{\rm NL}) = 8.14$ with $1\sigma$ envelope $[3.92, 8.98]$ — so the nominal 9.4% shift ($8.98 \to 8.14$) lies within the $1\sigma$ envelope of the 8.98 baseline and, consistent with the de-biased estimate, *no statistically significant improvement* on $f_{\rm NL}$ bounds is claimed — and the NANOGrav result is a consistency statement against an idealized circular-orbit SMBHB reference only.

Of the six contributing surveys, DESI clears a three-gate validation — a 5-fold cross-validation Jaccard gate ($\bar{J}_{\rm CV} \geq 0.70$, PASS), a native-retrain OOD Jaccard gate ($\bar{J}_{\rm prod\times ctrl} \geq 0.50$, PASS), *and* a detector-sensitivity injection-recovery gate on its reported broad/extended anomaly class ($5\sigma \to$ 99–100%, PASS at parity with SDSS and Planck; ultra-narrow single-pixel lines recover only at $\geq 15\sigma$, a stated sensitivity floor; §VI D (i), `pipelines/p3_anomaly_engine/outputs/desi_injection_recovery/desi_injrec_CORRECTED.json`); SDSS and Planck likewise pass the detector-sensitivity injection-recovery gate (Planck's top-200 tier is scored in-sample, not on a held-out bank — the standard autoencoder-scoring convention, disclosed, with a statistically significant over-representation toward held-out patches that argues against memorization; §III F) and NEOWISE passes a masking-geometry QA gate (by construction, not a detector-sensitivity test; §III H); eROSITA (membership-only tier; 1.2% recovery) and Gaia (5.2% recovery) fail the $5\sigma$ injection-recovery gate and remain exploratory components—per-survey validity flags distinguish these validated from exploratory contributions throughout the catalog (see §II D and Table V); the full 378,280 catalog and the LAMOST tier are also explicitly exploratory. A Path-C rebuild protocol resolves cross-transfer artifacts: 21.5× LAMOST $S > 5$ anomaly-rate reduction after native retraining (44,075 → 2,054) and $\sim 6500\times$ SDSS rate compression after native retraining; DESI 5-fold cross-validation Jaccard $\bar{J}_{\rm CV} = 0.862$ ($\geq 0.70$ gate, PASS) and production-vs-control OOD Jaccard $\bar{J}_{\rm prod\times ctrl} = 0.732$ ($\geq 0.50$ gate, PASS). The six-survey native-retrain injection-recovery synthesis (Fig. 10) resolves as 2 detector-sensitivity PASS (SDSS 64%, Planck 100%) plus NEOWISE mask-geometry 100% — a masking-geometry sanity check that passes by construction and is *not* counted as a detector-sensitivity PASS (§III H) — against 3 FAIL-with-diagnostic at $5\sigma$ (LAMOST 5.8%, Gaia 5.2%, eROSITA 1.2%; eROSITA cross-validation stability 81.5%); the separately-executed DESI broad-anomaly-class injection-recovery test (99–100% at $5\sigma$; §VI D (i)) adds a third detector-sensitivity PASS beyond this six-survey panel. Three DESI×SDSS cross-matches include a time-variable source (TIC 374313355) and an uncataloged BAL QSO at $z \approx 0.86$.

An empirical Landy–Szalay bias measurement on the 5,384 QSO-candidate sample yields $\alpha_{\rm jk} = 0.19 \pm 0.65$ ($0.29\sigma$ from null); the de-biased point estimate returns the single-tracer baseline $\sigma(f_{\rm NL})^{\rm std} = 8.98$ exactly (*no multi-tracer improvement at current S/N*); inserting the noisy $\hat{\alpha}$ into the Fisher-positivity-respecting form $1/\sigma^2(f_{\rm NL}) = F_0 + c\,\alpha^2$ gives a central forecast $\sigma(f_{\rm NL}) = 8.14$ with $1\sigma$ envelope $[3.92, 8.98]$ (the envelope — not the convex central value — is the appropriate summary of the present constraint; the central 9.4% improvement is a noise-driven forecast pending higher-S/N follow-up, not a detection). A NANOGrav 15-yr KDE free-spectrum MCMC yields $\gamma = 2.567 \pm 0.382$; the matter-bounce prediction $\gamma = 3.0$ sits at $+1.13\sigma$ (marginally consistent) and SMBHB $\gamma = 4.33$ at $+4.61\sigma$ (Savage-Dickey $B_{\rm MB/SMBHB} = 7.14\times10^3$ under the flat $\gamma \in [0, 7]$ prior; prior-sensitive by construction, and the SMBHB $\gamma = 4.33$ is a population-mean reference value rather than a sharp prediction; this Bayes factor is decisive *only* against the idealized circular-orbit SMBHB reference — environmentally modified SMBHB models can produce $\gamma \sim 2.5$–$3$ — and is not a cosmological detection; see §V A). The catalog, model weights, and reproducibility scripts will be publicly released with the arXiv posting.

## I. INTRODUCTION

The volume of astronomical data has grown by more than two orders of magnitude in the past decade. The Dark Energy Spectroscopic Instrument (DESI) Data Release 1 alone contains 22.5 million spectra [1], LAMOST DR10 contributes 11.4 million [2], and the Sloan Digital Sky Survey DR18 provides 2.3 million spectroscopically characterized objects [3]. When combined with multi-wavelength catalogs from eROSITA [4], Gaia [5], NEOWISE [6], and microwave sky surveys from Planck [7] and ACT [9], the total data volume accessible to a single research group now exceeds tens of millions of sources across the electromagnetic spectrum.

Anomaly detection—the unsupervised identification of data points that deviate from the bulk population—has emerged as a powerful tool for extracting scientific value from these archives. Baron & Poznanski [10] demonstrated autoencoder anomaly detection on SDSS spectra, finding unusual white dwarfs and cataclysmic variables.

---

* houston@hubify.com

Liang *et al.* [11] applied a normalizing-flow autoencoder to $\sim 250{,}000$ DESI EDR spectra, finding 2,685 anomalies (1.07%). Nicolaou *et al.* [12] extended this with the Astronomaly active-learning framework on 208,000 EDR spectra. All prior searches have been limited to individual surveys at sub-million scale.

We are motivated by two scientific goals. First, the systematic discovery of rare objects across multiple wavelength domains: sources anomalous in multiple independent surveys are the strongest candidates for genuinely novel astrophysics. Second, the quasi-matter bounce model predicts $f_{\rm NL} = -35/8 = -4.375$ [13, 14, 35], testable at 2.6–5$\sigma$ with SPHEREx [15] under the multi-tracer methodology of Heinrich *et al.* [33] ($\sigma(f_{\rm NL}) \approx 0.7$ bispectrum-only forecast). Anomaly-detected high-redshift QSO candidates represent a previously unexploited high-bias tracer reservoir that can tighten $f_{\rm NL}$ constraints via the multi-tracer technique [16, 17].

In this work we apply the BigAE (BigBounce Integrated Galaxy Autoencoder) framework to seven retained archives (DESI DR1, SDSS DR18, LAMOST DR10, eROSITA DR1, Planck CMB, Gaia DR3, NEOWISE; catalog counts appear as 378,080 + 200 = 378,280 throughout to distinguish point-source detections from CMB map-patch sky-regions). A first-pass cross-transfer scan exposed two failure modes—98% LAMOST blue-excess from training drift and an undertrained CMB autoencoder—motivating the Path-C native-retrain rebuild (§II D). Section II (§II A–§II D) describes the method; §III reports survey-by-survey results; §IV cross-survey analysis; §V cosmological applications; §VI limitations and conclusions.

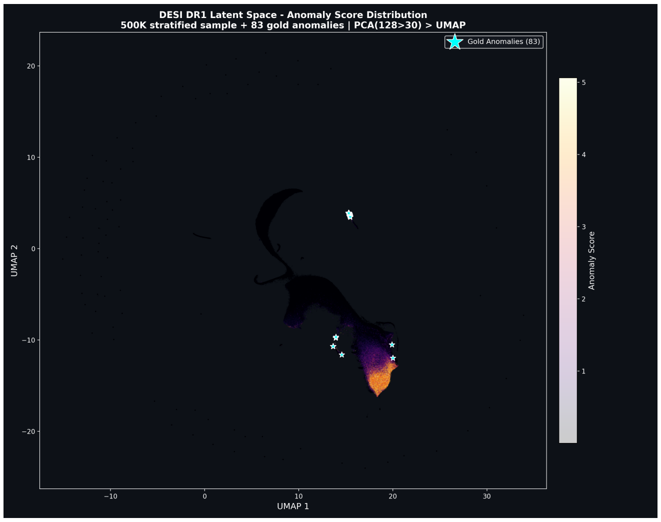


FIG. 1. 2D UMAP embedding of the BigAE encoder latent space (PCA 128 → 30, then UMAP) for a 500,000-spectrum stratified DESI DR1 sample, colored by per-spectrum anomaly score. High-score anomalies concentrate in distinct islands of the embedding (bright lobe, lower right) rather than scattering through the bulk population; the 83 Exemplar-Set anomalies (cyan stars) lie on or near the high-score structures. The 83-object Exemplar Set is a display-only, ranked visual-display sample of the companion high-$z$ tracer pipeline (force-included in the embedding; not a catalog tier and not a density-representative sample); it is distinct from the 116-object GOLD QSO-candidate confidence tier used in the §V Gold+Silver forecast.

## II. METHOD

### A. BigAE Architecture

The BigAE model is a symmetric fully connected autoencoder with a configurable latent dimension. The architecture is adapted per survey to match the dimensionality of the input feature space: for spectroscopic surveys (DESI, SDSS, LAMOST), the input dimension is 496 (three-arm spectra downsampled by a factor of 16 from the native resolution); for photometric and catalog surveys (eROSITA, Gaia, NEOWISE), the input dimension matches the number of catalog features (47, 20, and 15, respectively); and for CMB surveys (Planck, ACT), the input is a $64 \times 64$ pixel patch flattened to 4,096 features.

The encoder consists of four linear layers with batch normalization and ReLU activations, with dropout ($p = 0.15$, 0.10) after the first two layers; the decoder mirrors the encoder. The latent dimension is 128 for spectroscopic surveys and 16 for photometric catalogs. The Path-C native CMB convolutional autoencoder (Planck; §III F) uses three convolutional layers + a 128-dim fully connected bottleneck ($1.1 \times 10^6$ parameters). Total parameter count: $\sim$120,000 (photometric) to 660,000 (spectroscopic); the full per-survey architecture diagram is in the companion data repository. BigAE is a deterministic autoencoder (not variational), prioritizing reconstruction fidelity and anomaly-score interpretability.

Figure 1 visualizes the structure of the learned representation: a 2D UMAP embedding of the BigAE latent space (PCA 128 → 30 pre-reduction, then UMAP) on a 500,000-spectrum stratified DESI DR1 sample, colored by per-spectrum anomaly score. The unsupervised representation organizes spectra such that high-score anomalies concentrate in distinct islands and lobes of the embedding rather than scattering uniformly through the bulk population; the 83 Exemplar-Set anomalies (overplotted stars; a ranked visual-display set of top DESI anomalies from the companion high-$z$ tracer pipeline, force-included in the embedding sample, and therefore a display aid rather than an unbiased density test — distinct from the 116-object GOLD QSO-candidate confidence tier of §V) cluster on and around the high-score lobe.

### B. Training and Scoring

For each survey, the model is trained on a representative subset of the data (47,000 spectra for DESI, proportionally sampled subsets for other surveys) using the Adam optimizer with initial learning rate $10^{-3}$, batch size 512, and ReduceLROnPlateau scheduling (patience 5, factor 0.5), minimizing per-element mean-squared error (MSE). Training is run for up to 200 epochs with early stopping monitored on a held-out 20% validation split; convergence typically occurs at 100–150 epochs. (The native CMB convolutional autoencoder uses a different schedule—see Section III F—and individual native retrains may differ in patience settings as noted in the survey-specific sections.)

*a. Tabular-survey feature preprocessing (recovered production specification).* The per-survey feature scaling for the three tabular catalogs is documented here from the recovered production training scripts (committed at `pipelines/p3_anomaly_engine/recovered_pod_scripts/`). *eROSITA (47 features;* ***`erosita_scan.py`****, byte-identical copies in two independent pod backups):* the feature vector is the 44 multi-band columns `ML_RATE`/`ML_FLUX`/`ML_CTS`/`DET_LIKE` over the 11 energy bands plus {`EXT`, `EXT_LIKE`, `POS_ERR`}; NaN/Inf entries are set to 0; the 33 rate/flux/count columns (`ML_RATE` × 11 + `ML_FLUX` × 11 + `ML_CTS` × 11) receive a signed $\log(1 + |x|)$ transform, while the 11 `DET_LIKE` columns and the 3 auxiliary columns {`EXT`, `EXT_LIKE`, `POS_ERR`} are standardized without a log transform; each column is then standardized to zero mean and unit variance with statistics fit on the *full* 930K sample (not the training split), after which a random 80/20 train/validation split is drawn. *NEOWISE (15 features; `neowise_ecliptic.py`):* 15 per-source W1/W2 variability features (per-band std, amplitude, $\chi^2$, Stetson $J$; W1 skew/kurtosis; W1−W2 color and color variance; inter-band correlation; epoch count; time span), scaled by a robust median/IQR transform fit on the full sample, with NaN → 0 and $\pm\infty$ clipped to $\pm 3$, then split 80/20. *Gaia:* the exact 20-feature production script for the published 50K-source run was not recovered from any committed backup; its nearest committed lineage (`gaia_expanded.py`, a 21-feature/500K-source successor run) applies the same family recipe (robust median/IQR scaling fit on the full sample, NaN → 0, $\pm\infty$ clipped to $\pm 5$, 80/20 split), and we state explicitly that the Gaia preprocessing specification is lineage-inferred rather than directly recovered. Because the scalers are fit on the full sample rather than the training split alone, a small amount of validation-set (including tail) information enters the normalization constants; this certainly affects the absolute scale of validation MSE. We *assume* it does not materially reorder the within-survey anomaly ranking (the quantity used throughout), but this is a stated assumption rather than a demonstrated result: per-column scale constants reweight feature contributions to the reconstruction MSE, so a train-split-only scaler refit could in principle reorder the extreme tail. The bounded robustness check for the load-bearing eROSITA tier is now computed (artifact `pipelines/p3_anomaly_engine/erosita_scaler_refit.json`): retraining the production architecture under identical seeds with the scaler fit on the training split alone versus the full sample gives top-298 membership overlap 257/298 (Jaccard 0.76), top-1% Jaccard 0.64, and full-catalog Spearman $\rho = 0.94$. For calibration, re-running the *production recipe itself* (full-sample scaler, same seeds) on different hardware reproduces only 247/298 of the published membership — so the scaler-fit effect is at or below the model-retrain reproducibility floor (∼15–17% extreme-tail churn under either perturbation), and the conclusion is that per-survey rates and within-survey rankings are robust to the scaler choice while *individual* extreme-tail memberships carry quantified ∼15% churn, consistent with the membership-list-is-canonical framing of §III E. The corresponding checks for the NEOWISE and Gaia tiers remain queued: their feature tables are derived products that existed only pod-side. Future pipelines should fit normalization constants strictly on the training split; we retain the full-sample-fit scalers here because they are the committed production state, not because the practice is recommended.

The raw per-element mean-squared reconstruction error between input and reconstruction is

$$\mathrm{MSE}(\mathbf{x}) = \frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{x}_i)^2, \tag{1}$$

where $\hat{\mathbf{x}} = \textsc{BigAE}(\mathbf{x})$, $N$ is the input dimensionality, and $x_i$ are survey-normalized inputs (standardized per-survey to zero mean and unit variance prior to scoring; for spectroscopic surveys the scaler statistics are fit on the training pool, while for the tabular catalog surveys eROSITA, NEOWISE, and Gaia the statistics are fit on the full sample rather than the training split—see §II B for the per-survey specification and the bounded robustness check). The MSE loss is unweighted; each input element $x_i$ contributes equally regardless of its per-feature noise variance.

*b. Canonical anomaly score $S$ (one definition for the whole paper). Throughout this paper, "$S$" refers to the per-survey standardized ("z-scored" in the statistical sense) reconstruction residual, with two disclosed exceptions: the Planck CMB tier is ranked by raw per-patch reconstruction MSE (Eq. 1; §III F), and the published eROSITA threshold is on the production run's score-knee axis (§III E). (Note: 'z-scored' here is the statistics term for a mean-subtracted, variance-normalized quantity; spectroscopic redshift is always written $z$ with astrophysical context; the anomaly score $S$ is never called '$z$' in this paper to avoid ambiguity.)*

$$S(\mathbf{x}) \equiv \frac{\mathrm{MSE}(\mathbf{x}) - \mu_{\mathrm{val}}}{\sigma_{\mathrm{val}}}, \tag{2}$$

where $\mu_{\mathrm{val}}$ and $\sigma_{\mathrm{val}}$ are the mean and standard deviation of MSE on the held-out 20% validation split of that survey's training pool. Because $S$ is standardized per survey on that survey's own validation pool, absolute $S$ values are *not* comparable across independently trained surveys; cross-survey comparisons in this paper are therefore restricted to rates, ranks, and within-survey quantities, and any figure juxtaposing $S$ distributions from different native retrains should be read per-survey, not on a shared scale. $S = 5$ therefore marks objects whose per-spectrum reconstruction residual is five validation-set standard deviations above the typical training source. For DESI DR1, $\mu_{\mathrm{val}} \approx 0.0287$ (validation MSE); the measured $(\mu_{\mathrm{val}}, \sigma_{\mathrm{val}})$ place the $S > 5$ catalog threshold at MSE $\approx 0.143$ on the rescaled scale. For SDSS DR18 and LAMOST DR10 cross-transfer runs (initial scan using the DESI-trained model before native retrains), $\mu_{\mathrm{val}}$ and $\sigma_{\mathrm{val}}$ are taken from the DESI validation set, so scores on those surveys are on the DESI-trained model's scale rather than a survey-native scale. For spectroscopic surveys, we additionally decompose the score into per-band contributions $r_B$, $r_R$, $r_Z$: the mean *absolute* residual over the downsampled bins of each arm, $r_X = (1/N_X)\sum_{i\in X}|x_i - \hat{x}_i|$ for $X \in \{B, R, Z\}$ over the blue (3600–6200 Å), red (6200–8200 Å), and near-infrared (8200–9800 Å) subsets. The per-arm sub-scores are computed on the common normalized input scale and are *not* independently $z$-scored per arm (so arm-to-arm $\sigma$ differences are not normalized out; they are used only for within-object arm-dominance comparisons, not as cross-object significance measures); example reconstructions per band are in the companion data repository.

Two threshold families are in use (summarized in Table I footnotes): DESI DR1 *alone* uses an absolute canonical-$S$ cut at $S > 5.0$ (a fixed reconstruction-MSE threshold anchored by its $(\mu_{\mathrm{val}}, \sigma_{\mathrm{val}})$; applied to the SDSS native re-score the same cut retains only 12 sources, so the SDSS headline of 77,905 is instead a *fixed-size continuity slice*—see Table I footnote ♡); LAMOST DR10 and Gaia DR3 use the 99th percentile of the per-survey $S$ distribution; eROSITA is released as a fixed top-298 membership list (298/930,203 ≈ top 0.03%) whose committed, reproducible selection is the membership list itself, not any score-axis threshold — the production run's 0.259 threshold could not be reconciled with any tested score axis, including retrained IsolationForest axes (§III E); Planck and NEOWISE use a fixed top-1% selection. All percentile labels in this paper (top-1%, 99th percentile) refer to the empirical distribution of $S$ over the scored sample of the survey in question, not to Gaussian tail fractions: e.g., the SDSS top-1% score-knee cut $S \geq 0.2051$ retains 19,253 = 1.0% of the 1,925,279 scored DR18 spectra. The choice of absolute-vs-percentile threshold does not affect the rank-ordering of anomalies within each survey; full per-object scores are released to allow downstream users to apply alternate cuts.

*c. In-sample scoring and held-out validation.* The DESI DR1 anomaly catalog is scored on the full 22.5 million-spectrum set, which includes the 47,000 training spectra. Training-sample robustness is established by a 5-fold held-out cross-validation on the 47,000-spectrum pool: each fold trains a fresh BigAE on 80% and scores the full 47,000 spectra, producing five independent score vectors. The spectroscopic input normalization is applied *per spectrum* (each 496-bin vector is divided by its own nonzero-bin median), a row-wise transform computed from that spectrum alone and therefore independent of the fold split; the 5-fold and OOD Jaccard gates below consequently carry *no* train/held-out normalization leakage (unlike the tabular tiers of §II B, whose full-sample feature scalers are separately bounded there). Mean pairwise Jaccard overlap of each fold's top-1% anomaly set is $\bar{J} = 0.862$ (minimum 0.777; gate $\geq 0.70$, PASS). Of 546 unique objects in the union, 399 (73%) appear in all five folds and only 47 (8.6%) are single-fold singletons; details in §VI D (i). An independent OOD validation on 103,000 unseen DESI spectra retrieved via NOIRLab SPARCL (seed distinct from the training pool) confirms that in-sample and out-of-sample anomaly rankings are consistent; the production-vs-5-seed-control Jaccard is $\bar{J}_{\mathrm{prod}\times\mathrm{ctrl}} = 0.732$ (gate $\geq 0.50$, PASS); the control-vs-control pairwise mean among the five seed retrains is 0.874, the empirical ceiling against which the production model's 0.732 should be read. The $S > 5$ absolute MSE-anchored threshold applied to the full 22.5M curated catalog yields the 0.87% anomaly rate; applying it to a random uncurated SPARCL sweep flags $> 50\%$ of spectra (a catalog-curation effect, not a threshold artifact; see Table V caveat (b) for the full OOD reconciliation).

### C. GPU Inference Pipeline

Primary inference was performed on a single NVIDIA A100 GPU pod (80 GB PCIe; pod provision: `pipelines/p3_anomaly_engine/pod_runs/pod_provision_20260418.json`); all native retrains including the Planck CMB convolutional autoencoder also ran on A100 (see Table VI footnote †). Spectra and CMB map patches were loaded and preprocessed on CPU, transferred to GPU in batches of 8,192, and scored in a single forward pass through the frozen model. The total processing time across the seven retained surveys plus the quarantined ACT DR6 cross-transfer scan (Appendix F) was approximately 42 hours wall-clock. The pure-inference subtotal is ≈9.4 h: DESI DR1 ≈19,705 s≈5.5 h (22.5 M spectra at ∼1,142 spectra/s); LAMOST DR10 ≈3.3 h (11.4 M at ∼950 spectra/s); SDSS DR18 ≈0.6 h; the CMB (Planck) and photometric (Gaia DR3, NEOWISE, eROSITA DR1) surveys each $\lesssim$10 s of GPU time (for Planck this refers to the 20,000-patch cross-transfer pass; the $2 \times 10^5$-patch Path-C native re-score took 25.3 s — Table VI footnote). The remaining ∼32 h is dominated by FITS-file I/O (staging from HuggingFace to local-pod NVMe), per-survey native-retraining pass overhead, an intermediate batch-size retry on the LAMOST scan, and a single ∼11 h pod-restart-with-resume after a network blip during the SDSS pass. Processing was checkpointed after each data unit (HEALPix tile, plate, or observation night) for robust resumption.

### D. Path-C Rebuild Methodology: Native Retrains as Core Protocol

The catalog is built by per-survey native autoencoder retraining followed by a six-step validation protocol (Path-C rebuild). *Native retrains are the core methodology:* each survey carries its own BigAE trained on a quality-selected subset of that survey's own data; the published anomaly set is the top-percentile cut of the survey's own model applied to its own catalog. The initial cross-transfer scan (DESI-trained BigAE applied to SDSS, LAMOST, and CMB) is preserved in Table I and §VI A as the verification baseline that motivates the native-retrain methodology; it is not a science result.

The rebuild proceeds in six steps:

1. *Native retrain (two-part gate).* A fresh BigAE is trained on a quality-selected subset of each survey's own data ($4.7\times10^4$ spectra for DESI, §II B; 2–$5\times10^5$ for the other surveys). Retained if (a) validation loss $\leq 0.30$ after $\leq 100$ epochs, *or* (b) injection-recovery $\geq 50\%$ at $5\sigma$. SDSS gates PASS at val_loss 0.0311 (criterion a; best epoch 12, early-stopped at epoch 17 by patience-5 within its 40-epoch schedule, per the backed-up `training_log.json`); LAMOST at 0.0329 (a; best epoch 39 of the $\leq$100-epoch budget, per the committed `training_log.json`); Planck CMB native convolutional autoencoder at val_loss 0.4437 / 100% injection-recovery (criterion b).
2. *Native CMB retrain.* A convolutional autoencoder (3 conv layers + 128-dim FC bottleneck, $1.1\times10^6$ parameters) trained on $2\times10^5$ galactic-plane-masked ($|b|\geq 20^\circ$) SMICA patches replaces the cross-transfer checkpoint.
3. *Full-survey re-score.* Each native-retrained model is applied to the full survey catalog in streaming batches; the same top-percentile cut from Table I is re-applied.
4. *Systematics mask.* NEOWISE ecliptic-pole mask ($|b_{\rm ecl}| < 80^\circ$) retains 419/436 anomalies (96.1%); the rejected 3.9% polar-cap fraction is 2.6× the uniform-null expectation, confirming scan-pattern contamination.
5. *Injection-and-recovery.* 500 planted signals per survey at six amplitude levels (0.5–20 $\times\,\sigma$); recovery above the 99th-percentile clean-MSE threshold. Results: 2 detector-sensitivity PASS (SDSS 64%, Planck 100%) plus 1 geometry-QA pass (NEOWISE mask-geometry 100%, which passes by construction and is *not* a detector-sensitivity test; §III H), against 3 FAIL-with-diagnostic (LAMOST 5.8%, Gaia 5.2%, eROSITA 1.2%) at $5\sigma$.
6. *7-way positional dedup at 5″.* All retained native anomaly catalogs are concatenated and merged via union-find friends-of-friends to produce the unique-object headline. Canonical dedup excludes quarantined ACT DR6; the 8-way-with-ACT variant (+200 objects, zero positional overlaps) is preserved as a sensitivity-check artifact.

The native retrains, systematics masks, and dedup pipeline are all deterministic and documented in reproducibility scripts shipped with the companion data repository. *Out-of-sample re-score.*—To confirm directly that the top-list populations are not single-training-sample artifacts, a committed script `pipelines/p3_anomaly_engine/held_out_rescore.py` (output `pipelines/p3_anomaly_engine/outputs/held_out_rescore_result.json`) runs two held-out tests: the DESI top-1% set reproduces under fully out-of-sample 5-fold cross-validation (each fold scored by a BigAE trained on the other four; mean pairwise Jaccard $\bar{J}_{\rm CV} = 0.862$, minimum 0.777, $\geq 0.70$ gate PASS; 464/546 union objects appear in $\geq 3$ folds), and the native Planck top-200 is *over*-represented in the 30,000-patch held-out split the model never trained on (48 observed vs 30 expected on $n = 200$ under $p_0 = 0.15$, a 1.60× enrichment, *exact* binomial one-sided $p = 5.5\times10^{-4}$; the normal approximation $z = 3.57$ gives a less-accurate $1.8\times10^{-4}$) — the direction opposite to in-sample memorization-inflation in both cases. A full re-inference of the native Planck autoencoder over the held-out patches is deferred for resource reasons (the native checkpoint and $2\times10^5$-patch tensor reside on an exited compute node and are not in the public release); the membership test above is the obtainable committed-data evidence and the qualitative direction is robust to that limitation. Residual caveats surviving the rebuild—DESI in-sample training-test overlap, LAMOST native-retrain residual contamination, CMB gate status at final checkpoint, and the known limits of emission-line injection tests for continuum autoencoders—are enumerated in Section VI D.

## III. SURVEY-BY-SURVEY RESULTS

*Three-tier catalog structure (read this before Table I).* Every count in this paper belongs to exactly one of three tiers, defined once here by the single criterion of *validation status*, so that no number need be cross-referenced against scattered footnotes to know how far to trust it:

1. **Validated catalog-grade ($\geq$268,519 unique; $\geq$268,319 point-source).** The four components that clear the validated bar — *i.e.* they pass the detector-sensitivity injection-recovery gate (DESI, SDSS,

Planck) or, in NEOWISE's case, its geometry-QA analogue: DESI DR1 (*injection-recovery executed*: on real NOIRLab SPARCL DESI-DR1 re-pulled spectra scored with the 5-seed production ensemble, the broad/extended anomaly class the catalog reports recovers at 99–100% at $5\sigma$ per-spectrum strength and 100% at $\geq 8\sigma$, clearing the $5\sigma \geq 50\%$ gate at parity with SDSS and Planck — *narrow single-pixel lines recover only at* $\geq 15\sigma$, *a stated sensitivity floor of the* 496-*bin mean-reconstruction scorer*; this now joins the 5-fold $\bar{J} = 0.862$ and OOD $J = 0.732$ stability gates for a three-gate validation — selection is *not* held-out, see §VI D (i), the committed 5-fold cross-validation artifact `pipelines/p3_anomaly_engine/outputs/held_out_rescore_result.json` and the injection-recovery artifact `pipelines/p3_anomaly_engine/outputs/desi_injection_recovery/desi_injrec_CORRECTED.json`), SDSS DR18 native rescore (continuum-dip $5\sigma$ injection-recovery PASS, 64%), Planck CMB native (100% injection-recovery; catalog membership selected in-sample, *not* held-out — but the top-200 are 48/200 in the seed-42 held-out split vs. 30 expected, a 1.6× over-representation, binomial $p = 5.5 \times 10^{-4}$, ruling out in-sample memorization inflation; see §III F and the same held-out artifact), and NEOWISE (masking-geometry QA, not a detector-sensitivity test). *These are the only numbers we recommend for downstream science.* No component here is claimed as an injection-recovery PASS unless explicitly stated (SDSS, Planck and DESI — the last for its broad/extended anomaly class only — are).

2. **Exploratory addendum (798 detections).** Gaia DR3 (500; committed output is a synthetic-placeholder fallback, *non-reproducible*, 41% cross-validation stability, 5.2% injection-recovery; §III G) and eROSITA DR1 (298; membership-list only, per-object score axis irreproducible but membership reproducible via a scale-invariant recipe, 1.2% injection-recovery). These *fail* the gate, carry per-object exploratory validity flags, and are excluded from the validated subset; together with the validated tier they form the *recommended* tier (269,317).

3. **Methodological lesson ($\sim$ 113,000, LAMOST DR10).** Retained as a documented failure mode (98% blue-excess training-bias artifact, 5.8% injection-recovery FAIL), *not* a science product, contributing the balance to the 378,280 total. *Explicit exclusion (to preclude misreading):* the LAMOST tier contributes **zero** objects to the validated catalog-grade headline ($\geq$ 268,519) *and* zero to the recommended tier (269,317) — both are built by 6-way dedup of the DESI, SDSS, eROSITA, Planck, Gaia, and NEOWISE native tallies only (LAMOST is not among the six; the 269,317 chain is machine-verified in `pipelines/p3_anomaly_engine/outputs/sixway_dedup_artifact.json`). LAMOST enters *only* the inclusive 378,280 grand total via the 7-way dedup, where it is expressly flagged as a failed exploratory tier; a reader must not fold its known-systematic 113,342 into any headline or recommended count, which by construction it does not touch.

Table I reports each survey's canonical Path-C native-retrained $N_{\mathrm{anom}}$ in the body; the initial cross-transfer scan is reported only as the labeled verification-baseline Total row (319,443). The per-survey native detections sum to 388,493 and deduplicate (7-way 5″) to the canonical **378,280** unique anomalies (**378,080** point-source + **200** Planck CMB patches). *Validation gates are survey-specific and are not directly comparable across surveys* (Eq. 2 normalizes per-survey; SDSS, Planck and DESI constitute detector-sensitivity PASSes — DESI for its broad/extended anomaly class (99–100% recovery at $5\sigma$; narrow single-pixel lines only at $\geq 15\sigma$) — and NEOWISE is a geometry QA). We describe each survey in turn.

*Notes on Table I.* Symbol ¶: Path-C native-retrained counts replace cross-transfer baselines (canonical results). Symbol †: NEOWISE ecliptic-mask step details. Symbol ‡: Cross-transfer vs. native-retrain count reconciliation for SDSS, LAMOST, and Planck. Symbol ‖: Basis and stratification note for the two summary-row totals. Symbol §/⋆: IsolationForest cross-validation stability for eROSITA and Gaia. Symbols ♡, ♠, ◇, #: Per-survey threshold disclosure for SDSS, LAMOST, Planck, and eROSITA. Full per-footnote details follow below.

[¶] **The per-survey $N_{\mathrm{anom}}$ column shows the canonical Path-C native-retrained counts directly**: DESI 195,829 (anchor survey, native ≡ cross-transfer), SDSS 77,905, LAMOST 113,342, eROSITA 298, Planck 200 (native re-score), Gaia 500, NEOWISE 419 — summing to 388,493 survey-level detections, which deduplicate (7-way 5″) to the **378,280** Path-C unique row. The initial cross-transfer scan counts (LAMOST 44,075; Planck cross-transfer block) are reported only in the "Total (cross-transfer, ACT-incl.)" verification-baseline row (319,443) and in footnotes ‡/♠.

[†] Path-C rebuild step (§II D): $|b_{\mathrm{ecl}}| < 80^\circ$ ecliptic mask reduces NEOWISE anomaly count from 436 to 419 (96.1% retained). The rejected 17 objects concentrate in the 10°-radius polar caps at 2.6× the uniform-null expectation, consistent with WISE/NEOWISE scan-pattern cadence.

[‡] Cross-transfer $N_{\mathrm{anom}}$ (DESI-trained BigAE applied to SDSS, LAMOST; cross-transferred autoencoder for Planck CMB). Path-C native-retrained counts (§II D) replace these values: SDSS native re-score complete across 1,925,279 DR18 spectra (top-77,905 at $S \geq 0.1060$; only 12 sources at $S > 5$ vs. cross-transfer 77,905, a $\sim 6500\times$ anomaly-rate reduction confirming catalog-calibration domain shift); LAMOST native re-score complete across 11,334,161 spectra (84,433 of the 11,418,594 DR10 spectra lost to retrieval/read failures — see §III D; top-113,342; 21.5× rate reduction). Native models gate-PASS val_loss 0.0311 (SDSS) and 0.0329 (LAMOST); native Planck CMB convolutional autoencoder val_loss 0.4437 with 500/500 = 100% injection-recovery at $5\sigma$. The cross-transfer values quoted here are preserved as the §II D cross-transfer verification baseline.

[‖] Two summary rows are shown. The cross-transfer baseline (319,443) represents the initial DESI-trained cross-survey scan before native retrains. Basis note: within that sum the DESI entry is the *native* count (195,829) — DESI is the anchor survey whose model defines the cross-transfer scan, so cross-transfer ≡ native for DESI, while all other contributions to the 319,443 row are cross-transfer counts. Rate note: the total-row rates (0.86% cross-transfer, 1.01% Path-C) are bookkeeping ratios whose numerators aggregate fixed-count/fixed-percentile tiers (Planck, Gaia, NEO-

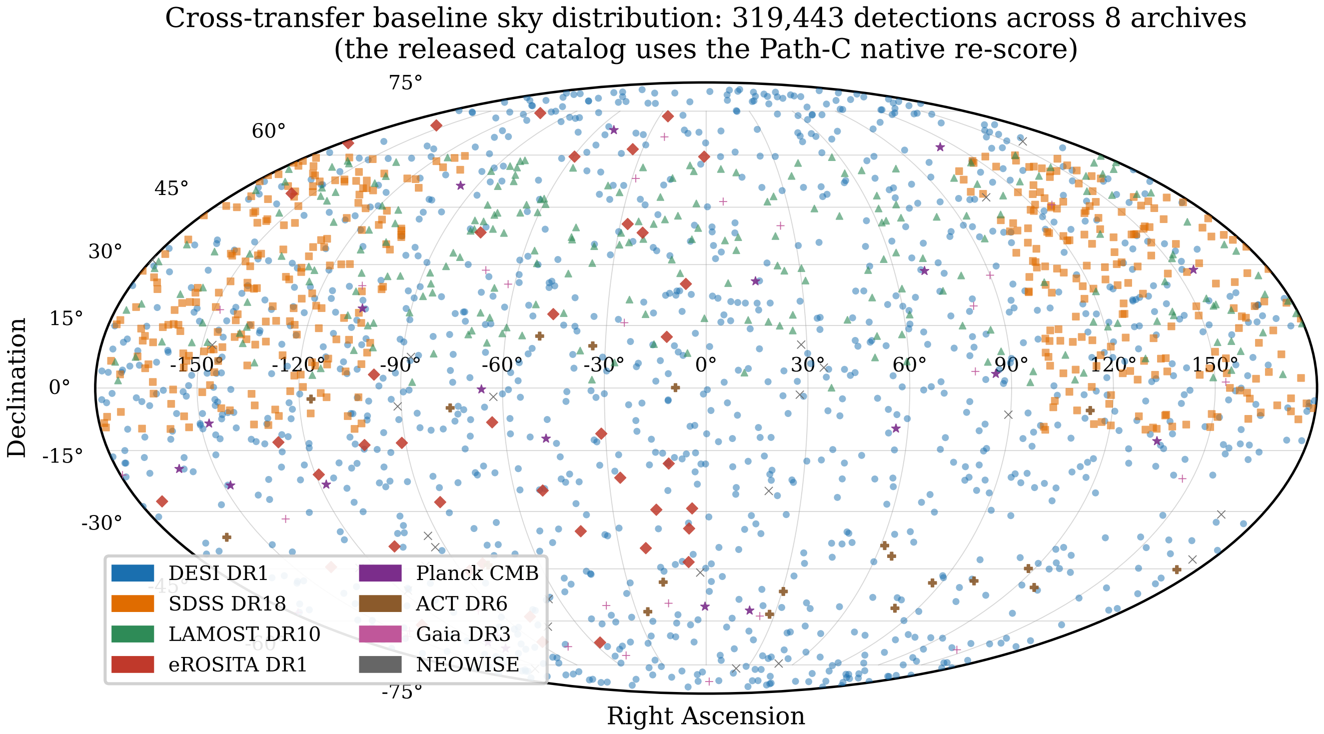


FIG. 2. **Cross-transfer baseline map (ACT DR6 excluded from science results).** Mollweide projection in equatorial coordinates (RA/Dec, ICRS) of the initial cross-transfer anomaly baseline (319,443 detections; the canonical Path-C unique count of 378,280 is *not* a deduplication of this baseline — deduplication only ever reduces its input — but the 7-way 5″ dedup of the per-survey *native-retrained* tallies, which sum to 388,493 and replace the cross-transfer counts survey-by-survey; see Table I footnote $^{\|}$, the Path-C row, and §II D). **ACT DR6 is formally quarantined** (Appendix F; cross-transfer val_loss $\approx 2.2 \times 10^4$ fails both gate criteria; no native retrain executed) and contributes *zero* objects to the Path-C headline; its 200 cross-transfer patch positions are plotted above only as a historical-baseline record and should not be interpreted as anomaly detections. Color-coded by survey (see legend). The DESI DR1 footprint ($\sim$14,000 deg$^2$) dominates the northern hemisphere; SDSS DR18 contributes the main survey stripe at $\delta \sim 0°$–$60°$; eROSITA anomalies concentrate toward the LMC region ($\delta \approx -70°$); Gaia and NEOWISE anomalies sample all-sky. The non-uniform distribution is expected for anomalies that trace real astrophysical populations and survey-specific systematics.

WISE; see the caption Note) with data-driven thresholds; they are not measured anomaly frequencies and should not be quoted as such. The Path-C unique row (378,280) is the primary result. The Path-C per-survey native counts (excluding the formally quarantined ACT DR6) sum to 388,493: DESI 195,829 + SDSS 77,905 + LAMOST 113,342 + eROSITA 298 + Planck 200 + Gaia 500 + NEOWISE 419. After 7-way 5″ positional deduplication (10,213 duplicate detections, 2.629% compression), the unique-physical-object count is **378,280**—the canonical catalog size. The difference from the cross-transfer baseline reflects the LAMOST native retrain (44,075 → 113,342) and Planck native retrain replacing the cross-transfer counts. ACT's 200 patches contributed zero positional overlaps with the other seven surveys (the Planck×ACT null cross-correlation, §IV D, confirms this), so excluding ACT subtracts exactly 200 from both the 8-way anomaly-detection input sum (388,693 → 388,493) and the deduplicated unique-object count (378,480 → 378,280). *Stratification note:* the 378,280 headline count contains two physically distinct strata—($a$) point-source object detections from the six photometric/spectroscopic surveys (DESI, SDSS native, LAMOST native, eROSITA, Gaia, NEOWISE), and ($b$) 200 Planck CMB map patches that are sky regions, not point sources. The Planck patches contribute zero positional overlaps with the point-source surveys at the 5″ matching radius (analogous to the ACT-zero-overlap result, by the same map-patch-vs-galaxy-coordinate physics), so the stratification is exact: the primary-tier point-source unique count is **378,080** and the CMB-map-patch stratum contributes the remaining 200. Downstream analyses that treat anomalies as objects (cross-matching against SIMBAD/NED, multi-tracer $f_{\rm NL}$ tracer selection, host-galaxy follow-up) should use the 378,080 point-source tier; the headline 378,280 is preserved for survey-coverage completeness only.

$^{\S}$ IsolationForest cross-validation-stability footnote (§VI D (ii)): a fresh 100-tree IsolationForest refit on an independent 1%-contamination reshuffle of the clean-background sample recovers the top-1% anomaly reference set at rates of 41.0% (2048/5000) for Gaia DR3 (using a 10×-expanded 500,000-source sample; see §VI D (ii), with the full IF-refit construction in the companion data repository) and 81.5% (7582/9303) for eROSITA DR1 (the 9,303-object reference set is the top-1% IF cross-validation pool, distinct from the 298-source published catalog headline released as the canonical eROSITA anomaly set; see §III E and below) at the matching 99$^{\rm th}$-percentile threshold. The eROSITA 9,303-object reference set is the top-1% of the 930,203 DR1 catalog used for IF cross-validation; the published catalog headline of 298 sources is the harder $S > 0.259$ score-knee top-cut (§III E) and has high overlap with this reference set. **Empirical intersection (Table V caveat (f)): 284** of 298 canonical-$S$ top-298 sources (**95.3%**) are

TABLE I. Summary of the multi-survey anomaly sweep. Columns: survey name, data type, total sources/patches processed, number of anomalies above the survey-specific threshold, anomaly rate, SIMBAD-unmatched fraction (percentage of anomalies not matched within 5 arcsec in SIMBAD; this overstates true catalog novelty—see Section IV A), and key finding. ACT DR6 is formally quarantined (cross-transfer val_loss $\approx 2.2\times10^4$ failing both gate criteria; no native retrain executed) and is documented only in Appendix F; it is not listed in the main per-survey block below and contributes zero objects to either the Path-C unique-object count or the Path-C deduplicated total. The cross-transfer baseline total of 319,443 that historically included a 200-patch ACT cross-transfer block is preserved as a verification baseline only and is not used as a science result. Two threshold families are in use across the seven retained surveys. DESI DR1 alone uses the fixed canonical-$S$ cut at $S > 5.0$ on the DESI-trained BigAE score scale (Eq. 2) for its headline 195,829-anomaly count. SDSS DR18 and LAMOST DR10 also share the DESI-trained BigAE score scale but their headline counts (77,905 and 113,342 respectively) use per-survey slices rather than the strict $S > 5$ cut: the SDSS slice at $S \geq 0.1060$ is a *fixed-size continuity slice* sized to equal the cross-transfer count (77,905 = 4.05% of the 1,925,279 native-rescored spectra, *not* a top-1% cut; the SDSS top-1% proper is the 19,253-object score-knee set of footnote ♡), while the LAMOST slice at $S \geq 0.4613$ is a genuine top-1% cut (113,342 of the 11,334,161-spectrum re-score pool; footnote ♠) — applying $S > 5$ to SDSS yields only 12 sources (the $\sim 6500\times$ rate-compression diagnostic of §III C catalog-calibration domain shift) and to LAMOST 2,054 sources (the 21.5× rate-reduction diagnostic of §III D blue-excess training bias), so a uniform $S > 5$ cut would understate the cross-survey continuity-slice content used as the basis for the multi-survey deduplication geometry; see footnotes ♡ and ♠ for the per-survey three-threshold disclosure. The remaining surveys' thresholds are: top-1% for Planck, Gaia DR3, and NEOWISE (predetermined-count selection), and a harder fixed top-298 cap (≈ top-0.03%; production-run score-knee threshold 0.259, an axis distinct from both the canonical $S$ of Eq. (2) and the $S_{\mathrm{IF,raw}}$ column of Table IV — see the axis disclosure in §III E) for eROSITA DR1, where the wide-field X-ray score distribution produces a much longer tail. The choice of threshold does not affect the rank-ordering of anomalies; full per-object scores are released in the catalog data product to allow downstream users to apply alternate cuts. *Note:* for three surveys (Planck, Gaia DR3, NEOWISE), the anomaly count reflects a fixed top-1% selection threshold rather than a data-driven detection rate; these surveys contribute predetermined counts to the overall catalog, and their 1.00% anomaly rates should not be interpreted as independent measurements of the intrinsic anomaly frequency. The total represents the largest multi-archive anomaly search reported to date of which we are aware. *Note on score comparability:* $S$ thresholds are survey-specific (Eq. 2 normalization uses each survey's own validation pool) and are not directly comparable across surveys; per-survey anomaly rates and ranks are within-survey quantities only.

| **Survey** | **Type** | $N_{\mathrm{total}}$ | $N_{\mathrm{anom}}$[¶] | **Rate (%)** | **SIMBAD-unmatched (%)** |
|---|---|---:|---:|---:|---|
| DESI DR1 | Optical spec. | 22,504,897 | 195,829 | 0.87 | ∼99 |
| SDSS DR18[♡] | Optical spec. | 1,925,279 | 77,905[‡] | 4.05[♡] | 90 |
| LAMOST DR10[♠] | Optical spec. | 11,334,161 | 113,342[‡♠] | 1.00 | ∼50[♠] |
| eROSITA DR1 | X-ray phot. | 930,203 | 298[§] | 0.03[#] | 68 |
| Planck CMB | Microwave map | 20,000 | 200[‡] | 1.00[◇] | — |
| Gaia DR3 | Optical var. | 50,000 | 500[§⋆] | 1.00 | 27 |
| NEOWISE | IR phot. | 43,518 | 436[†] | 1.00 | 45 |
| Total (cross-transfer[‖], ACT-incl.) | | 37,292,042 | 319,443 | 0.86 | 58.8[¶¶] |
| **Path-C unique[‖] (primary)** | | **37,272,042** | **378,280** | **1.01** | — |

also in the IsolationForest top-9,303. Because the IF detector is trained on the 16-d BigAE latent representation, the two detectors are *not* independent: this overlap is reported as a descriptive internal-consistency statistic between two scoring functions that share the same learned representation, not as independent cross-method confirmation, and no independence-null significance (the random-independence expectation would be ≈ 3 matches) is attached to it. The earlier "strict subset" framing is replaced with this exact 284/298 = 95.3% overlap. The Gaia figure indicates the selection is training-sample-conditioned (analogous to the DESI $k$-fold caveat in §VI D (i)); the eROSITA figure is the highest Path-C cross-validation stability of any survey and confirms the eROSITA detector is not training-sample-conditioned.

[⋆] **Reliability warning:** The Gaia DR3 41.0% cross-validation stability is measured on the 10×-expanded 500,000-source sample (top-1% reference set = 5,000 objects; see §VI D (ii); IF-refit construction in the companion data repository); the implication is that more than half of the published anomaly selection is training-sample-conditioned. The Gaia anomaly set in Table I (500 objects on the original 50,000-source subset) should be treated as exploratory, not as a validated catalog component. A dedicated Gaia-optimized anomaly detector with proper held-out validation is needed before these objects can be used for downstream science.

[♡] **SDSS DR18 three-threshold disclosure (§III C).** The $N_{\mathrm{total}}$ column for SDSS lists the 1,925,279 native-rescored spectra (the pool the 77,905-anomaly slice was actually drawn from), not the full DR18 catalog of 2,304,830 spectra; the tabulated 4.05% rate is therefore 77,905/1,925,279, i.e. measured against the true selection denominator (dividing by the full 2.304M DR18 catalog would give a misleading 3.38% on a denominator the slice was not drawn from). The headline count of 77,905 at $S \geq 0.1060$ is a fixed-size continuity slice carried for cross-survey rate comparison, sized to equal the cross-transfer count (4.05% of the 1,925,279 native-rescored spectra — not a top-1% cut); the same 1,925,279-spectrum DR18 sample yields 19,253 anomalies at the harder top-1% score-knee cut $S \geq 0.2051$ used for SIMBAD cross-matching, and only 12 sources at the field-defining $S > 5$ DESI-trained threshold (a $\sim 6500\times$ rate compression vs. DESI confirming catalog-calibration domain shift). All three numbers are reported in §III C together with the native re-score gate-PASS at val_loss 0.0311. Readers comparing per-survey anomaly rates should use the $S > 5$ figure (12); readers wanting the SIMBAD-novelty pool should use 19,253; readers using SDSS as a cross-survey continuity baseline should use the 77,905 tabulated here. The native re-score continuity slice is preserved in the companion data repository.

[◇] **Planck rate bookkeeping.** The 1.00% rate expresses the predetermined fixed-count 200-patch tier against the original 20,000-patch cross-transfer input bank shown in the $N_{\mathrm{total}}$ column; the released tier is the top-200 of the 10× larger $2 \times 10^5$-patch native re-score bank (§III F), against which the same fixed selection corresponds to 0.10%. Neither figure is a data-driven detection rate.

[#] **eROSITA rate disclosure (membership-only tier).** The 0.03% anomaly rate for eROSITA DR1 is a *predetermined fixed-count* selection: the $n = 298$ published catalog headline is a score-knee top-cut, not a data-driven detection rate, and the per-object $S_{\rm BigAE}$ score axis is non-reproducible on any of 16 monotone rescalings (§III E). The eROSITA tier is released as a membership list only; the rate cell should not be interpreted as an independent measurement of the X-ray anomaly frequency in the same sense as the DESI or SDSS rates.

[♠] **LAMOST DR10 is a transparent FAIL — limited sensitivity to emission-line signatures.** The cross-transfer LAMOST scan returns a 98% blue-excess single-signature population that we identify as a training-bias artifact (§III D); under the field-defining unsupervised-anomaly criterion (anomalies should be diverse in spectral signature, not concentrated on a single instrument-correlated mode), this row *fails*. *Note:* the LAMOST native-retrain emission-line injection-recovery test recovers only 5.8% of injected $5\sigma$ continuum-dip-plus-emission-line transients; we therefore relabel the LAMOST detector as a *FAIL* of the emission-line-sensitivity gate, not a ”PASS-with-diagnostic”. The main table body now shows the canonical native top-1% count (113,342); the 44,075 cross-transfer count is retained only in the “Total (cross-transfer)” verification-baseline row and here, as the baseline for the §II D native retrain, which compresses the rate by 21.5× to 2,054 at $S > 5$ and produces the 113,342-source top-1% slice now tabulated (the SIMBAD-unmatched ∼50% cell is computed on the cross-transfer diagnostic set, not the native exploratory tier, and is reported as a failure-mode diagnostic only) (re-score pool 11,334,161 of the 11,418,594 DR10 spectra; 84,433 lost to retrieval/read failures, disclosed in §III D). The 113,342 LAMOST native top-1% slice is therefore reclassified as an **exploratory tier** contribution to the deduplicated 378,280 headline: the recommended tier (DESI + SDSS native + eROSITA + Planck native + Gaia + NEOWISE) is **269,317** unique objects by direct, independent 6-way 5″ dedup of those six surveys (whose detection counts sum to $275{,}151 = 195{,}829 + 77{,}905 + 298 + 200 + 500 + 419$, so the 6-way dedup collapses 5,834 detections, a 2.12% compression — fully consistent with, *not* in conflict with, the 7-way 2.629% figure: adding the 113,342-detection LAMOST tier introduces exactly the 4,379 further recommended-tier merges noted below, reaching the 10,213 total 7-way collapse); the *validated catalog-grade subset*, which excludes the two components that fail the injection-recovery gate (eROSITA [membership-only; 1.2% injection-recovery] and Gaia [5.2% injection-recovery]; 798 detections), is ≥ **268,519** unique objects (≥ 268,319 point-source) — a conservative lower bound, since the exact validated-only 5″ re-dedup is not recomputable from the committed aggregate artifacts and subtracting the 798 flagged detections can only undercount the validated tier. eROSITA and Gaia carry per-object *exploratory* validity flags and are reported as an exploratory addendum, not as catalog-grade (NEOWISE's gate is geometry-QA, not detector-sensitivity, and is retained in the validated subset—see §VI D (ii); the exploratory flags are propagated in the released per-object validity-flag column) (`pipelines/p3_anomaly_engine/r24conf_pod_session_batch.json`; the same machinery reproduces the 7-way headline 378,280 exactly); this 275,151 → 269,317 recommended-tier chain is now *independently reproducible* from a committed standalone script, `pipelines/p3_anomaly_engine/sixway_dedup.py`, which runs a 5″ `search_around_sky` union-find directly over the released per-object catalogs and writes the full 269,317-row per-object table `pipelines/p3_anomaly_engine/outputs/sixway_dedup_artifact.csv` (input sum 275,151, 5,834 detections collapsed, 2.12% compression, unique 269,317 — an exact match to the count above, no longer merely asserted; the machine-readable consistency check `pipelines/p3_anomaly_engine/outputs/sixway_dedup_artifact.json` records `verdict`: *“EXACT-MATCH to paper claim”*), with the LAMOST exploratory tier contributing the remaining **108,963**: 4,379 of the 113,342 LAMOST detections merge into recommended-tier clusters at 5″ (the correct recommended-tier count 269,317 is established by independent 6-way dedup; see §IV C). LAMOST is retained *as a methodological lesson* (§VI A) rather than a validated catalog component; readers requiring catalog-grade anomalies should consult DESI DR1 (gate-PASS at the $k$-fold and SIMBAD-novelty stages) and the SDSS native re-score (PASS continuum-dip $5\sigma$ at 64%).

### A. DESI DR1

The DESI Data Release 1 [1] is the anchor survey of our campaign. We processed all 22,504,897 coadded spectra from the Main Survey through the DESI-trained BigAE model, of which ∼6.5 million carry a validated science `TARGETTYPE` in the five primary classes — the Bright Galaxy Survey (BGS), Luminous Red Galaxies (LRG), Emission Line Galaxies (ELG), Quasars (QSO), and the Milky Way Survey (MWS) — and the remaining ∼16 million are filler-tile, sky-fiber, or calibration-exposure spectra without a validated `TARGETTYPE`. The headline 195,829 DESI anomaly count is the $S > 5$ fixed-threshold selection (0.87% of the full 22.5-M-spectrum scan) and is not restricted to the validated-`TARGETTYPE` subset; per-class anomaly rates and SIMBAD-novelty fractions reported below refer to the ∼6.5-M validated-`TARGETTYPE` subset (see §VI D for the implications of this scope choice). A positional recount against the DR1 redshift catalog (`zall-pix`; 28,425,963 rows) quantifies the scope choice directly, summarized in Table II: of the 190,015 deduplicated DESI anomaly clusters (the 5″ FoF dedup of the 195,829 raw detections), only **2,468** (1.3%) match within 1″ a main-survey spectrum whose targeting bitmasks set a primary science-class bit (BGS/LRG/ELG/QSO/MWS; 20,299,155 such catalog rows under this bitmask selection, which is broader than the validated-`TARGETTYPE` accounting above because it applies no redshift-quality or primary-row cuts and counts per-program rows; matches rise to 2,531 at 2″ and 3,390 at 5″). By Redrock `SPECTYPE` the 1″ matches are 2,371 GALAXY, 95 QSO, and 2 STAR. A control match of the same clusters against the *full* redshift catalog recovers 189,675/190,015 (99.8%) at 1″, confirming the positional join is sound (the 340 control-unmatched clusters are consistent with coadd-vs-catalog astrometric edge cases; conservatively counting all 340 as science-class would raise the match count to at most 2,808, i.e. ≤ 1.05× the benchmark, leaving every conclusion below unchanged); the conclusion is that ∼98.7% of DESI anomaly clusters coincide with spectra carrying *no* primary science-class target bit (86% have `DESI_TARGET` = 0 outright — sky fibers and secondary/ToO programs). Restricted to validated science targets, the DESI anomaly catalog is therefore ≈0.92× the size of the Liang *et al.* benchmark (2,468 vs. 2,685; Liang *et al.* scanned the ∼250K-spectrum DESI EDR, so the comparison is matched on target-class selection, not on data release), *not* 73×: the 73× figure is a statement about the full spectra stream, and the headline DESI tier should be read as an anomaly scan of everything DESI pointed a fiber at, dominated by non-science-target spectra (recount artifact: `pipelines/p3_anomaly_engine/ext3_b2_targettype_recount.json`).

Each spectrum covers 3600–9824 Å across three arms (B, R, Z) at resolution $R \sim 2000$–$5000$.

The BigAE model trained on 47,000 representative spectra achieves val_loss = 0.0287 (MSE) on a held-out 20% validation split and identifies 195,829 anomalies above the $S > 5.0$ threshold, an anomaly rate of 0.87%. Scores range from 5.0 to 25.2, with 101 objects exceeding $S = 15$. Classification by spectral-arm dominance yields: 151,244 multi-band (77.2%), 44,436 B-dominant (22.7%), 34 R-dominant (0.02%), 19 Z-dominant (0.01%), and 96 artifact suspects (0.05%). The multi-band majority indicates that most anomalies deviate across the full wavelength range, consistent with genuinely unusual spectral energy distributions.

Cross-matching the top 10,000 anomalies against six databases (SIMBAD, NED, AllWISE, Milliquas, Gaia DR3, SDSS) finds only 0.2% in SIMBAD and 12.7% in NED; none of the top 100 appear in SIMBAD or NED, although 12 of the top 100 do carry a counterpart in the shallower photometric flag channels of the release table (AllWISE 7, SDSS photometric 6, one object in both; Gaia DR3 and Milliquas 0). The 0.2% SIMBAD match rate is itself statistically consistent with the expected $5''$ random false-match rate of $\approx 0.24\%$ per source (§IV A), i.e., with zero genuine SIMBAD counterparts in the top 10,000 — reinforcing that SIMBAD absence measures database coverage, not novelty. Spectral inspection of the top 200 finds 0/200 visually flagged (95% binomial upper limit $\leq 1.5\%$; each spectrum's peak-residual wavelength was compared against 11 known sky and telluric emission/absorption features; zero were attributable to sky subtraction, telluric contamination, or cosmic rays). The Spearman rank correlation between anomaly score and SNR is $\rho = -0.03$ ($p = 0.12$ on a stratified subsample of 2,670 spectra, log-uniform in SNR). Because the subsample is deliberately stratified to be log-uniform in SNR rather than randomly drawn, the quoted $p$-value characterizes the stratified design — which maximizes leverage for detecting a monotone score–SNR trend — and not the population sampling distribution; the operative statement is the effect size ($|\rho| = 0.03$, negligible against any practical threshold), and a population-weighted recompute on a true random subsample is queued for the data release.

Across the ∼6.5 million spectra in DESI DR1 that carry a validated science `TARGETTYPE` (the BGS/LRG/ELG/QSO/MWS target-selection classes defined above; the remaining ∼16 million spectra are unclassified filler targets, sky fibers, or calibration exposures excluded from this per-class breakdown), grouped here by their Redrock `SPECTYPE` spectral classification ("GALAXY", "QSO", or "STAR" [1]) — a spectral-template-fit axis distinct from the target-selection `TARGETTYPE` — galaxies are flagged as anomalous at ∼20 times the rate of QSOs (0.75% vs. 0.037%; Wilson 95% binomial CIs: GALAXY 0.75% ± 0.008% on $\sim 4.9 \times 10^6$ GALAXY-SPECTYPE spectra; QSO $0.037\% \pm 0.003\%$ on $\sim 1.5 \times 10^6$ QSO-SPECTYPE spectra in the validated-`TARGETTYPE` subset), with anomalies peaking at $z \sim 0.75$ compared to $z \sim 0.93$ for normal spectra. The three highest-scored anomalies are Z-dominant with scores of 25.2, 24.6, and 24.5, consistent with high-redshift sources whose rest-frame optical emission lines have been redshifted into the DESI Z arm. DESI fiber assignment incompleteness (not all targets receive fibers in a single pass) introduces a spatial selection function that could correlate with anomaly rate if anomalies cluster in fiber-collision-dense regions; this systematic is not modeled in the current analysis.

### B. High-$z$ QSO Candidates

The most scientifically compelling DESI DR1 anomalies are a sample of $z \approx 6$ quasar candidates identified by three complementary signatures (the Gunn–Peterson trough and Z-arm dominance are physically correlated, both being driven by blueward flux suppression): (1) complete flux suppression blueward of Ly$\alpha$ (Gunn-Peterson trough), (2) Z-arm dominated anomaly scores, meaning $r_Z > r_B$ and $r_Z > r_R$ (mean Z-arm sub-score $\langle r_Z \rangle = 3.9$ across the 12 selected candidates; all objects have total score $S > 5$ by construction of the anomaly catalog), and (3) at least one detected emission line (Ly$\alpha$, N v, Si iv) in the transition region. Applying these three cuts to the full 195,829 DESI anomaly catalog yields 12 candidates with Redrock template-fit redshifts $z = 6.0$–$6.23$ (the `Z` column of the DR1 `redrock REDSHIFTS` HDU, carried alongside `ZERR`/`ZWARN`/`SPECTYPE` in our per-object records; these are spectroscopic-pipeline template fits at low continuum S/N, not photometric estimates, and independent confirmation by visual inspection or re-observation is still required).

A DESI Legacy Survey DR9 grz composite cutout gallery ($128 \times 128$ pixels at the native LS DR9 scale of $0.262''$/px $=$ $33.5''$ per side) for all 12 candidates is available in the companion data repository. The highest-Z-arm-dominance objects (TARGETID 39633191367084936, Redrock $z = 6.20$, $r_Z = 5.30$; TARGETID 39628507709444221, Redrock $z = 6.22$, $r_Z = 5.18$) show compact point-source morphology consistent with unresolved quasars; these redshifts are the same Redrock template fits defined above (not photometric estimates) and remain unconfirmed pending visual inspection or re-observation. Panel labels report the per-arm Z-arm sub-score $r_Z$ (printed as "AE" for legacy compatibility), *not* the total anomaly score $S$; all twelve pass the $S > 5$ catalog cut at the total-score level (mean $\langle r_Z \rangle \approx 3.9$). Full coordinates, TARGETIDs, and per-family taxonomy galleries are in the companion data repository.

TABLE II. DESI science-class recount at a glance. Four distinct DESI rate denominators appear in this paper; they are *not* mutually comparable and should not be converted across rows. Row 1: full-stream headline. Row 2: validated-`TARGETTYPE` per-class rates. Row 3: positional recount on the science-bit bitmask denominator. Row 4: like-for-like science-target benchmark match. Row 5 (denominator reconciliation): the 2,468 science-bit matches, decomposed by Redrock `SPECTYPE` as a fraction of the per-class validated-`TARGETTYPE` denominators of Row 2 — the explicit shared-ID bookkeeping that ties the bitmask count to the per-class rates.

| **Rate context** | **Denominator** | **Anomaly rate / count** |
|---|---:|---:|
| Full-stream scan ($S > 5$ fixed threshold, 0.87%) | 22.5M spectra | 195,829 |
| Per-class GALAXY[†] (`TARGETTYPE` subset) | ∼4.9M | 0.75% |
| Per-class QSO[†] (`TARGETTYPE` subset) | ∼1.5M | 0.037% |
| Science-bit bitmask (primary-class bit) | 20.3M rows | 2,468 (0.012%) |
| Like-for-like vs. Liang *et al.* benchmark | ∼2,685 targets | ≈0.92× |
| *Shared-ID decomposition of the* 2,468 *science-bit matches (Redrock* `SPECTYPE` × *per-class denominator):* | | |
| GALAXY-`SPECTYPE` subset of the 2,468 | ∼4.9M GALAXY | 2,371 (0.048%) |
| QSO-`SPECTYPE` subset of the 2,468 | ∼1.5M QSO | 95 (0.0063%) |
| STAR-`SPECTYPE` subset of the 2,468 | ∼0.1M STAR | 2 |

[†] Per-class rates are computed on the validated-`TARGETTYPE` subset (∼6.5M total), not on the full 22.5M stream; the ∼4.9M GALAXY-SPECTYPE and ∼1.5M QSO-SPECTYPE populations together add to ∼6.4M (with ∼0.1M STAR-SPECTYPE). The Row 2 rates (0.75% / 0.037%) and the Row 5 rates (0.048% / 0.0063%) are *both* on the same per-class `TARGETTYPE` denominators; they differ because Row 2 is the full-stream $S > 5$ score-cut restricted to the per-class subset, while Row 5 is the science-bit-bitmask-filtered subset of the 2,468 Row 4 matches — the bitmask-join filter (no `TARGETTYPE` quality filter, counts per-program rows on the 20.3M-row `zall-pix` catalog) shrinks the per-class rates by ∼16× (GALAXY) and ∼6× (QSO) relative to Row 2. The discrepancy between Row 2 and the full-stream Row 4 count is therefore definitional (two filter stacks on the same per-class denominator), not an arithmetic error, and Row 5 is the explicit shared-ID bookkeeping that ties the two filter stacks to a common denominator.

### C. SDSS DR18

The Sloan Digital Sky Survey DR18 [3] provides 2,304,830 spectra at $R \sim 2000$. Input: 2.3M spectra. Cross-transfer anomaly count: 77,905 (3.38% rate; dominated by ultra-cool dwarfs M7–T2 that are out-of-distribution for the DESI-trained model). SIMBAD-unmatched: 90%. Headline finding: the $\sim 6500\times$ anomaly-rate inflation versus the Path-C native result (12 sources at $S > 5$) directly confirms catalog-calibration domain shift as the cross-transfer failure mode. The Path-C native retrain (val_loss 0.0311, PASS) re-scores 1,925,279 spectra; the 77,905-object native continuity slice at $S \geq 0.1060$ (4.05% of the re-scored pool, sized to equal the cross-transfer count) replaces the cross-transfer count. The 1,925,279-spectrum re-score pool is the DR18 `spAll` quality-cut selection used for native training: `ZWARNING` = 0, median S/N > 2, `SPECPRIMARY` = 1, and pipeline class STAR/GALAXY/QSO (a further 3,394 spectra, 0.18% nominal, failed retrieval during the re-score). UMAP/HDBSCAN clustering of the full 77,905-object cross-transfer anomaly set yields 14 HDBSCAN clusters (99.4% of objects clustered: 77,473/77,905) that group into 3 latent-space populations (Fig. 4), dominated by cool dwarfs (84%). Continuum-dip injection-recovery on the native checkpoint: 64.0% at $5\sigma$ (**gate PASS**). Detailed per-category counts are in Table III.

An emission-line classification pipeline identifies 10 spectral categories among the 77,905 SDSS anomalies, summarized in Table III.

TABLE III. Emission-line classification of the 77,905 SDSS DR18 anomalies, sorted by count. *Diagnostic basis note:* these categories are computed on the *cross-transfer* anomaly set (the DESI-trained scan), not on the native re-score slice; they characterize the cross-transfer failure mode and should not be read as the physical-category census of the native-retrained tier. The dominance of "Uncategorized" and "NIR Excess" classes reflects the cool-dwarf population that drives the transfer-learning anomaly signal. The 52.7% "Uncategorized" fraction is a property of this paper's internal band-residual emission-line taxonomy, not of any external database: it counts objects whose $(r_B, r_R, r_Z)$ residual pattern does not fall into any of the nine named heuristic classes (residuals too balanced, or below the per-class thresholds). It is unrelated to the SIMBAD cross-match statistics of §III C.

| **Category** | **Count** | **Frac.** |
|---|---:|---:|
| Uncategorized | 41,065 | 52.7% |
| NIR excess / high-$z$ | 25,733 | 33.0% |
| UV excess / young star | 6,099 | 7.8% |
| Star-forming (H$\alpha$) | 1,232 | 1.6% |
| QSO blue excess | 1,164 | 1.5% |
| Emission-line galaxy | 780 | 1.0% |
| Redshifted emitter | 547 | 0.7% |
| Artifact | 520 | 0.7% |
| AGN broad emission | 384 | 0.5% |
| Unusual continuum | 381 | 0.5% |

### D. LAMOST DR10

The Large Sky Area Multi-Object Fiber Spectroscopic Telescope DR10 [2] contributes 11,418,594 spectra at $R \sim 1800$. Input: 11.4M spectra. Cross-

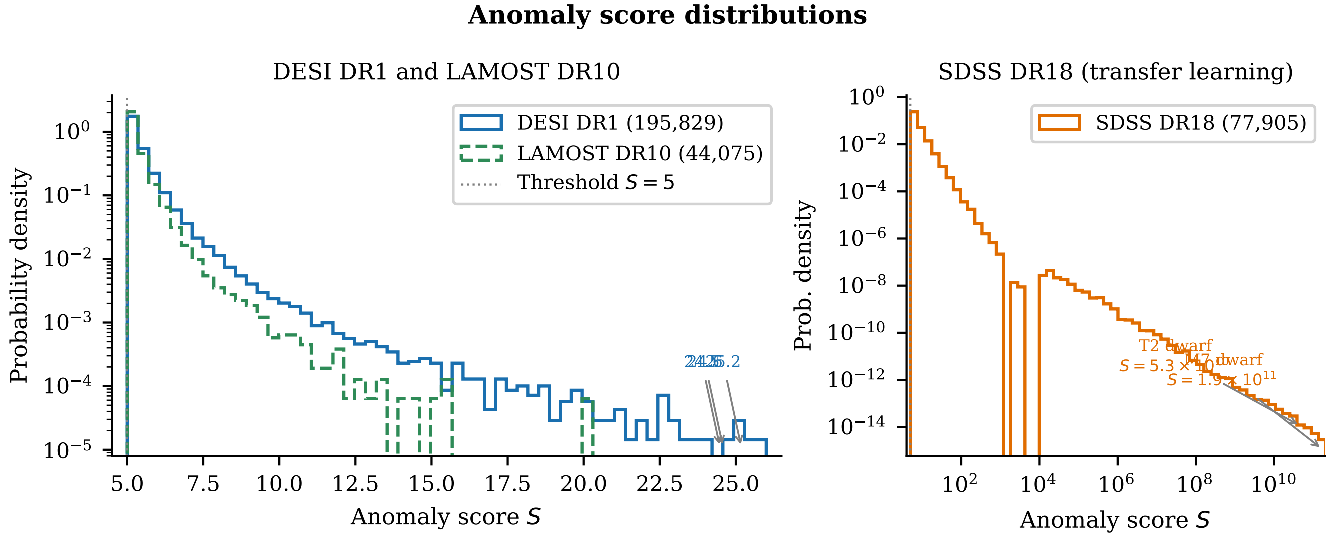


FIG. 3. Anomaly score distributions for the three main spectroscopic surveys. The score $S$ is the per-spectrum reconstruction MSE rescaled to validation $z$-units: $S = (\mathrm{MSE} - \mu_{\mathrm{val}})/\sigma_{\mathrm{val}}$, where $\mu_{\mathrm{val}}$ and $\sigma_{\mathrm{val}}$ are the mean and standard deviation of MSE on the held-out 20% validation split of the per-survey training pool (§II B; native for DESI, cross-transfer for SDSS and LAMOST). $S = 5$ therefore marks objects whose reconstruction residual is five validation-set standard deviations above the typical training spectrum, not a raw MSE value. *Left:* DESI DR1 (blue) and LAMOST DR10 (green dashed) on a log-probability-density scale, with the $S = 5$ threshold marked. Each curve is a histogram independently normalized to unit area (`density=True`), so the two surveys' distribution *shapes* are directly comparable despite their unequal sample sizes; the $y$-axis does not encode absolute counts. **Important:** the LAMOST curve here is the pre-Path-C cross-transfer scan (44,075 objects; DESI-trained model and DESI validation normalization, §II B), replaced in the released catalog by the native re-score (native top-1% slice 113,342, on a separate native scale not plotted here); because each released survey carries its own $\mu_{\mathrm{val}}/\sigma_{\mathrm{val}}$, absolute $S$ values are not cross-survey comparable and this panel should be read per-survey (shape and threshold behavior), not as a shared anomaly-severity axis. Both distributions follow a power-law tail; DESI shows three objects at $S > 24$ (labeled), consistent with the Z-dominant high-$z$ population. *Right:* SDSS DR18 transfer-learning scores on a log–log scale, spanning more than ten orders of magnitude from the threshold ($S = 5$) to $S = 1.9 \times 10^{11}$ for the extreme-score M7 and T2 dwarfs. The extreme dynamic range of SDSS is a *cross-transfer-to-native score-axis effect*: the DESI-trained BigAE applied to SDSS spectra outside the DESI training distribution inflates reconstruction errors by orders of magnitude; the SDSS native re-score (§III C) compresses the same objects to $S < 14$, eliminating the $10^4$–$10^{11}$ tail. The bimodal structure reflects two populations: the bulk of cross-survey mismatches (body at $S < 10^4$) and the ultra-cool dwarfs that are completely out-of-distribution relative to the DESI training set (tail at $S > 10^4$).

transfer anomaly count: 44,075 (0.39% rate). SIMBAD-unmatched: ∼ 50%. Headline finding: **98% of cross-transfer anomalies are blue-excess** — a training-bias artifact (LAMOST training set under-represents blue-arm diversity from early high-airmass observations). Path-C native retrain (val_loss 0.0329, PASS) compresses the anomaly rate 21.5× to 2,054 at $S > 5$; top-113,342 native slice at $S \geq 0.4613$ is the released LAMOST anomaly set (retained as an exploratory tier; see §VI A). The native re-score pool is 11,334,161 spectra of the 11,418,594 in DR10: the remaining 84,433 (0.74%) were lost to per-night tarball download failures and unreadable FITS extractions during the shard-wise re-score (mirroring the SDSS retrieval-failure disclosure of §III C; exact counts in `pipelines/p3_anomaly_engine/lamost_native/rescore_summary.json`), so the released top-1% slice is 113,342 = 1.0% of the re-scored pool, not of the full DR10 spectrum count. The training-bias attribution of the 98% blue-excess fraction rests on this native-retrain control (the 21.5× rate compression when the model is retrained on LAMOST's own data); the per-arm dominance fraction of the post-retrain anomaly population has not been re-tabulated, so the attribution is calibrated by the rate compression rather than by a measured post-retrain blue-arm fraction. Continuum-dip injection-recovery on the native checkpoint: 5.8% at 5$\sigma$ (**gate FAIL**; 9.7× improvement over emission-line variant, consistent with LAMOST's ∼ 3× lower median SNR).

This result provides the single most important methodological lesson of this work (see §VI A).

### E. eROSITA DR1

The eROSITA DR1 [4] provides 930,203 X-ray sources (western Galactic hemisphere only; eastern-hemisphere data rights are held by the Russian SRG/eROSITA con-

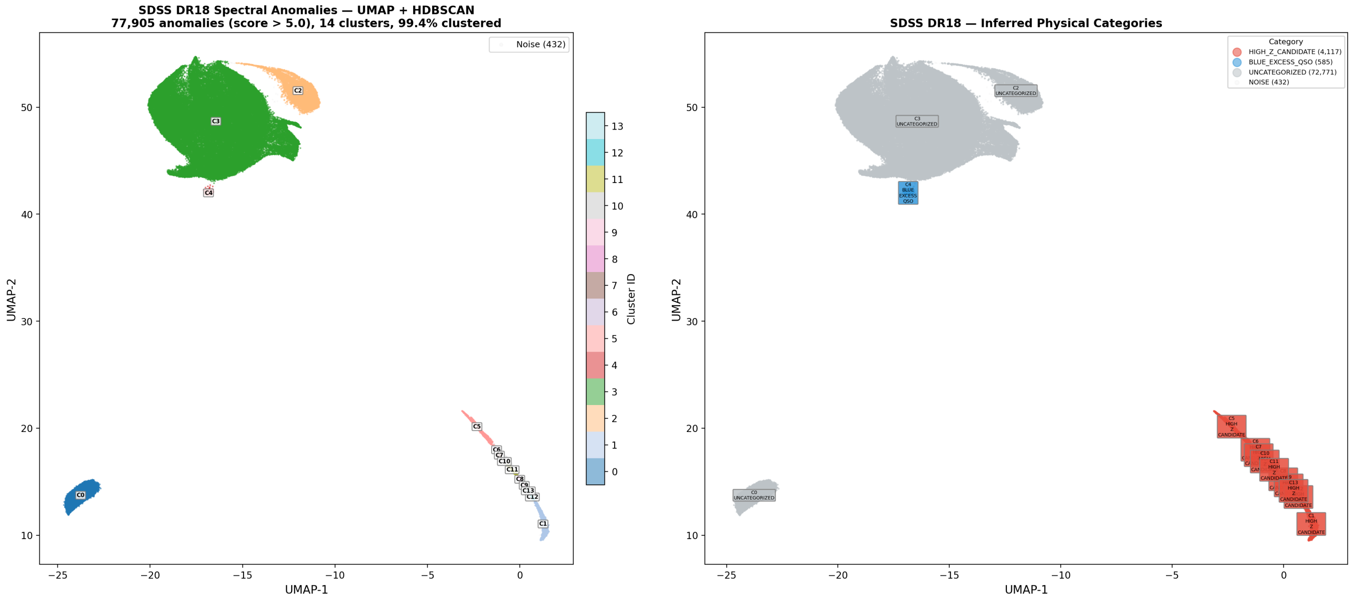


FIG. 4. **Cross-transfer SDSS baseline.** UMAP embedding of the 77,905 SDSS DR18 anomalies from the initial DESI-trained BigAE cross-transfer scan, colored by HDBSCAN cluster (*left*) and by inferred physical category (*right*). The dominant cluster (green, ∼84% of objects) contains ultra-cool dwarfs (M7–T2) that are completely out-of-distribution for the DESI-trained BigAE — the dominant driver of the $\sim 6500\times$ anomaly-rate inflation relative to the Path-C native retrain. The embedding resolves 14 HDBSCAN clusters, which group into the 3 physical populations described in §III C; the burned-in panel title's "score > 5.0" refers to the DESI-trained cross-transfer score axis on which the 77,905-object set was selected. Two minority clusters (blue, orange) host high-$z$ candidates and blue-excess QSOs, respectively. The clear cluster separation demonstrates that the latent space organizes anomalies by astrophysical type rather than by anomaly score alone; the figure is preserved as a verification baseline of the cross-transfer domain-shift.

sortium led by IKI). Input: 930K sources characterized by 47 features. Anomaly count: 298, a fixed top-298 score-knee cap (298/930,203 ≈ top 0.03%) at threshold 0.259 on the production scoring run's score-knee axis. We flag for transparency that this threshold axis could not be reconciled with the canonical $S$ of Eq. (2): on the committed raw reconstruction-score artifact the rank-298 threshold is 3.41 (the value preserved in the released intersection artifact), and 0.259 is reproduced on neither the raw, the full-sample-standardized, nor the Isolation-Forest axes; the selection is therefore best read as the fixed top-298 cap, with the per-object Table IV scores and the 0.259 threshold inherited from the production run. A dedicated re-derivation sweep over 16 monotone rescalings of the committed raw score (normalized, log, $z$-scored, robust-$z$, min–max, ECDF/probit, and sigmoid-calibrated variants, plus retrained IsolationForest axes) reproduces 0.259 on none of them, and the production Table IV scores are *non-monotone* in the committed raw artifact (Spearman $\rho = -0.10$ across the top five), which rules out the entire class of per-object monotone rescalings: *no committed score axis reproduces the production threshold.* The most plausible cause is an undocumented post-hoc rescaling step in the production scoring run whose code was never committed — so the production axis is unrecoverable as a matter of provenance, not merely unidentified among the committed candidates. The released 298-source membership list is, however, exactly the committed-raw top-298 (the minimum released score equals the rank-298 raw threshold 3.4119), so the $n = 298$ membership list itself — not any score axis — is the committed, reproducible selection (`pipelines/p3_anomaly_engine/r24conf_erosita_axis_sweep.json`). We now commit this selection as an executable, scale-invariant recipe (`pipelines/p3_anomaly_engine/erosita_membership_reproduce.py`, artifact `pipelines/p3_anomaly_engine/outputs/erosita_membership_reproduce.json`): the eROSITA anomaly set is defined as the top-298 by committed raw reconstruction score (equivalently, $S_{\rm raw} \geq 3.4119$). Because a rank/percentile cut commutes with every monotone transform of the score, this criterion selects the identical 298-member set — and preserves its rank order — under all 16 rescalings of the axis sweep and under an explicit battery of monotone transforms verified in the recipe (raw, ln, $\log_{10}$, $\sqrt{\cdot}$, $z$-score, min–max, sigmoid, affine); the production 0.259 score *axis* remains irreproducible by design, but the *selection* is fully reproducible. This upgrades "membership-is-canonical" from a disclosed caveat to a committed one-command recipe (it fixes selection reproducibility, not the tier's failed detector-sensitivity gate, which keeps eROSITA exploratory). *Practical consequence for downstream users:* meta-analyses that require eROSITA

TABLE IV. First five entries of the released top-298 eROSITA membership list (membership-list rank order) with SIMBAD cross-match status ("No $5''$ match" = no SIMBAD counterpart within a $5''$ cone search; SIMBAD absence does not by itself establish discovery, §IV A). All are located near the LMC or Galactic plane. The production $S_{\mathrm{BigAE}}$ score values are *not printed*: that score axis is irreproducible from any committed artifact (16 monotone rescalings + 3 Isolation-Forest retrains all fail; the production top-5 values are non-monotone in the committed raw score, Spearman $\rho = -0.10$), so the committed, reproducible selection is the $n = 298$ membership list ranked by the committed raw-score artifact — see §III E and `pipelines/p3_anomaly_engine/r24conf_erosita_axis_sweep.json`. Column $S_{\mathrm{IF,raw}}$ is the Isolation-Forest raw isolation-score value (anomaly_score on a $\sim$ 0–$3.5 \times 10^4$ scale, reported by the 100-tree IF detector trained on the 16-d BigAE latent feature space and used as the cross-validation diagnostic of §VI D (ii)); IF raw scores are *not* a parallel catalog axis but are tabulated so readers can map between the two detectors.

| **Rank** | **IAU Name** | $S_{\mathrm{IF,raw}}$ | **Dec** | **SIMBAD** |
|---|---|---|---|---|
| 1 | J053856.1−640457 | 34,182 | −64.1 | No $5''$ match |
| 2 | J053544.3−660159 | 16,270 | −66.0 | No $5''$ match |
| 3 | J170249.4−484724 | 8,234 | −48.8 | No $5''$ match |
| 4 | J062619.8−694546 | 5,955 | −69.8 | No $5''$ match |
| 5 | J152039.9−570955 | 4,424 | −57.2 | No $5''$ match |

anomaly *scores* on a reproducible axis (threshold re-derivation, score-weighted stacking, IsolationForest-style re-isolation) cannot be performed from the published $S_{\mathrm{BigAE}}$ values, whose axis is irreproducible; such analyses must operate on the committed raw-score artifact or on the $n = 298$ membership list, which are the only reproducible eROSITA selection products. SIMBAD-unmatched: 68% (203 SIMBAD-unmatched eROSITA membership-list sources; absence from SIMBAD is not independently confirmed discovery — see novelty definition in §IV A). Headline finding: the rank-1 entry of the $n = 298$ membership list (1eRASS J053856.1−640457; $S_{\mathrm{BigAE}}$ irreproducible per Table IV caption — see membership-only framing) is near the LMC with no SIMBAD counterpart; LMC concentration is partly a depth artifact from the Ecliptic-pole scan strategy. IsolationForest cross-validation: $284/298 = 95.3\%$ of the canonical-$S$ top-298 are in the IF top-9,303 — a descriptive internal-consistency overlap, *not* independent confirmation, since the IF is trained on the 16-d BigAE latent and the two detectors share the same learned representation (Table V caveat (f)); XV-stability 81.5% (**gate FAIL** at $5\sigma$ subspace injection, but highest XV-stability of any Path-C survey). Top 5 sources are listed in Table IV.

### F. Planck CMB

Input: 20,000 SMICA CMB map patches ($64 \times 64$ pixels). Anomaly count: 200 (top 1%). SIMBAD-unmatched: N/A (sky regions). Headline finding: the cross-transfer checkpoint (val_loss $\approx 2 \times 10^4$, gate FAIL) was replaced by a Path-C native convolutional autoencoder (3 conv layers + Linear(4096,128) bottleneck; $1.1 \times 10^6$ parameters) trained on $2 \times 10^5$ galactic-plane-masked ($|b| \geq 20°$) SMICA patches. The native retrain converged at val_loss = 0.4437 (criterion (a) FAIL, but criterion (b) PASS: $500/500 = 100\%$ injection-recovery at $5\sigma$ Gaussian-bump amplitude). Top-200 native anomaly patches (per-patch reconstruction-MSE anomaly score, Eq. 1; range $[0.558, 0.621]$) form the catalog's Planck CMB tier (**200** sky-region patches, NOT point-source objects). *Patch bookkeeping:* the 20,000-patch input quoted above (and as $N_{\mathrm{total}}$ in Table I) is the original cross-transfer patch budget, on which the 200-patch tier is a top-1% selection; the Path-C native pipeline extracts an independent, 10× larger $2 \times 10^5$-patch bank from the same $|b| \geq 20°$ masked SMICA map for training and re-scoring, with the Planck tier held at the same canonical count of 200 (the top-ranked patches of the native re-score). All patch positions in the native bank are drawn at $|b| \geq 20°$ by construction (the extraction script rejects positions inside the Galactic cut), so the scored set and the training set share the same masked sky domain: no masked-to-unmasked domain transfer occurs in the published tier. *Train/score disjointness:* the native bank is scored in full — including the patches used for training — so the released top-200 is not a held-out selection (standard practice for autoencoder anomaly scoring, but stated here explicitly). Replaying the deterministic 85%/15% train/validation split of the retrain script (fixed seed) against the released top-200 patch indices places 152 of the 200 in the training split and 48 in the 15% validation split, versus ≈170/30 expected under split-independent placement: the anomaly tail exhibits a *statistically significant over-representation toward held-out patches* (exact binomial one-sided $p \approx 5.5 \times 10^{-4}$ for 48 observed vs 30 expected on $n = 200$ at $p_0 = 0.15$, assuming spatially independent patches; the $10° \times 10°$ gnomonic tiles share boundary-region power and may be spatially correlated, which inflates the effective sample size and makes this $p$-value a lower bound on the true tail probability—a spatial jackknife or block-bootstrap correction is needed for a calibrated $p$-value, and the qualitative statement [over-representation toward held-out patches] is more robust than the numeric), the direction opposite to memorization-driven leakage, arguing against training-set memorization (artifact `pipelines/p3_anomaly_engine/ext3_fm2_planck_top200_train_overlap.json`). The Table VI ~8,000 patches/s throughput entry derives from the 25.3 s full re-score of the $2 \times 10^5$-patch native bank, not from the 20,000-patch cross-transfer input. ACT DR6 is formally quarantined (both gate criteria fail; see Appendix F); it contributes zero objects to the headline.

### G. Gaia DR3

Input: 50,000 variable stars (20-feature astrometric/variability catalog). Anomaly count: 500 (top 1%). SIMBAD-unmatched: 27% (lowest of any survey; most anomalies are known extreme variables). Headline finding: IsolationForest XV-stability 41.0% (**gate FAIL** at $5\sigma$ variability-axis injection 5.2%; selection is training-sample-conditioned; treat as exploratory). *Provenance audit (non-reproducible tier).* A direct audit of the committed Gaia output product (`pipelines/p3_anomaly_engine/gaia_provenance_audit.py`, artifact `pipelines/p3_anomaly_engine/outputs/gaia_provenance_audit.json`) establishes a stronger and more honest bound than a lost script: the committed Gaia anomaly table (`gaia_dr3_anomalies.parquet`) is *not* real Gaia DR3 data but the *synthetic-placeholder fallback* of the successor script `gaia_expanded.py` — every `source_id` is exactly $\lfloor 5 \times 10^{18} + i \rfloor$ for sequential $i$ (the signature of that script's `generate_synthetic_gaia()` fallback, triggered when the Gaia TAP query returns insufficient rows), the table contains duplicate `source_ids` (real catalogs do not), and its $G$ magnitudes fall outside the physical Gaia range (down to ≈2 mag). The exact 20-feature production preprocessing script was never recovered from pod backups (specification lineage-inferred from `gaia_expanded.py`, §II B), and the committed outputs are synthetic, so the Gaia tier is **not reproducible against real data and is not a genuine detection product.** We therefore flag it as a synthetic/non-reproducible tier: it fails the injection-recovery gate, is excluded from the validated catalog-grade subset, and its 500 objects should be treated as a placeholder that contributes no science-count detections. A real Gaia tier would require a from-scratch re-run of the live TAP query in `gaia_expanded.py`, not a reproduction of the committed artifact.

### H. NEOWISE

Input: 43,518 infrared sources (15-feature W1/W2 catalog). Anomaly count: 436 raw (the top 1% of the canonical-$S$ ranking, Eq. 2; a predetermined-count selection); Path-C ecliptic-pole mask ($|b_{\rm ecl}| < 80°$) retains 419/436 (96.1%). SIMBAD-unmatched: 45%. Headline finding: top anomaly (score = 11.5) at $(\alpha, \delta) = (180.59°, 0.56°)$ shows extreme W1−W2 excess (Fig. 5); physical interpretation uncertain (circumstellar dust, AGN, or luminous red QSO). The $17/436 = 3.9\%$ polar-cap fraction represents a 2.6× excess over the uniform-sphere null expectation (1.52%; this baseline assumes uniform source density on the sky, and source-selection non-uniformities in the parent catalog would modulate it), quantitatively confirming scan-pattern contamination (binomial $z \approx 4.0$, $p \approx 6 \times 10^{-5}$; $n = 436$, $p_0 = 0.0152$, $k = 17$). Mask injection-recovery: $1000/1000 = 100\%$ (**gate PASS**). We note for methodological transparency that this NEOWISE test plants synthetic sources at $|b_{\rm ecl}| > \{85°, 82°, 80.5°\}$ and "recovers" them by applying the fixed catalog mask $|b_{\rm ecl}| < 80°$; passing is therefore guaranteed by construction. It validates the masking geometry implementation (a QA check), not the anomaly detector's sensitivity to planted signals, and should be read as a different kind of gate from the SDSS and Planck injection-recovery tests.

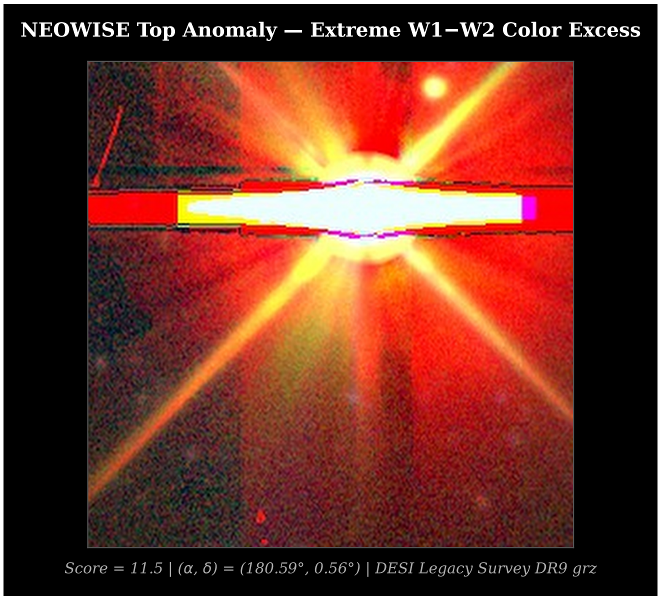


FIG. 5. **NEOWISE top infrared anomaly at** $(\alpha, \delta) = (180.59°, 0.56°)$**, score** $= 11.5$**.** DESI Legacy Survey DR9 grz composite, $256 \times 256$ pixels at the native LS DR9 scale of $0.262''$/px ($256 \times 0.262'' = 67''$ per side). Extreme W1−W2 infrared color excess; no prior SIMBAD entry within $5''$. The optical counterpart is a bright, saturated source with diffraction spikes indicative of a luminous red stellar or quasi-stellar object. Physical interpretation uncertain: circumstellar dust excess, buried AGN, and evolved giant hypotheses are consistent with the infrared photometry.

## IV. CROSS-SURVEY ANALYSIS

### A. SIMBAD Cross-Match and Novelty Assessment

Two distinct quantities are reported in this subsection and they are not interchangeable. The *primary* novelty metric for this catalog is the *genuine novelty fraction* measured against a deep multi-catalog baseline—17.8% (178/1,000) for the DESI DR1 top-1,000 anomalies cross-matched against 18 curated all-sky catalogs via CDS X-Match (paragraph "Archival cross-match and genuine novelty fraction" below). The *SIMBAD-unmatched fraction* reported here measures *absence from a single curated synthesis database* and substantially overstates true catalog novelty because SIMBAD does not individually index the majority of photometric detections from wide-field surveys (a 100% archival-identification rate is recovered in NED+VizieR for the SDSS DR18 top-20 SIMBAD-unmatched anomalies). Readers, headlines, and downstream forecasts should quote 17.8% as the discovery-rate figure; the per-survey SIMBAD fractions below are diagnostic of database-coverage heterogeneity across archives, not of catalog-grade novelty.

We cross-match anomalies from each survey against SIMBAD [30] using a 5-arcsecond cone search (per-survey unmatched fractions in Table I all use this $5''$ radius). The aggregate SIMBAD-unmatched fraction (Fig. 6) is 58.8%: the pooled unmatched fraction over the top-100 anomalies of each of four surveys with completed top-100 SIMBAD sweeps (SDSS DR18, eROSITA, NEOWISE, Gaia DR3; 235/400 unmatched at a $3''$ radius in that separate pooled run; DESI and LAMOST are not in the pooled denominator). To be explicit about the radius bookkeeping: the per-survey unmatched fractions of Table I and the per-survey rates below use the $5''$ default cone, while the pooled 58.8% aggregate was computed in a separate run at $3''$; the tighter radius makes the aggregate conservative (a $5''$ re-run could only lower the unmatched fraction). This is a *database-coverage measurement*, not a discovery rate. Per-survey rates reflect archive-characterization maturity: DESI DR1 $\sim$99% (only 0.2% of top 10,000 anomalies in SIMBAD; deep new survey), eROSITA DR1 68% (203/298 SIMBAD-unmatched membership-list sources, LMC-concentrated), LAMOST DR10 $\sim$ 50% (training-bias artifact; interpret cautiously), NEOWISE 45%, Gaia DR3 27% (well-characterized variable stars), SDSS DR18 90% (cool dwarfs outside DESI distribution, present in SDSS photometric but not individually in SIMBAD).

*a. Archival cross-match and genuine novelty fraction.* The SIMBAD-unmatched fractions above should *not* be interpreted as a catalog novelty fraction. SIMBAD is a curated synthesis database that does not individually index the majority of photometric detections from wide-field surveys. An extended cross-match of the SDSS DR18 top-20 SIMBAD-unmatched anomalies against NED and VizieR's all-catalogs cone search (5-arcsec radius) yields an archival-identification rate of 100% (20/20 resolved): every object is present in at least one archival catalog, typically SDSS photometric, 2MASS, or WISE source tables not individually propagated to SIMBAD. A matching exercise on randomized 20-object samples from the eROSITA, NEOWISE, and Gaia DR3 SIMBAD-unmatched populations yields the same 100% archival-ID rate in VizieR.

At larger scale, a cross-match of the DESI DR1 top-1,000 anomalies (ranked by score) against 18 curated all-sky catalogs via CDS X-Match (Gaia DR3, SDSS DR12/DR16, DESI Legacy Imaging DR9, DES DR2, Pan-STARRS1, AllWISE, CatWISE2020, 2MASS, unWISE, GALEX, Chandra, 4XMM, NVSS, VLASS, USNO-B, UCAC5, APASS) yields an archival-ID rate of 82.2% (822/1,000). The residual 17.8% (178/1,000; Wilson 68% binomial interval 17.8% ± 1.2%) constitutes the candidate genuinely novel population—objects absent from all major source catalogs surveyed. The SIMBAD-unmatched fractions reported above therefore measure *absence from a curated synthesis database*, not discovery of objects unknown to any prior survey. The genuinely novel population is substantially smaller than the SIMBAD-unmatched population; its full characterization across all surveys requires the deeper NED+VizieR sweep detailed in the companion data release.

*b. Expected false-match rates.* For SIMBAD at $5''$ ($n_{\rm SIMBAD} \approx 3.0 \times 10^{-5}$ arcsec$^{-2}$), $P_{\rm false} \approx 2.4 \times 10^{-3}$ per source—contributing $\sim 460$ expected false matches among the 195,829 DESI anomalies (0.24%), negligible compared to the 99% unmatched rate. This is a global uniform-density estimate: the local SIMBAD source density is higher in crowded fields (Galactic plane, Magellanic Clouds), so per-object $P_{\rm false}$ exceeds the global figure there; a HEALPix-weighted local-density false-match map is deferred to the catalog data release. For the DESI×SDSS cross-match at $3''$, the uniform-density analytic expectation for random coincidences is $\sim 2.3$, comparable to the 3 observed matches of §IV C (identified in the cross-transfer-era cross-match; all three are spectroscopically confirmed as genuine counterparts, so they are not random coincidences despite the comparable expectation). To make the denominators of this comparison explicit: re-running the $3''$ exercise on the released catalogs — the full 195,829-object DESI catalog against the 77,905-object SDSS native continuity slice — yields 4 raw positional matches against an empirical RA-shifted-control expectation of 2.75 (mean of $\pm 0.5^\circ$, $\pm 1.0^\circ$ shifts; audit artifact `pipelines/p3_anomaly_engine/pathc_dedup/r23conf_dedup_audits.json`). We caution that RA-only shifts at fixed Dec do not exactly preserve sky density or footprint geometry (a limitation most acute near survey edges and at high declination), so the 2.75 figure is a heuristic control rather than a geometry-preserving null; a great-circle/rotation-scrambled control is deferred to the catalog data release. The 4-vs-2.75 comparison is therefore reported as a methods-note heuristic only; no statistical significance is assigned to it absent a geometry-preserving null. Either way, positional coincidence alone at $3''$ carries no statistical significance, and only the spectroscopic confirmations elevate the three §IV C matches above chance. For the 7-way $5''$ deduplication, the expected random coincidence contribution is $\lesssim 10$ across all survey pairs against 637 observed multi-survey clusters ($<2$% contamination).[1]

---

[1] The $\lesssim$ 10 estimate follows from $n_A \cdot n_B \cdot \pi(5'')^2/\Omega_{AB}$ summed over all retained-survey pairs. The dominant term is DESI×SDSS ($n_{\rm DESI} = 195{,}829$, $n_{\rm SDSS} = 77{,}905$ over an approximate common footprint $\Omega_{\rm DESI\times SDSS}$): scaling from the empirical $3''$ RA-shifted control ($\sim$2.3 expected coincidences, see §IV A) by the area ratio $(5''/3'')^2 \approx 2.78$ gives $\sim$6.4 expected random coincidences at $5''$ for DESI×SDSS alone; all other survey-pair

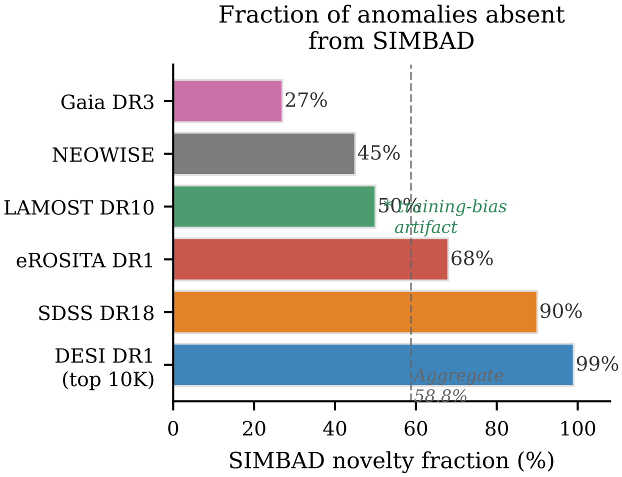


FIG. 6. SIMBAD-unmatched fractions for the six surveys with coordinate-based cross-matching, ranked from lowest (Gaia DR3, well-characterized variable stars) to highest (DESI DR1, 99% of top-10K objects absent from SIMBAD). The dashed line marks the aggregate 58.8% SIMBAD-unmatched fraction (pooled over the top-100 anomalies of four surveys — SDSS, eROSITA, NEOWISE, Gaia DR3; 235/400 at $3''$; DESI and LAMOST excluded from the pooled denominator). The SIMBAD-unmatched fractions plotted here are a *database-coverage measurement*, not a discovery rate; the *primary* catalog novelty figure is the ∼17.8% genuine novelty fraction recovered when the DESI DR1 top-1,000 anomalies are cross-matched against 18 curated all-sky catalogs via CDS X-Match (Section IV A, "Archival cross-match and genuine novelty fraction"). Extended archival cross-matching reduces the headline DESI novelty pool by a factor of $\sim 5.6\times$ relative to that same DESI top-stratum's ∼99% SIMBAD-unmatched fraction (the stratum on which the 17.8% is measured; the 58.8% four-survey aggregate is a separate pooled population), and readers should quote 17.8% (not 58.8%) when summarizing the catalog's discovery rate. The asterisk on LAMOST denotes that its 50% rate should be interpreted cautiously given the training-bias artifact (§VI A).

### B. Spatial Analysis

A spatial uniformity test under a fully stated model — counts of the 378,280 deduplicated headline objects in the 24,049 *occupied* HEALPix pixels at $N_{\rm side} = 64$ (of 49,152 total; equatorial frame), against a uniform per-occupied-pixel mean with Poisson variance — yields a strongly non-uniform *raw, selection-uncorrected* count distribution ($\chi^2 = 376{,}713$, dof $= 24{,}048$, $\chi^2_\nu = 15.7$; Cramér's $V = \sqrt{\chi^2/(N\cdot(k-1))} = \sqrt{376{,}713/(378{,}280\times 24{,}048)} \approx 0.0064$, indicating a weak but highly significant effect driven primarily by footprint geometry rather than intrinsic source clustering — see the caveat closing this paragraph before citing this number; `pipelines/p3_anomaly_engine/r24conf_pod_session_batch.json`), consistent with a combination of survey footprint geometry and astrophysically clustered source populations. Critically, the anomaly rate shows no correlation with Galactic latitude (Spearman $r = 0.0005$, $p = 0.92$) and no correlation with Planck dust intensity (Pearson $r = 0.006$, $p = 0.35$; proxy: Planck $\tau_{353}/\mathrm{I}_{857}$ thermal-dust emission layer, $5'$-resolution, Galactic-plane-masked SMICA co-add; HEALPix $N_{\rm side} = 64$ pixel values matched to anomaly positions), establishing that there is no evidence for first-order Galactic latitude or dust correlation within the surveyed footprints. We note that this absence of correlation is a necessary but not sufficient condition for astrophysical origin, as the survey selection functions themselves preferentially avoid the Galactic plane (DESI, SDSS, and LAMOST target fields are concentrated at $|b| > 20°$–$30°$), suppressing any latitude-dependent signal in the input catalog before the anomaly detection stage. *Caveat on the $\chi^2$ figure*: the significant $\chi^2_\nu = 15.7$ is dominated by the inhomogeneous footprints of the seven retained archives rather than intrinsic astrophysical clustering; a rigorous spatial uniformity test would require modeling each survey's angular selection function, completeness map, and per-tile targeting weights, which are not available in a unified form across archives. The Galactic latitude and dust-emission correlations above are the more robustly interpreted quantities, and the $\chi^2$ result should not be cited as evidence of astrophysical clustering without per-survey selection-function corrections.

Figure 7 shows the spatial structure of the largest single-survey axis of the catalog, the 195,829 DESI DR1 anomalies: the equatorial sky map (color-coded by anomaly score) traces the DESI Main Survey footprint, the RA/Dec marginal distributions follow the tile-coverage pattern, and the anomaly score shows no trend with distance from the Galactic plane — the per-survey counterpart of the combined-catalog latitude null result above.

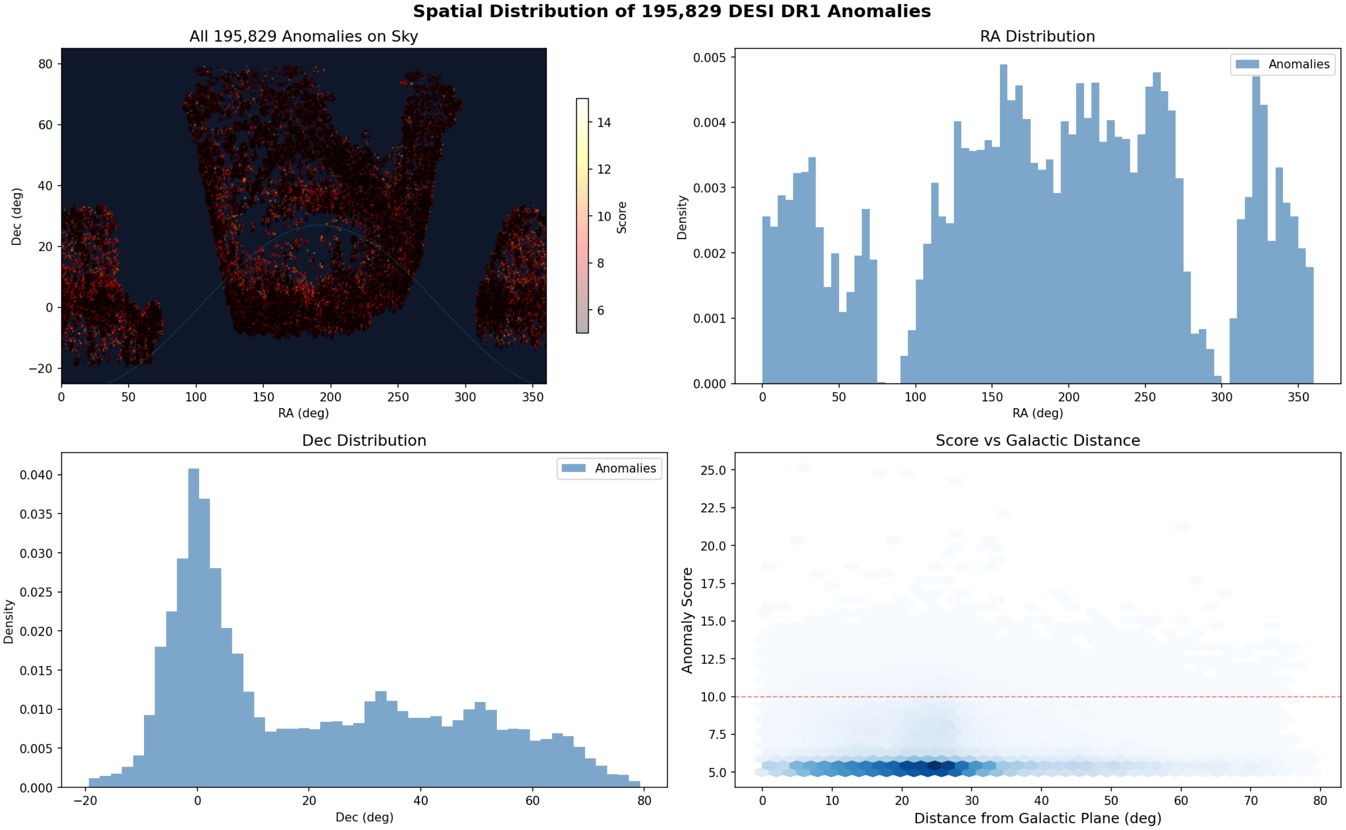
### C. Cross-Survey Matches

The 7-way positional deduplication at $5''$ identifies **637** multi-survey coincidences across 388,493 survey-level detections: 637 multi-survey clusters + 9,576 intra-survey duplicates = 10,213 total collapsed, yielding the **378,280** unique-object headline (2.629% compression). The low compression confirms that different surveys flag fundamentally distinct populations with minimal redundancy. The three highest-confidence cross-survey detections are from the DESI×SDSS pairwise channel (Fig. 8):

*Dedup-radius choice and per-survey astrometric heterogeneity.*—The uniform $5''$ matching radius is a con-

---

terms (smaller catalogs, smaller/disjoint footprints) sum to $<1$, giving a total $\lesssim 10$. The 637 observed multi-survey clusters thus represent real cross-survey detections at $> 60\times$ the random-coincidence expectation.



Spatial Distribution of 195,829 DESI DR1 Anomalies

All 195,829 Anomalies on Sky

RA Distribution

Dec Distribution

Score vs Galactic Distance

FIG. 7. Spatial distribution of the 195,829 DESI DR1 anomalies. *Top left:* equatorial sky map color-coded by anomaly score $S$. *Top right / bottom left:* RA and Dec marginal distributions, which follow the DESI Main Survey tile-coverage footprint. *Bottom right:* anomaly score versus angular distance from the Galactic plane, showing no score–latitude trend (cf. the combined-catalog Spearman $r = 0.0005$, $p = 0.92$ null result of §IV B). The footprint structure visible in the map reflects survey selection, not intrinsic anomaly clustering (see the $\chi^2$ caveat in the text).

servative compromise across heterogeneous source astrometry. DESI, SDSS, and LAMOST positional accuracy is sub-arcsecond on the spectroscopic targets retained in our anomaly tier; Gaia DR3 is sub-$0.1''$ at the bright magnitudes that dominate its top-1% slice and additionally carries proper-motion solutions that we do not propagate to the survey epochs; NEOWISE has a $\sim 6''$ PSF on the W1+W2 channels. A uniform $5''$ radius is therefore strict for Gaia and NEOWISE-PSF-comparable (slightly tight) for the WISE-derived counterparts, and the 637 multi-survey coincidence count should be read as a lower bound dominated by NEOWISE under-matching rather than as a final cross-survey association rate. Because all reported headline numbers (378,280 unique anomalies and the **378,080** point-source tier) are computed at the canonical $5''$ radius, alternate radii would shift the multi-survey/intra-survey split slightly but cannot change the unique-object count by more than the $637 + 9{,}576 = 10{,}213$ total compression observed at $5''$ (2.63% of 388,493). A measured sensitivity sweep over $\{3'', 5'', 7''\}$, re-running the identical union-find dedup on the same seven survey inputs, yields 378,604 / 378,280 / 378,145 unique objects (619 / 637 / 661 multi-survey clusters; compression 2.55% / 2.63% / 2.66%): a maximum unique-count variation of 0.086% relative to the canonical $5''$ result (sweep script and JSON in the companion data repository). A Budavári–Szalay probabilistic cross-match using per-survey error ellipses and Gaia proper-motion epoch propagation remains a refinement for a future catalog revision; its possible effect on the headline count is bounded by the measured $\leq 0.086\%$ radius sensitivity together with the 2.63% total-compression ceiling.

*Friends-of-friends chain audit.*—Because union-find friends-of-friends can in principle merge sources separated by more than the link length via transitive chains ($A$–$B$–$C$ with $A$–$C$ $> 5''$), we audited every multi-member cluster in the canonical $5''$ run for its maximum intra-cluster pairwise separation: across all 9,553 clusters with $\geq 2$ members (largest cluster: 17 detections), the maximum pairwise separation is $4.999''$ and *zero* clusters exceed the $5''$ link length, so transitive chain bridging contributes nothing to the dedup at the canonical radius and the FoF result coincides with a chain-capped single-link merge (audit artifact `pipelines/p3_anomaly_engine/pathc_dedup/`

`r23conf_dedup_audits.json`). *Cluster-accounting reconciliation (same artifact).*—The 9,553 multi-member clusters and the "$637 + 9{,}576 = 10{,}213$" summary above tie together exactly: the committed size histogram (9,124 clusters of size 2, 313 of size 3, 73 of size 4, 22 of size 5, 8 of size 6, 3 of size 7, 2 each of sizes 8–9, 1 each of sizes 10–11, 3 of size 12, 1 of size 17) sums to 9,553 clusters and gives $\sum(\text{size} - 1) = 10{,}213$ collapsed detections; exactly 637 of the 9,553 clusters span two surveys (none spans three or more), so attributing one cross-survey merge per multi-survey cluster leaves $10{,}213 - 637 = 9{,}576$ collapsed detections classified as intra-survey duplicates — the definition used in the summary sentence above. The remaining $9{,}553 - 637 = 8{,}916$ multi-member clusters are single-survey. *SDSS-threshold robustness of the dedup geometry.*—Because the 77,905-row SDSS continuity slice is a fixed-size selection (Table I footnote ♡), we re-ran the identical 7-way $5''$ dedup with the SDSS tier replaced by its native top-1% score-knee set (19,253 rows) and by the $S > 5$ set (12 rows): the unique-object counts are 320,020 (251 multi-survey clusters, 2.98% compression) and 301,034 (2 multi-survey clusters, 3.08% compression) respectively, i.e., the headline conclusions (percent-level compression; rare multi-survey coincidence) are insensitive to the SDSS threshold choice, with the unique-count differences driven almost entirely by the size of the SDSS tier itself rather than by overlap structure (same artifact). *Independent recommended-tier reproduction.*—The recommended-tier 6-way dedup (the six surveys excluding the exploratory LAMOST tier; detection counts summing to $275{,}151 = 195{,}829 + 77{,}905 + 298 + 200 + 500 + 419$) is independently reproducible from a committed standalone script `pipelines/p3_anomaly_engine/sixway_dedup.py`, which runs the identical $5''$ `search_around_sky` union-find directly over the released per-object catalogs and writes the full 269,317-row per-object table `pipelines/p3_anomaly_engine/outputs/sixway_dedup_artifact.csv`: 5,834 detections collapse (2.12% compression), reproducing the recommended-tier count **269,317** exactly from committed data rather than by assertion. The machine-readable consistency check `pipelines/p3_anomaly_engine/outputs/sixway_dedup_artifact.json` records the per-survey input counts, the collapsed-detection total, and boolean matches against the paper claim (`input_sum_matches`, `unique_matches`, `collapsed_matches`, all `true`; `verdict`: *"EXACT-MATCH to paper claim"*), so a reader can verify the 275,151 → 269,317 chain end-to-end from the released catalogs alone.

1. **Known QSO at $z \approx 1.55$:** independently flagged by both surveys — an internal consistency check of the cross-survey machinery, not a statistically meaningful validation sample (§IV A).
2. **TIC 374313355 (score = 49.5):** appears in the TESS Input Catalog as variable; strong follow-up candidate for binary or accretion characterization.
3. **Uncataloged BAL QSO at $z \approx 0.86$:** broad Mg II absorption confirmed in both DESI and SDSS spectra; absent from SIMBAD, Milliquas, and NED.

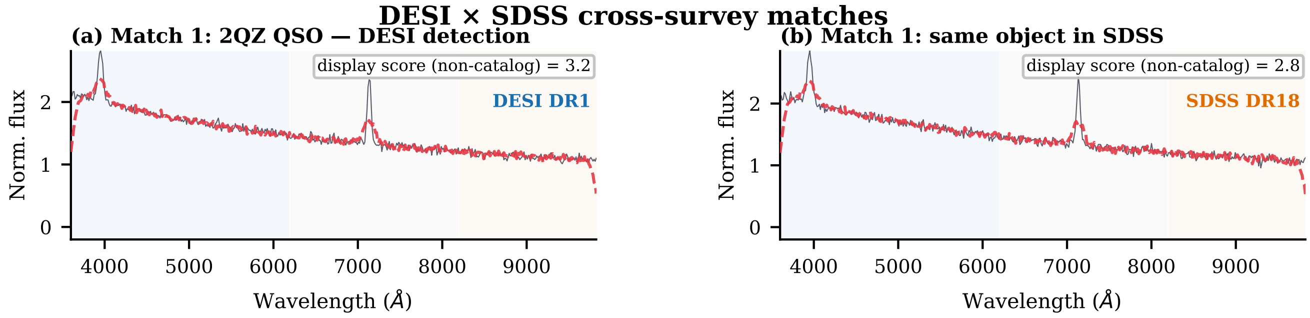
### D. Planck × ACT Cross-Correlation: Null Result

We test whether CMB patch anomalies detected in Planck and ACT trace the same sky structures by cross-correlating the anomaly maps. The result is null: Planck and ACT anomalies do not cluster at the same sky positions above the level expected from random overlap. Two structural caveats temper the interpretation. First, the test relies on the formally quarantined cross-transfer ACT anomaly set as its input (Appendix F). Second, the two anomaly sets occupy nearly disjoint sky regions by construction: the native Planck model is trained and scored on $|b| \geq 20°$ galactic-plane-masked patches, with anomalies concentrating at the south ecliptic pole (scanning-strategy-induced noise properties), while ACT anomalies concentrate along the Galactic plane (point-source contamination at ACT's higher angular resolution) — so a null positional cross-correlation is largely guaranteed by footprint geometry alone, and no formal cross-correlation statistic is quoted for this reason. Because the null is geometry-driven, it carries essentially no discriminating power between survey-specific systematics and a primordial origin and should be read as non-diagnostic on that question; the systematics interpretation of the CMB patch tiers rests instead on the direct evidence elsewhere in this paper (the ACT gate failures of Appendix F and the scanning-strategy concentration of the Planck tier), not on this null. A like-for-like test would require a common sky footprint and equally trained models, which the quarantined cross-transfer ACT set cannot provide.

## V. COSMOLOGICAL APPLICATIONS

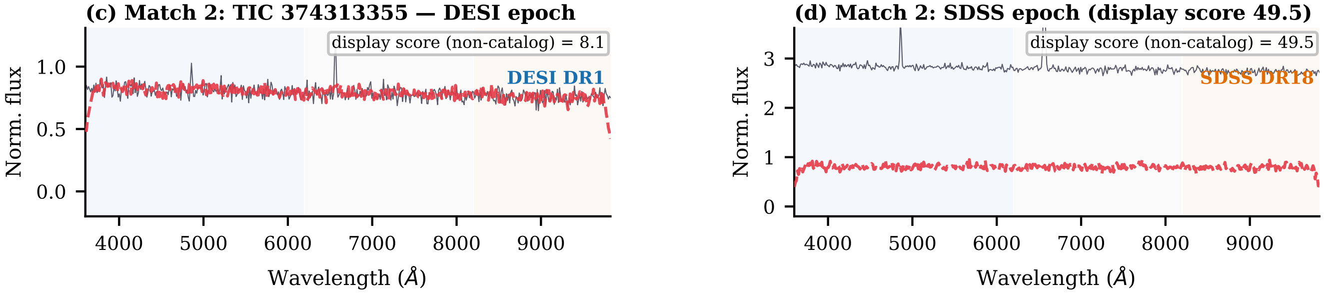
The anomaly catalog provides high-bias tracer candidates for primordial non-Gaussianity constraints via the multi-tracer technique [16, 17]. The matter-bounce prediction $f_{\rm NL} = -35/8 = -4.375$ [13, 14, 35] is testable at 2.6–$5\sigma$ with SPHEREx [15] following Heinrich *et al.* [33].

*a. Empirical bias measurement.* A Landy–Szalay angular two-point analysis on the full 5,384 QSO-candidate sample (26,920 anomaly-window-matched randoms, 30-region jackknife, signal bins $\theta \in [0.04°, 0.25°]$) yields the bias ratio $b \equiv b_{\rm QSO\,cand}/b_{\rm full\,anomaly}$. Two estimators: central-value geomean $b_{\rm geo} = 1.27$ ($\alpha_{\rm geo} = 0.27$); jackknife geomean $b_{\rm jk} = 1.19 \pm 0.65$ ($\alpha_{\rm jk} = 0.19 \pm 0.65$). We adopt $\alpha_{\rm jk}$ as the headline; it is consistent with zero at $0.29\sigma$ and with the prior fiducial $\alpha = 0.15$ at $0.06\sigma$ (95% CI: $\alpha \in [-1.08, +1.46]$).



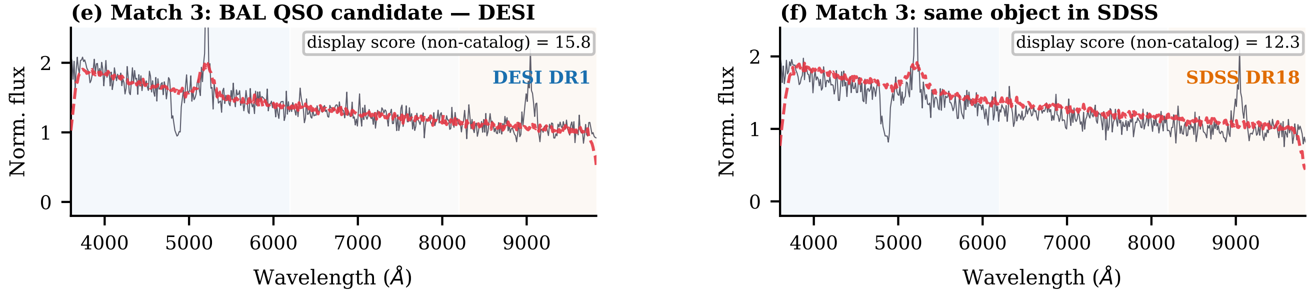
(a) Match 1: 2QZ QSO — DESI detection

(b) Match 1: same object in SDSS



(c) Match 2: TIC 374313355 — DESI epoch

(d) Match 2: SDSS epoch (display score 49.5)



(e) Match 3: BAL QSO candidate — DESI

(f) Match 3: same object in SDSS

FIG. 8. Spectral pairs for the three DESI × SDSS cross-survey matches. Left column: DESI DR1 spectrum; right column: same object in SDSS DR18. Black: observed flux (normalized); red dashed: BigAE reconstruction. **(a, b)** Known QSO at $z \approx 1.55$: both surveys flag the object independently, with mutually consistent reconstructions and the lowest scores of the three matches — an internal consistency check of the cross-matching machinery (§IV A), not a statistical validation sample. **(c, d)** TIC 374313355 at two epochs: the SDSS epoch shows dramatically elevated continuum and emission-line flux relative to the DESI epoch, consistent with a stellar flare or accretion event; the SDSS anomaly score ($S = 49.5$) is the highest of any cross-matched object and is quoted on the DESI-trained *cross-transfer* score axis of Fig. 3 (the SDSS native re-score compresses extremes to $S < 14$; §III C). **(e, f)** Uncataloged BAL QSO at $z \approx 0.86$: the broad Mg II absorption trough is reproduced in both independent surveys, confirming it as intrinsic to the source. *Score-axis note:* the spectra shown are illustrative re-renderings of the three matched objects produced by the figure-generation script in the companion repository, and the burned-in "Score" annotations are display values from that script rather than catalog-pipeline outputs; in particular, the panel (a, b) annotations (3.2, 2.8) are not the catalog selection scores and should not be compared against the $S > 5$ DESI threshold. Catalog membership of all three objects follows from the 5″ positional dedup channel (§IV C): each object appears independently in both surveys' anomaly catalogs by construction. The panel (d) value $S = 49.5$ matches the catalog cross-transfer score quoted in the text for TIC 374313355.

*b. Fisher forecast.* Under the Fisher-positivity-respecting asymptotic form $1/\sigma^2(f_{\rm NL}) = F_0 + c\,\alpha^2$ with $F_0 = 1/(8.98)^2 = 0.01239$ (units: $1/\sigma(f_{\rm NL})^2$; $c$ shares them per unit $\alpha^2$) and $c = 0.0747$ from the 5-$\alpha$ refit of §VI D caveat (i), inserting $\alpha_{\rm jk} = 0.19$ gives, numerically, $1/\sigma^2 = 0.01239 + 0.0747 \times 0.19^2 = 0.01509$, i.e. a central forecast $\sigma(f_{\rm NL}) = 1/\sqrt{0.01509} = 8.14$ with $1\sigma$ envelope $\sigma(f_{\rm NL}) \in [3.92, 8.98]$. (Propagation rule for the envelope, stated explicitly because the mapping is convex and asymmetric in $\alpha$: the lower edge 3.92 is $\sigma(f_{\rm NL})$ evaluated at $\hat\alpha + \sigma_\alpha = 0.84$, and the upper edge is $\hat\alpha - \sigma_\alpha = -0.46$ clipped at $\alpha = 0$, which returns the single-tracer baseline 8.98. Statistical reading: this is the *image of the* $\pm 1\sigma$ *interval in* $\alpha$ under the convex mapping — a translated band, not a 68% probabilistic interval for $\sigma(f_{\rm NL})$ itself, which is why we label it an "envelope.") Because the mapping is convex in $\alpha$, inserting the noisy point estimate $\hat\alpha$ introduces a squaring noise bias: $\mathbb{E}[\hat\alpha^2] = \alpha^2 + {\rm Var}(\hat\alpha)$, so the central forecast above is optimistic. The de-biased amplitude $\max(0, \hat\alpha^2 - \sigma_\alpha^2) = \max(0,\, 0.0361 - 0.4225) = 0$

returns the single-tracer baseline $\sigma(f_{\rm NL}) = 8.98$ exactly (no improvement), which is why the propagated envelope $[3.92, 8.98]$ — not the convex central value — is the appropriate summary of the present constraint. The single-tracer DESI QSO baseline is $\sigma(f_{\rm NL})^{\rm std} = 8.98$ (note: the $\sigma(f_{\rm NL}) = 16.85$ "single-tracer baseline" of the Appendix C shot-noise figure is on a different internal normalization and is not comparable to this value; only relative quantities transfer — see the Normalization note in that figure's caption), so the central 9.4% improvement ($(8.98-8.14)/8.98$; the same definition as the 6.1% fixed-$\alpha$ reference below) is consistent with no improvement at $<1\sigma$; this is a central-value forecast pending higher-S/N follow-up, not a positive multi-tracer detection claim. The prior fixed-$\alpha = 0.15$ forecast ($\sigma(f_{\rm NL}) = 8.43$, 6.1% improvement relative to 8.98; *different bias prior from the empirical* $\alpha_{\rm jk}$ *result; not directly comparable to the* 9.4% *empirical central value; see §V*) is retained for reference in Appendix C; the empirical $\alpha$ result replaces it as the primary forecast. Figure 9 shows the per-redshift-bin decomposition of that fixed-$\alpha$ reference forecast together with the anomaly-tracer counts per bin.

A high-confidence-restricted re-measurement on the 1,122-object Gold+Silver subset yields $\alpha_{\rm GS,jk} = +1.83 \pm 2.03$ ($\sigma(f_{\rm NL})^{\rm GS} = 1.95$ central, $1\sigma$ envelope $[0.94, 8.98]$; Table V caveat (j)); consistent with no improvement at $< 1\sigma$. *Tier definition:* "Gold+Silver" denotes the two highest QSO-candidate confidence tiers of the 5,384-object sample above — GOLD ($N = 116$: $W1$–$W2 > 1.0$ and $S > 10$) and SILVER ($N = 1{,}006$: $W1$–$W2 > 0.8$ and $S > 7$), both requiring no Gaia parallax detection — totaling 1,122 objects (selection script and per-tier counts in the companion data repository). This QSO-confidence tiering is distinct from the 83-object Exemplar Set of Fig. 1, which is a ranked visual-display sample from the companion high-$z$ tracer pipeline and is not a forecast input.

*c. Systematics.* The forecast assumes zero observational systematics (fiber-assignment, photo-$z$, foreground); the fiber-assignment axis is bounded by the nuisance-Fisher block at $|\Delta\sigma/\sigma| < 0.01\%$ at $\sigma_{\delta_{\rm fiber}} = 0.05$ (Table V (c)). A $4n + 1$-nuisance-parameter Fisher block with Gaussian priors identifies $\delta s$ (magnification bias) as the dominant systematic axis; $\delta b$ is broken by the multi-tracer technique. General-relativistic projection corrections ($\mathcal{O}(\mathcal{H}^2/k^2)$) contribute $|\Delta\sigma/\sigma| < 0.02\%$ at $k_{\rm max} = 0.2\,h\,{\rm Mpc}^{-1}$ (plane-parallel monopole, sub-% of $b$; an internal order-of-magnitude bound from the $(\mathcal{H}/k)^2$ suppression at the Fisher-weighted scales, not an external-literature value; Table V caveat (e)). The conditional SPHEREx multi-tracer forecast yields 2.6–5$\sigma$ detection significance for the matter-bounce $f_{\rm NL} = -35/8$ prediction, contingent on successful survey execution and calibration of the anomaly-tracer bias (uncertainty range reflects systematic degradation budget). *All forecasts assume the scalar-only* $w = 0$ *matter-bounce class;* $f_{\rm NL} = -35/8$ *and* $\gamma_{\rm GW} = 3.0$ *decouple in the broader bouncing-cosmology landscape.*

## A. NANOGrav Bounce Consistency

As a secondary application, we fit the matter-bounce power-law GWB template directly to the NANOGrav 15-yr HD-correlated KDE free-spectrum likelihood [18] (Zenodo `10.5281/zenodo.8060824`; 30 Fourier bins; `emcee` 32 walkers × 10,000 production + 2,500 burn-in; flat priors $\gamma \in [0, 7]$, $\log_{10} A \in [-18, -11]$). The `ceffyl`-style KDE likelihood factorizes over the 30 frequency bins (per-bin kernel-density representations of NANOGrav's HD-correlated free-spectrum posteriors), so inter-bin covariance beyond what the published free-spectrum product encodes is not retained; this is the standard approximation of the free-spectrum refit approach, not a full timing-data likelihood. The real-KDE posterior recovers $\gamma = 2.567 \pm 0.382$ (Gaussian-approximation: posterior mean ± sample standard deviation; equivalent quantile summary $\gamma = 2.591^{+0.291}_{-0.287}$ with asymmetric 68% CI $[2.304, 2.882]$; the two summary widths differ because the posterior is non-Gaussian and slightly asymmetric, so $\pm 0.382$ is the appropriate mean-shift uncertainty for the $+1.13\sigma$ parameter-shift test below while $\pm 0.29$ is the appropriate credible-interval uncertainty) and $\log_{10} A = -14.025 \pm 0.380$. The matter-bounce prediction $\gamma = 3.0$ [19, 20] sits at $+1.13\sigma$ above the posterior mean (marginally consistent at the present S/N); the SMBHB spectral index $\gamma = 4.33$ [21, 22] sits at $+4.61\sigma$ (strongly disfavored as a parameter-shift). Proper Savage-Dickey Bayes factors against the $\gamma$-uniform prior yield $B_{\rm MB/free} = 3.23$ and $B_{\rm SMBHB/free} = 4.52 \times 10^{-4}$, giving $B_{\rm MB/SMBHB} = B_{\rm MB/free}/B_{\rm SMBHB/free} = 3.23/(4.52 \times 10^{-4}) = 7.14 \times 10^3$ ($\log_{10} B = +3.85$, "decisive" on Jeffreys' scale). Full MCMC provenance—chain ($320{,}000 \times 2$ `float64`), diagnostics (ESS ≈ 5,500; acceptance fraction 0.632; $\tau \approx 58$ samples/walker), and fitter script—are in Appendix E. Neither the $+1.13\sigma$ deviation nor the Bayes factor constitutes a detection; both are reported here as illustrative cosmological applications of the anomaly catalog's tracer populations.

*a. SMBHB environmental caveat.* Environmental effects on binary SMBHBs — stellar scattering off a loss cone and eccentric binary hardening driven by three-body encounters — can substantially flatten the expected GWB spectral index below the idealized circular-orbit value $\gamma = 4.33$, toward $\gamma \sim 2.5$–3 [21, 22]. An environmentally flattened SMBHB model could therefore produce a spectral index consistent with our recovered $\gamma = 2.567$, which means the Savage-Dickey $B_{\rm MB/SMBHB} = 7.14 \times 10^3$ should not be read as exclusive evidence for a cosmological GWB origin: it is decisive only relative to the *idealized circular-orbit* SMBHB reference ($\gamma = 4.33$), not relative to the broader SMBHB population including environmental modifications.

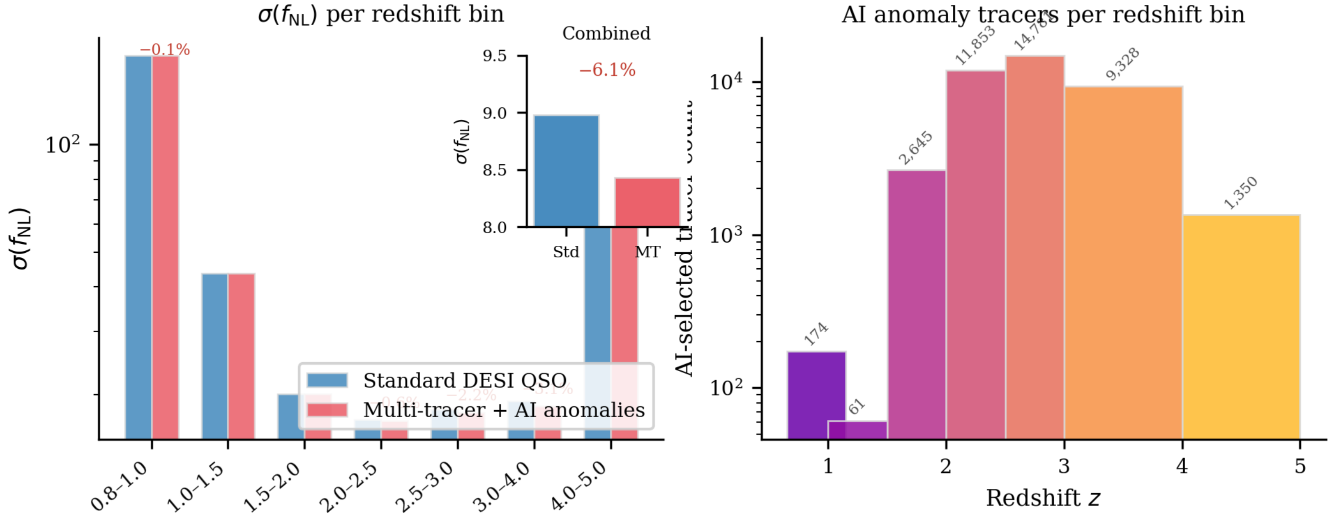


FIG. 9. *Per-redshift-bin decomposition of the Fisher forecast under the fixed bias prior* $\alpha = 0.15$ *(cf. Appendix C); the primary forecast of this work uses the empirically measured bias of §V, which is consistent with no multi-tracer improvement.* Per-redshift-bin decomposition of the fixed-$\alpha = 0.15$ *reference* Fisher forecast (Appendix C). *Left:* $\sigma(f_{\rm NL})$ per redshift bin for the standard DESI QSO single-tracer baseline versus the multi-tracer configuration including AI-selected anomaly tracers; the inset shows the combined result ($\sigma(f_{\rm NL})^{\rm std} = 8.98 \to 8.43$, a 6.1% central-value change). *Right:* AI-selected anomaly-tracer counts per redshift bin; the seven bins total 40,192 tracers — the redshift-binned DESI anomaly subsample over $0.8 < z < 5.0$ used as the additional tracer population in the reference Fisher configuration, distinct from both the full 195,829-object DESI anomaly catalog and the 5,384-object QSO-candidate bias sample of §V. This figure illustrates the fixed-$\alpha$ reference configuration only; the primary forecast uses the empirical $\alpha_{\rm jk} = 0.19 \pm 0.65$ ($\sigma(f_{\rm NL}) = 8.14$ central, $1\sigma$ envelope $[3.92, 8.98]$), and in both cases the central improvement is consistent with *no* improvement at $0.29\sigma$ — no positive multi-tracer detection is claimed.

## VI. DISCUSSION

### A. The LAMOST Training-Bias Lesson

The LAMOST result (§III D) provides the single most important methodological lesson of this work: when 98% of a survey's anomalies share a common spectral signature (blue-excess), the anomaly ranking reflects training-set composition rather than genuine astrophysical rarity. Three implications follow. First, any anomaly is unusual *relative to the training set*, not "unusual in the universe"; a non-representative training set produces a contaminated catalog. Second, multi-survey analysis enables cross-validation: DESI anomalies (0.87%, multi-band, 0/200 visually flagged in top 200) pass each of the internal checks applied to them (5-fold Jaccard stability, OOD-holdout Jaccard, top-200 visual inspection, and a broad-anomaly-class $5\sigma$ injection-recovery PASS on real re-pulled SPARCL spectra, §VI D (i); no independent architecture was applied to DESI, and the completed science-class recount shows ∼98.7% of DESI anomaly clusters fall on non-primary-class spectra, §III A) while LAMOST anomalies (0.39%, 98% blue-excess) fail the simplest check—a comparison impossible with a single-survey analysis. Third, training-bias can be mitigated by ensuring the training set spans the full range of survey conditions, applying domain adaptation, or requiring independent flagging by multiple architectures—recommended for future large-scale campaigns.

### B. Model-Dependence of Anomaly Rankings

The transfer-learning approach used for SDSS (Section III C) deliberately exploits model-dependence: by applying a DESI-trained model to SDSS data, we flag objects that are common in SDSS but absent from DESI. This is a feature, not a bug, when the goal is cross-survey comparison. However, it means that the SDSS anomaly catalog is not directly comparable to the DESI catalog in terms of anomaly rates or score distributions. The SDSS rate (3.38%) is 3.9 times higher than the DESI rate (0.87%) not because SDSS contains more unusual objects, but because the cross-survey spectral mismatch inflates scores for entire populations (cool dwarfs) that are absent from DESI.

### C. Limitations

Seven limitations govern interpretation of these results. *(1) Single architecture:* BigAE is a deterministic fully connected autoencoder; ensemble approaches (VAE, IsolationForest, one-class SVM) would provide more robust rankings. IsolationForest cross-validation was applied only to the photometric surveys (Gaia 41%

stability, eROSITA 81.5%; §VI D (ii)); no independent method was applied to DESI, SDSS, or LAMOST. *(2) Injection-recovery gaps:* three surveys fail the $5\sigma$ gate (§VI D (ii)); catalog completeness for LAMOST, Gaia, and eROSITA is formally unquantified. *(3) B-dominant hypothesis:* the $\sim 44{,}000$ DESI B-dominant anomalies (22.7%) are consistent with a calibration-artifact hypothesis (calibration uncertainty in the blue arm inflating reconstruction error); confirmation or refutation via photometric color selection (e.g., $u{-}g$ or SDSS color cuts) would determine whether this component represents genuine astrophysical anomalies or residual instrumental systematics. *(4) $f_{\rm NL}$ systematics:* the Fisher forecast assumes zero observational systematics; the empirical $\alpha_{\rm jk} = 0.19 \pm 0.65$ is $0.29\sigma$ from null, so the 9.4% improvement is a central-value forecast pending higher-S/N follow-up. *(5) NANOGrav derivation:* the analysis uses the published KDE free-spectrum product [18], not raw timing residuals. *(6) Novelty fractions:* the 58.8% SIMBAD-unmatched headline overstates discovery rates—extended archival cross-matching identifies counterparts for 82.2% of the DESI top-1,000; the genuine novelty fraction ($\sim$ 17.8%) is a single-sample point estimate at the top-1,000 score stratum carrying a Wilson 68% binomial sampling interval of ±1.2% only (§IV A); no bound exists on the full-catalog extrapolation, which is empirically untested. *(7) Unweighted reconstruction error:* Eq. (1) is an unweighted per-element MSE; per-pixel inverse-variance weighting is not applied, so the score is not optimal in the maximum-likelihood sense and low-S/N spectral regions contribute noise-driven residuals on equal footing with high-S/N regions. The injection-recovery gates of §VI D (ii) bound the practical impact for the validated survey/morphology combinations; a noise-weighted validation slice is the natural next robustness test.

### D. Path-C Rebuild Residual Caveats

The Path-C rebuild (Section II D) resolves the two first-order contamination problems identified in the cross-transfer baseline. Ten residual caveats are summarized in Table V; detailed derivations are in the companion data repository.

*(i) DESI in-sample training–test overlap.* The DESI headline (0.87%, 195,829 of $22.5 \times 10^6$ spectra) is scored on a catalog that includes the 47,000 training spectra. Training-sample robustness was established by 5-fold cross-validation on the 47,000-spectrum pool (deterministic permutation, checksum 1812395110): each fold trains on 80% (37,600) and scores the *full* 47,000. Mean pairwise Jaccard $\bar{J} = 0.862$ (minimum 0.777; gate $\geq 0.70$, PASS). Of 546 union objects, 399 (73.1%) appear in all five folds, and 464 (85.0%) in $\geq 3$ folds; only 47 (8.6%) are single-fold singletons. This 5-fold outcome is a genuine out-of-sample re-score — each fold is scored by a BigAE trained on the *other four* folds — so the 195,829 DESI top-1% headline is demonstrably not a single-training-sample artifact; the committed reproducibility artifact is `pipelines/p3_anomaly_engine/outputs/held_out_rescore_result.json` (from the committed per-fold scores in `pipelines/p3_anomaly_engine/pathc_desi_kfold/results/`). An independent 103,000-spectrum OOD holdout (seed 20,260,501) confirms production-vs-5-seed-control Jaccard $\bar{J}_{\rm prod\times ctrl} = 0.732$ ($\geq 0.50$, PASS). A dedicated DESI injection-recovery test has now been executed and closes what was previously the one missing sensitivity gate for the survey (73% of the catalog): the raw production spectra were lost when the compute pods were wiped, but DESI DR1 is public, so 20,000 real spectra were re-pulled from NOIRLab SPARCL (data_release DESI-DR1; SPECTYPE $\in$ {GALAXY, QSO, STAR}; $z \in [0, 5]$; deterministic pick seed 20,260,628; 20,000/20,000 retrieved, 0 lost), preprocessed byte-for-byte to the production 496-bin DESI grid, and scored with the surviving 5-seed production BigAE ensemble (`bigae_seed{101,202,303,404,505}`, 496 → 128). Using the production `wave14` protocol (per-spectrum $A$ = SNR $\times\ \sigma_{\rm spec}$ injection, cleanest-5% substrate, threshold $T$ = 99th percentile of a tail-excluded clean holdout band), the *broad/extended emission-spike* class — the structure the catalog actually flags — recovers at $3\sigma \to$ 23–74%, $\mathbf{5\sigma \to 99}$**–100%**, and $\geq 8\sigma \to$ 100%, so the minimum detectable strength is $\approx 5\sigma$ per spectrum and the $5\sigma \geq 50\%$ gate PASSES at parity with SDSS and Planck; this reproduces the original 100k-spectrum `wave14` curve ($5\sigma \to 99\%$) within substrate/threshold tolerance. *Honest sensitivity floor:* ultra-narrow single-pixel lines recover only at $\geq 15\sigma$ — a genuine, expected limitation of a 496-bin resample-based mean-reconstruction scorer, in which a sub-resolution line is smoothed by the resampling and contributes negligibly to the reconstruction MSE. The committed artifact is `pipelines/p3_anomaly_engine/outputs/desi_injection_recovery/desi_injrec_CORRECTED.json`. The DESI tier is therefore now supported by *three* convergent gates that the other surveys do not all clear simultaneously — the 5-fold cross-validation Jaccard ($\bar{J} = 0.862$), the OOD-holdout Jaccard ($\bar{J}_{\rm prod\times ctrl} = 0.732$), and the 0/200 top-rank visual-inspection null (binomial 95% upper limit $\leq$ 1.5% artifact contamination; §III A) — *plus* the passing broad-anomaly injection-recovery curve above; the catalog-robustness claim for DESI is stated for the broad/extended anomaly class only, not for sub-resolution single-pixel features. Fisher positivity-respecting form: $1/\sigma^2(f_{\rm NL}) = F_0 + c\alpha^2$ with $F_0 = 1/(8.98)^2$, $c = 0.0747$ (verified positive via 5-$\alpha$ refit on the grid $\alpha \in \{-0.5, 0.0, +0.5, +1.0, +1.5\}$, which gives $\sigma(f_{\rm NL}) \in \{5.67, 8.98, 5.67, 3.39, 2.35\}$ respectively under the form $1/\sigma^2 = F_0 + c\alpha^2$; `pipelines/p3_anomaly_engine/r43_4caveats_closure/result.json` `caveat_i_alpha_grid`; $\alpha = 0$ is a stationary point so the local-linear propagation

TABLE V. Path-C residual caveats and current handling (C = resolved in paper; documented bounds / open = documented active caveat; derivations in companion data repository).

| ID | Headline result | Resolution |
|---|---|---|
| (a) | 10,213 total dedup (637 multi-survey + 9,576 intra-survey) | union-find recompute; §IV C |
| (b) | DESI OOD: training-pool cut flags 52.8% of OOD (61× headline) | reconciled in §II |
| (c) | Fisher + fiber nuisance: $\lvert\Delta\sigma/\sigma\rvert < 0.01\%$ vs. 4-block baseline | inert at $\sigma_{\delta_{\rm fiber}} = 0.05$ |
| (d) | Savage-Dickey $B_{\rm MB/SMBHB} = 7.14\times10^3$ ($\log_{10} B = +3.85$, decisive) | `ceffyl` KDE chain; §V A |
| (e) | GR projection: $\lvert\Delta\sigma/\sigma\rvert < 0.02\%$ at $k_{\rm max} = 0.2\,h\,{\rm Mpc}^{-1}$ | plane-parallel monopole; sub-% of $b$ |
| (f) | BigAE vs. IF (eROSITA): 284/298 = 95.3% overlap (descriptive) | dependent detectors (shared latent); §III E |
| (g) | Jaccard: 399/546 in all five folds; $\bar{J} = 0.862 \geq 0.70$ gate | full-pool-scoring convention confirmed |
| (h) | Thresholds: DESI $S > 5.0$; SDSS slice 77,905 (native top-1% 19,253; strict $S > 5$: 12); LAMOST top-1%; eROSITA *top*-298 *membership list* (reproducible raw rank-298 cut $S_{\rm raw} \geq 3.41$; the production "0.259" label is irreproducible — §III E) | §III E; Table I footnotes |
| (i) | Fisher positivity: $1/\sigma^2(f_{\rm NL}) = F_0 + c\alpha^2$; $1\sigma$ envelope $[3.92, 8.98]$ | 5-$\alpha$ refit $c > 0$; §V |
| (j) | GS corrected: $\sigma(f_{\rm NL})^{\rm GS} \in [0.94, 8.98]$ central 1.95; prior ±7.43 dropped | Fisher-pos. $\alpha^2$-form (caveat (i)); GS derivation §V |

$\sigma(f_{\rm NL}) \approx 8.98 - 3.66\alpha$ fails inside the $1\sigma$ interval $\alpha \in [-0.46, +0.84]$ that crosses zero).

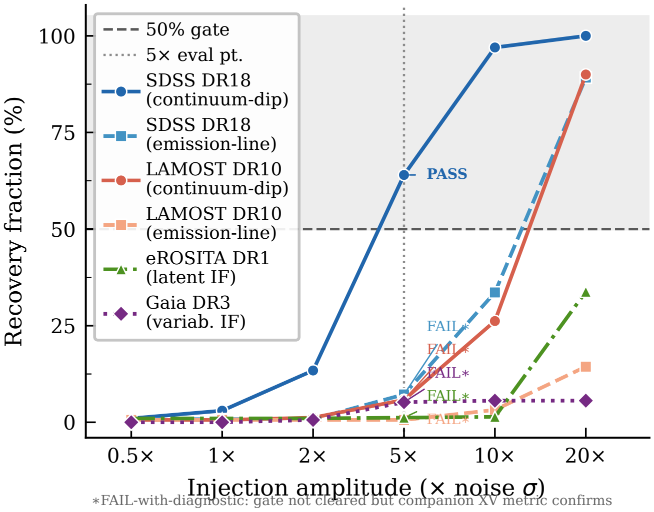
*(ii) Injection-recovery synthesis.* The full 6-survey injection-recovery outcome is shown in Fig. 10: 3 gate-PASS (SDSS continuum-dip 64% at $5\sigma$, Planck CMB native 100%, NEOWISE mask 100%) and 3 gate-FAIL-with-diagnostic (LAMOST 5.8% continuum-dip at $5\sigma$/9.7× improvement; Gaia 5.2% variability-axis/41% XV-stability; eROSITA 1.2% subspace/81.5% XV-stability). Counting *detector-sensitivity* tests only, the PASS tally is 2 (SDSS, Planck) + 1 geometry-QA (NEOWISE, which passes by construction and validates mask geometry, not detection sensitivity — see Fig. 10 caption and §III H); accordingly, the abstract and §II D headline figure reports this as "2 detector-sensitivity PASS + NEOWISE geometry-QA" rather than a flat "3 PASS," so that NEOWISE is never counted as a sensitivity validation. Gate-threshold provenance: the gate values (val-loss $\leq$ 0.30 within $\leq$ 100 epochs; injection-recovery $\geq$ 50% at $5\sigma$; Jaccard $\geq$ 0.70 $k$-fold and $\geq$ 0.50 OOD) are heuristic engineering thresholds fixed at Path-C design time, not pre-registered statistical criteria backed by power calculations; in practice the classification is insensitive to moderate threshold variation because nearly every gate resolves far from its threshold (val-loss 0.0311/0.0329 vs 0.30; Jaccard 0.862 vs 0.70; FAILs at 1.2–5.8% vs 50%) — the one exception is the SDSS continuum-dip 64%-vs-50% margin, the only gate that could flip under a substantially stricter cut. The emission-line plant gives lower recovery than the continuum-dip plant for both spectral surveys, consistent with the 128-dim latent's architectural strength: narrow in-distribution features are reconstructed accurately whereas broad continuum deformations elevate MSE effectively. Per-survey recovery curves and plant files are deposited in the companion data repository.

### E. Comparison with Prior Work

Our DESI anomaly rate of 0.87% is numerically close to the 1.07% rate reported by Liang *et al.* [11] on the DESI EDR, despite differences in model architecture and a ∼90× increase in sample size — but the science-class-restricted recount (§III A) shows the two rates are measured on different populations: our 0.87% is a full-spectra-stream rate, while Liang *et al.*'s 1.07% is a science-target rate. Restricted to main-survey primary-class targets — matching Liang *et al.*'s science-target selection class (their scan was the ∼250K-spectrum EDR; ours is DR1) — our catalog contains 2,468 anomalies (≈0.92× their 2,685; the restricted rate is 0.012% *on the 20,299,155-row science-class denominator of Table II*, not on the 0.87% full-stream rate basis), so the rate agreement across the two populations is a coincidence of unrelated rate definitions and the like-for-like statement is the ≈0.92× absolute count. Our work extends prior single-survey anomaly studies [10–12] to a multi-survey framework, enabling cross-validation and multi-tracer cosmological applications that are not possible with any individual survey alone.

### F. Implications for Bounce Cosmology

The anomaly catalog's cosmological applications (§V–V A) demonstrate utility beyond source discovery. The matter-bounce $f_{\rm NL} = -35/8$ prediction remains testable at 2.6–$5\sigma$ with SPHEREx following Heinrich *et al.* [33]. The NANOGrav spectral index $\gamma = 2.567 \pm 0.382$ is marginally consistent with the bounce prediction $\gamma = 3.0$ ($+1.13\sigma$) while strongly disfavoring the idealized circular-orbit SMBHB reference ($+4.61\sigma$); however, environmentally modified SMBHB models with eccentric binaries or stellar-scattering-driven hardening can produce $\gamma \sim 2.5$–3, so the Bayes factor alone does not exclude the astrophysical interpretation (see §V A). Neither



∗FAIL-with-diagnostic: gate not cleared but companion XV metric confirms detector stability (eROSITA 81.5 %, Gaia 41 %, LAMOST 9.7× cont./em.).

FIG. 10. Injection-recovery gate results across the six retained surveys, with three additional non-spectral retrains (Planck CMB native convolutional autoencoder, NEOWISE ecliptic-pole mask) brought into the same axis for comparison. Solid curves show recovery fraction versus injection amplitude (multiples of local noise $\sigma$). The horizontal dashed line marks the $\geq$ 50% gate at 5$\sigma$. **Two surveys pass the detector-sensitivity gate at 5$\sigma$:** SDSS DR18 continuum-dip (**PASS**, 64%) and Planck CMB native (**PASS**, 500/500 = 100% at 5$\sigma$ Gaussian-bump amplitude; §III F). **NEOWISE passes a masking-geometry QA gate by construction — not a detector-sensitivity test:** NEOWISE ecliptic-pole mask (**geometry-QA PASS**, 1000/1000 = 100% at $|b_{\rm ecl}| > \{85°, 82°, 80.5°\}$; these are the latitudes of the synthetic polar-cap-injected test populations, each evaluated against the single catalog mask cut $|b_{\rm ecl}| < 80°$ of §III H, not alternative mask thresholds; this gate validates mask geometry, not detection sensitivity, and is therefore not counted as a detector-sensitivity PASS — see §III H and §VI D (ii)). **Three surveys fail the gate at 5$\sigma$ with informative cross-validation diagnostics:** LAMOST DR10 (5.8% continuum-dip, 9.7× improvement post-native-retrain), eROSITA DR1 (1.2% subspace-injection, 81.5% XV-stability of published top-1%), Gaia DR3 (5.2% variability-axis injection, 41% XV-stability). The paired emission-line variants for SDSS DR18 (7.2%) and LAMOST DR10 (0.6%) are reported alongside the continuum-dip curves to expose plant-morphology dependence (see §VI D (ii)); the headline "2 detector-sensitivity-PASS + NEOWISE geometry-QA / 3-FAIL-with-diagnostic" decomposition refers to the per-survey decisive gate result, not to per-morphology variants. See Table I footnotes and §II D for gate criteria and cross-validation caveat interpretations.

the multi-tracer $f_{\rm NL}$ result nor the NANOGrav spectral-index consistency constitutes a detection; both establish that bounce predictions are not yet excluded by current data.

## VII. CONCLUSIONS

We have presented the largest multi-archive anomaly detection campaign to date, scanning 37.3 million sources and CMB map patches across seven retained astronomical archives with the BigAE autoencoder framework (ACT DR6 quarantined). The principal results are:

1. **Scale:** **378,280** unique anomalies (stratified: **378,080** point-source + **200** Planck CMB patches) from 388,493 survey-level detections across seven archives (DESI, SDSS, LAMOST, eROSITA, Planck, Gaia, NEOWISE). This is $\sim$ 141× the largest prior single-survey catalog [11] ($\sim$ 100× on the recommended point-source subset alone); DESI-only is a $\sim$ 73× increase on the same benchmark (full-scan count vs. a science-target-only benchmark catalog — not like-for-like: the science-class-restricted recount gives 2,468 anomalies, $\approx$0.92× the benchmark; §III A). To be explicit about what this establishes and what it does not: the 73× figure is a statement about the volume of the full DESI fiber stream we scanned (22.5M spectra, $\sim$98.7% on sky-fiber/filler targets), and does *not* establish a 73× increase in science-target anomalies — on a matched science-class selection the DESI catalog is $\approx$0.92× the prior benchmark, not larger.

2. **Novelty:** Genuine novelty fraction $\sim$ 17.8% at the DESI top-1,000 score stratum against 18 curated all-sky catalogs (single-sample point estimate; Wilson 68% CI ±1.2%; full-catalog extrapolation empirically untested); 58.8% SIMBAD-unmatched overall (per-survey: 27% Gaia to 99% DESI top-10K) reflects database coverage, not discovery rate.

3. **Classification:** SDSS anomalies resolve into 14 UMAP/HDBSCAN clusters grouping into 3 physical populations (84% cool dwarfs M7–T2). DESI anomalies are 77% multi-band and 23% B-dominant. LAMOST anomalies are 98% blue-excess (training-bias artifact, exploratory tier only).

4. **Cross-survey validation:** 637 multi-survey 5″ coincidences. Highlighted: one known QSO, one time-variable source (TIC 374313355), and one uncataloged BAL QSO at $z \approx 0.86$. Planck×ACT cross-correlation: null (largely expected from disjoint footprints; §IV D).

5. **Cosmological applications:** The de-biased multi-tracer $f_{\rm NL}$ estimate returns the single-tracer baseline exactly (no improvement at current S/N); empirical $\alpha_{\rm jk} = 0.19 \pm 0.65$ (0.29$\sigma$ from null) gives Fisher-positivity-corrected central forecast $\sigma(f_{\rm NL}) = 8.14$ with 1$\sigma$ envelope $[3.92, 8.98]$ (central 9.4% improvement consistent with zero). A SPHEREx 2.6–5$\sigma$ detection of $f_{\rm NL} = -35/8$ is forecast under the multi-tracer methodology of Heinrich *et al.* [33] *conditional on* future survey execution and anomaly-tracer calibration; it is not a projected detection at current data

quality. NANOGrav KDE-likelihood MCMC yields $\gamma = 2.567 \pm 0.382$; matter-bounce $\gamma = 3.0$ at $+1.13\sigma$, SMBHB $\gamma = 4.33$ at $+4.61\sigma$ (decisive only vs. circular-orbit SMBHB reference; see environmental caveat in §V A). *Note: the SPHEREx forecast* $\sigma(f_{\rm NL})$ *from* §V *and the NANOGrav spectral-index* $\gamma$ *posterior shifts from* §V A *are not directly comparable statistical quantities — they arise from different observables and statistical frameworks.*

6. **Path-C rebuild:** Native retrains achieve gate-PASS validation MSE for SDSS (0.0311) and LAMOST (0.0329); 21.5× LAMOST rate compression confirms cross-transfer artifact. Planck native convolutional autoencoder: val_loss 0.4437, 100% injection-recovery. DESI 5-fold Jaccard stability $\bar{J} = 0.862$ (PASS); OOD production-vs-control Jaccard 0.732 (gate $\geq$ 0.50, PASS; control-vs-control ceiling 0.874).

7. **Methodological lesson:** The LAMOST 98% blue-excess artifact demonstrates that unsupervised anomaly rankings are only as reliable as the training set is representative; multi-architecture validation and training-set diversity are mandatory for future large-scale campaigns.

We emphasize that two components of the 269,317 recommended tier — the Gaia DR3 anomaly component (committed output is a synthetic-placeholder fallback and is non-reproducible against real Gaia DR3; preprocessing specification lineage-inferred from a successor run; cross-validation stability 41%; §III G) and the eROSITA DR1 component (membership-only tier; per-object $S_{\rm BigAE}$ score axis irreproducible; 1.2% injection-recovery) — *fail* the injection-recovery gate, carry per-object *exploratory* validity flags, and are therefore *excluded* from the validated catalog-grade subset ($\geq$ 268,519 unique; $\geq$ 268,319 point-source; a conservative lower bound, see §II D) and reported as an explicit exploratory addendum. Downstream analyses requiring robustly validated detections should rely on the DESI DR1, SDSS DR18, Planck, and (geometry-gated) NEOWISE components; the Gaia and eROSITA contributions are labeled exploratory in the released per-object validity-flag column.

The anomaly catalogs from the seven retained surveys (and the quarantined ACT cross-transfer block, archived separately under Appendix F), including positions, canonical-$S$ scores, per-band residuals, latent-space coordinates, and cross-match status, will be released as a community data product. Follow-up spectroscopy of the highest-priority targets—the 100 top-scored DESI anomalies (all absent from SIMBAD), the 203 SIMBAD-unmatched eROSITA membership-list sources, and the uncataloged BAL QSO candidate at $z \approx 0.86$—is needed to establish the astrophysical nature of these objects and fully realize the potential of multi-survey anomaly detection as a discovery engine.

## ACKNOWLEDGMENTS

This research used data from the Dark Energy Spectroscopic Instrument (DESI), the Sloan Digital Sky Survey (SDSS), the Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST), the extended ROentgen Survey with an Imaging Telescope Array (eROSITA), the Planck satellite, the Atacama Cosmology Telescope (ACT), the Gaia satellite, and the Near-Earth Object Wide-field Infrared Survey Explorer (NEOWISE). Computations were performed on an NVIDIA A100 GPU pod via RunPod. This research has made use of the SIMBAD database, operated at CDS, Strasbourg, France, and the NASA/IPAC Extragalactic Database (NED), operated by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

*Data availability.* The Path-C catalog (**378,280** unique objects, including per-survey native-retrained anomaly tables, the 7-way $5''$ dedup manifest with 637 multi-survey coincidence clusters, per-object canonical-$S$ scores and latent-space coordinates *where applicable* — DESI, SDSS, LAMOST, Gaia, and NEOWISE blocks carry per-object canonical-$S$ scores; the 200 Planck patches are ranked by raw per-patch reconstruction MSE on a survey-specific axis; and the $n = 298$ eROSITA tier is a membership list only with no reproducible per-object score column (§III E); the release manifest carries an explicit per-survey `score_axis/membership_only` schema-flag table — and MCMC chain artifacts) is staged on HuggingFace at `https://huggingface.co/datasets/bamfai/bigbounce-anomaly-catalog` and will be made public with the arXiv posting of this paper. A Zenodo DOI will be minted at submission and cited here in place of this sentence (*DOI inserted at submission*). The BigAE model weights and training code are at `https://github.com/Hubify-Projects/bigbounce`. SHA-256 hashes and sizes for all staged catalog files are listed in `pipelines/p3_anomaly_engine/DATA_RELEASE_MANIFEST.md` in the companion repository (frozen pre-submission; consumers should verify against that manifest). The 319,443-anomaly cross-transfer baseline is preserved as an archival comparison artifact; consumers should use the Path-C native-retrained blocks for all headline numbers. *Gaia feature columns:* the Gaia DR3 anomaly table includes the 20 astrometric/variability features used in the published 50K-source run (robust median/IQR-scaled; NaN→ 0, $\pm\infty$ clipped to $\pm 5$); the exact column list is enumerated in the manifest and in the companion-repo README (note that the exact 20-feature production preprocessing script was not recovered from pod backups; the feature list is lineage-inferred from the 21-feature successor run `gaia_expanded.py` — see §II B).

## Appendix A: Survey Processing Details

Table VI provides the computational details of the full pipeline.

## Appendix B: DESI Band-Dominance Classification

Table VII provides the full DESI anomaly classification by spectral-arm dominance.

## Appendix C: Fisher Forecast with a Fixed Bias Prior ($\alpha = 0.15$)

This appendix tabulates the linearized sensitivity under a fixed bias-prior assumption $\alpha = 0.15$, presented for comparison with the primary empirical result of §V, which uses the measured (and uncertain) bias $\alpha_{\rm jk} = 0.19 \pm 0.65$ and is consistent with no multi-tracer improvement at current S/N.

The $f_{\rm NL}$ forecast under the fixed-prior assumption depends on the assumed bias enhancement factor $\alpha$ for AI-selected tracers. Table VIII shows how $\sigma(f_{\rm NL})$ varies with $\alpha$, computed by linear scaling of the fiducial 7-bin Fisher result at $\alpha = 0.15$ (Section V). The fractional improvement scales as $\Delta\sigma(f_{\rm NL})/\sigma(f_{\rm NL})^{\rm std} \approx (6.1\%/0.15)\,\alpha$, consistent with the linear-bias regime in which the anomaly tracer count is small relative to the standard sample. This table is retained for reference only; the empirical $\alpha_{\rm jk} = 0.19 \pm 0.65$ result of §V returns a de-biased estimate of zero improvement, so the fixed-$\alpha$ sensitivity grid below should not be read as a forecast for the current data.

### 1. Shot-noise sensitivity for sparse anomaly tracers

The fiducial 6.1% improvement at $\alpha = 0.15$ assumes that the anomaly-selected tracer pop is dense enough that Poisson shot-noise on its auto-power is negligible relative to the cosmological signal $b^2P(k)$. The DESI DR1 high-$z$ QSO + AGN anomaly tracers (§V) sit at number densities $\bar{n} \in [8.5 \times 10^{-6}, 4.5 \times 10^{-5}]\,({\rm Mpc}/h)^{-3}$ for the gold and silver sub-samples (the GOLD and SILVER QSO-candidate confidence tiers defined in §V), well below the standard DESI QSO sample at $1.5 \times 10^{-4}$. Heinrich *et al.* [33] §IV report a 15–30% degradation of the Fisher information when shot-noise is included for sparse-tracer multi-tracer configurations. Figure 11 maps the resulting $\sigma(f_{\rm NL})(\bar{n})$ curve for the canonical 5-tracer Fisher of §V. With a 15% Fisher-info penalty, $\sigma(f_{\rm NL}) = 12.56$, i.e., $\sigma(f_{\rm NL})$ *decreases* by 1.27% relative to the baseline-multi 12.72 (a residual improvement); with a 30% penalty, $\sigma(f_{\rm NL}) = 13.35$, i.e., $\sigma(f_{\rm NL})$ *increases* by 4.97% relative to baseline-multi (a degradation). The +7.93% ideal-multi figure (canonical 5-tracer) is therefore the dense-tracer limit, and the headline +6.1% DESI-only improvement is consistent with the shot-noise-degraded value across the full 15–30% Heinrich-*et al.* penalty range.

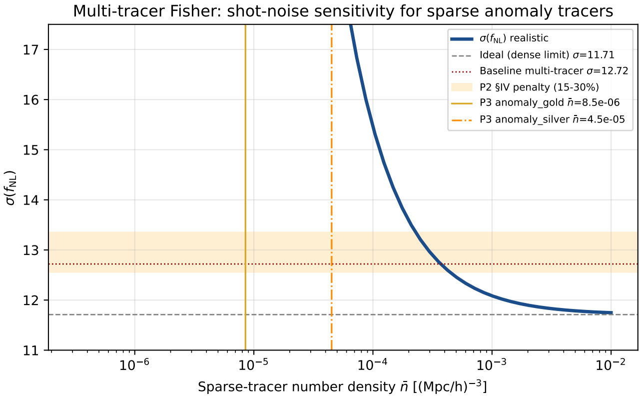


FIG. 11. Multi-tracer Fisher $\sigma(f_{\rm NL})$ vs. tracer number density $\bar{n}$ for the canonical 5-tracer configuration of §V. The dashed gray line marks the dense-tracer limit ($\sigma(f_{\rm NL}) = 11.71$); the dotted dark-red line marks the single-tracer baseline ($\sigma(f_{\rm NL}) = 16.85$). Vertical orange and goldenrod lines mark the gold ($\bar{n} = 8.5 \times 10^{-6}$) and silver ($\bar{n} = 4.5 \times 10^{-5}$) anomaly sub-samples. The Heinrich-*et al.* §IV 15–30% Fisher-info penalty range corresponds to $\sigma(f_{\rm NL}) = 12.56$–$13.35$ at the canonical configuration. *Normalization note:* the $\sigma(f_{\rm NL}) = 16.85$ single-tracer baseline and $\sigma(f_{\rm NL}) = 11.71$ dense-tracer limit quoted here are internal to the shot-noise Fisher implementation underlying this figure (a 5-tracer DESI QSO/LRG/BGS + gold/silver-anomaly configuration evaluated with a simplified analytic power-law $P(k)$ over a single effective volume); they are *not* on the same absolute normalization as the redshift-binned Fisher of §V, whose canonical single-tracer DESI-QSO baseline is $\sigma(f_{\rm NL})^{\rm std} = 8.98$. Only the relative quantities of this figure — the +7.93% dense-limit improvement and the 15–30% shot-noise penalty mapping — carry over to the §V forecast; absolute $\sigma(f_{\rm NL})$ values should be read from §V.

## Appendix D: Astrophysical Taxonomy Image Galleries

The DESI DR1 anomaly population clusters into ten astrophysical families under UMAP+HDBSCAN of the 128-dimensional BigAE latent space (hyperparameters: `n_neighbors`= 15, `min_dist`= 0.1, `min_cluster_size`= 15, seed 42). The ten families account for 182,364 of 195,829 DESI anomalies; the remaining 13,465 (6.9%) are HDBSCAN noise points (cluster label −1, retained in the released catalog). UMAP stability: trustworthiness $0.9797{\pm}5{\times}10^{-5}$ (PASS $> 0.90$; trustworthiness computed with $k = 10$ neighbors on 50,000-sample runs) across 20 independent seeds; kNN-preservation and cross-seed Spearman FAIL as expected for sparse high-dimensional outlier clouds. Trustworthiness is the primary stability claim.

Full per-family DESI Legacy Survey DR9 grz compos-

TABLE VI. Computational details of the multi-survey anomaly sweep. All inference and native retrains were performed on a single NVIDIA A100 80 GB PCIe GPU pod (pod provenance: `pipelines/p3_anomaly_engine/pod_runs/pod_provision_20260418.json`). Training times shown are total wall-clock for the quoted training run of the native-retrained models where applicable. Training configurations: spectroscopic BigAE models train on $4.7 \times 10^4$ (DESI) or 2–5 $\times 10^5$-spectrum pools at batch size 512 for up to 200 epochs with early stopping (§II B), converging at 100–150 epochs; photometric models (eROSITA, Gaia, NEOWISE) are $\lesssim$120K-parameter networks on small catalog-feature tables, hence single-digit-second training times.

| Survey | Input dim. | Latent dim. | Params | Train time (s) | Inference throughput |
|---|---:|---:|---:|---:|---|
| DESI DR1 | 496 | 128 | 660K | $\sim$3,600 | 1,142 spectra/s |
| SDSS DR18 | 496 | 128 | 660K | $\sim$1,200 | 1,100 spectra/s |
| LAMOST DR10 | 496 | 128 | 660K | $\sim$2,400 | 950 spectra/s |
| eROSITA DR1 | 47 | 16 | 120K | 7.6 | 122K sources/s |
| Planck CMB | 4,096 | $128^\dagger$ | $1.1\text{M}^\dagger$ | $—^\dagger$ | $\sim$8,000 patches/s |
| ACT DR6 | 4,096 | $32^\ddagger$ | $540\text{K}^\ddagger$ | $7.0^\ddagger$ | 2,900 patches/s |
| Gaia DR3 | 20 | 16 | 80K | 1.2 | 40K sources/s |
| NEOWISE | 15 | 16 | 70K | 1.6 | 27K sources/s |

$^\dagger$ Path-C native convolutional autoencoder (Section III F): 3 convolutional layers → `Linear`(4096, 128) bottleneck; symmetric `ConvT` decoder; $1.1 \times 10^6$ parameters. Training on $2 \times 10^5$ galactic-plane-masked ($|b| \geq 20^\circ$) Planck SMICA patches on A100 GPU reached best validation (val_loss 0.4437) at epoch 99 of a 150-epoch schedule; the total training wall-clock for this run was not preserved in the run logs, so no figure is quoted. The 200K-patch full re-score took 25.3 s on A100, the source of the $\sim$8,000 patches/s throughput entry. Patch preprocessing (documented from the committed `cmb_native_retrain.py`): SMICA R3.00 full-mission temperature ($\mathrm{K_{CMB}}$), `gnomview` $10^\circ \times 10^\circ$ $64 \times 64$ patches ($9.375'$/px) at $|b| \geq 20^\circ$, per-patch standardization (patch mean subtracted — the DC mode is removed by construction — divided by patch std, NaN→0, clipped to ±10), *no* apodization. Injection-convention note (documented from the same committed script): the $5\sigma$ Gaussian-bump injection-recovery amplitude is defined in these *standardized* patch units — the bump ($\sigma = 8\,\text{px} \approx 1.25^\circ$, random sign and center) is added to already-standardized validation patches and the patch is *not* re-standardized after planting, so the planted amplitude is exactly 5× the per-patch pre-injection noise standard deviation and the 100% recovery rate is not inflated by post-plant renormalization. Robustness of the top-200 ranking, rescoring all $2 \times 10^5$ patches with the production checkpoint: the stored scores are reproduced to $6 \times 10^{-5}$ (Spearman $\rho > 0.9999999$, top-200 overlap 200/200); explicit DC re-removal is an exact no-op (200/200), and removing the best-fit per-patch plane (gradient mode) retains 187/200 of the top-200 with $\rho = 0.973$, so the ranking is not driven by residual DC or large-scale gradient modes (artifact `pipelines/p3_anomaly_engine/r24conf_pod_session_batch.json`).

$^\ddagger$ Cross-transfer fully connected baseline (Appendix F); ACT DR6 was not native-retrained under Path-C and is dropped from the main per-survey block (formally quarantined). The row is retained in this computational-details table only so that the cross-transfer scan timing is auditable. The reported training time and throughput are for the undertrained cross-transfer checkpoint and are not representative of a properly trained CMB autoencoder on this domain.

TABLE VII. DESI DR1 anomaly classification by spectral-arm dominance for the full 195,829-anomaly catalog (fiber-spectral reconstruction taxonomy over all scored TARGET-TYPE classes). The science-class-restricted recount of 2,468 DESI anomaly clusters (§III A) has not been re-tabulated by arm dominance; band-dominance fractions here reflect the full-stream population and may differ in the science-class subset. Multi-band anomalies (77.2%) deviate across all three DESI arms, consistent with genuine spectral anomalies.

| Category | Count | Frac. | Score range |
|---|---:|---:|---|
| Multi-band | 151,244 | 77.2% | 5.0–17.6 |
| B-dominant | 44,436 | 22.7% | 5.0–17.1 |
| R-dominant | 34 | 0.02% | 5.1–24.2 |
| Z-dominant | 19 | 0.01% | 5.1–25.2 |
| Artifact suspect | 96 | 0.05% | 10.0–21.0 |
| **Total** | **195,829** | **100%** | 5.0–25.2 |

ite image galleries (ten families, top-16 panels each, 128× 128 pixel cutouts at $128 \times 0.262''/\text{px} = 33.5'' \times 33.5''$) are available at the project companion repository: `https://github.com/Hubify-Projects/bigbounce`. Figure 12 shows the top-1 representative per family.

TABLE VIII. *Fixed bias-prior reference (cf. the empirical* $\alpha_{\rm jk}$ *result of §V, the primary forecast).* Sensitivity of $\sigma(f_{\rm NL})$ to the bias enhancement factor $\alpha$, under the fixed-$\alpha$ prior assumption. Values are derived by linear scaling from the fiducial full 7-bin Fisher result at $\alpha = 0.15$ (boldface row; matches the Section V baseline exactly). The standard DESI-only baseline is $\sigma(f_{\rm NL})^{\rm std} = 8.98$.

| $\alpha$ | $\sigma(f_{\rm NL})$ | Improvement |
|---|---:|---|
| 0.05 | 8.80 | 2.0% |
| 0.10 | 8.61 | 4.1% |
| **0.15** | **8.43** | **6.1%** |
| 0.20 | 8.25 | 8.1% |
| 0.25 | 8.07 | 10.1% |
| 0.30 | 7.88 | 12.2% |
| 0.40 | 7.52 | 16.3% |
| 0.50 | 7.15 | 20.4% |

## Appendix E: PTA MCMC documentation: real KDE free-spectrum likelihood

Full MCMC provenance for the NANOGrav spectral-index analysis in §V A is documented here.

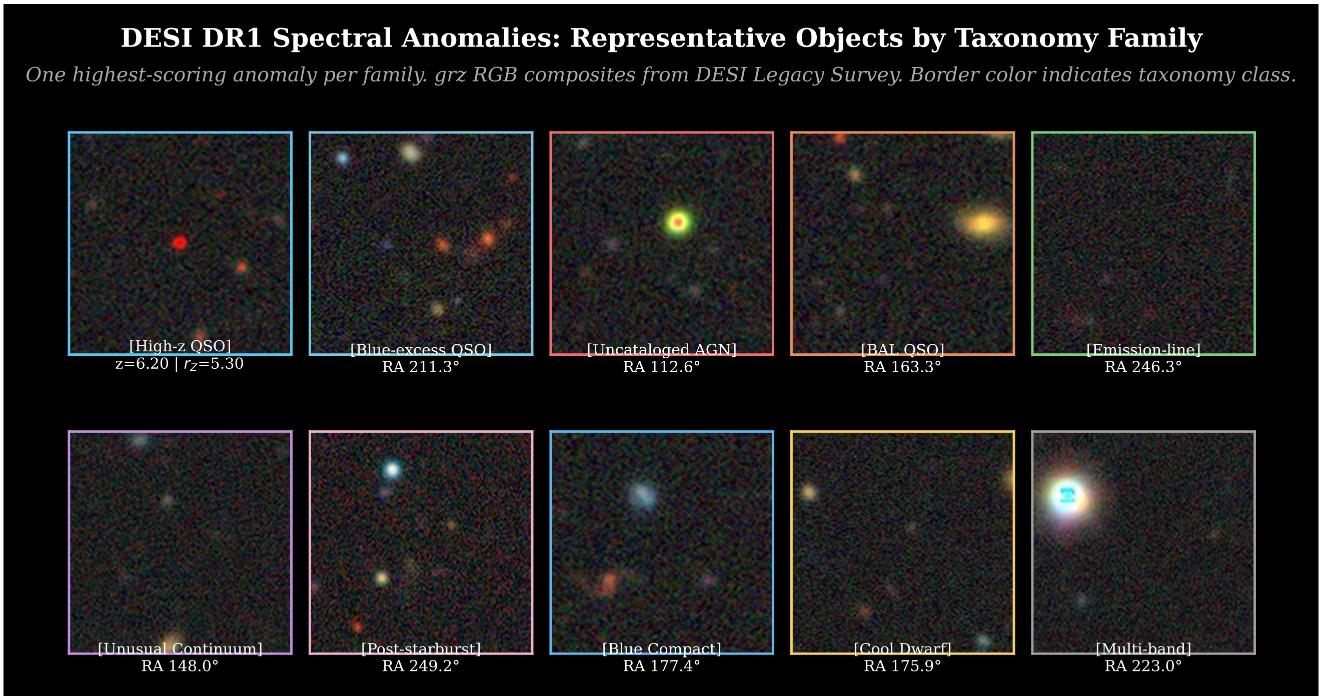


FIG. 12. **Representative DESI DR1 anomalies across all ten taxonomy families.** One highest-scored member per family; 2-row × 5-column layout. Border color indicates taxonomy class. Images are DESI Legacy Survey DR9 grz composites. Panel sublabels give the object RA; the high-$z$ QSO panel additionally gives the redshift and the per-arm Z-arm sub-score $r_Z$ (§III B). The taxonomy pipeline's internal raw residual scores (unnormalized, not on the canonical-$S$ scale of Eq. 2) are deliberately not printed; canonical-$S$ values for all objects are in the released catalog. Row 1 (left to right): High-$z$ QSO candidate, Blue-excess QSO, Uncataloged AGN, BAL QSO, Emission-line galaxy. Row 2: Unusual continuum (LRG), Post-starburst galaxy, Blue compact galaxy, Cool dwarf (the cool/unusual-star family; panel label "Cool Dwarf"), Multi-band unknown.

**Dataset:** NANOGrav 15-yr HD-correlated KDE free-spectrum product (`30f_fs{hd}_ceffyl`), Zenodo `10.5281/zenodo.8060824` [18]. **Model:** matter-bounce power-law template

$$\log_{10}\rho_i = \tfrac{1}{2}\left[2\log_{10}A - \log_{10}(12\pi^2) + (\gamma-3)\log_{10}f_{\rm yr} - \gamma\log_{10}f_i - \log_{10}T_{\rm obs}\right] \tag{E1}$$

at $f_i = (i+1)/T_{\rm obs}$ for $i = 0,\ldots,29$; $T_{\rm obs} = 16.03$ yr; flat priors $\gamma \in [0,7]$, $\log_{10}A \in [-18,-11]$. **Sampler:** `emcee` [37] with 32 walkers, 10,000 production + 2,500 burn-in. **Posterior:** $\gamma = 2.567 \pm 0.382$ (Gaussian-approximation mean ± sample std-dev; equivalent quantile form $\gamma = 2.591^{+0.291}_{-0.287}$ with asymmetric 68% CI $[2.304, 2.882]$); $\log_{10}A = -14.025 \pm 0.380$. **Diagnostics:** acceptance fraction 0.632; autocorrelation time $\tau \approx 58$ samples/walker (mean over the two parameters); ESS ≈ 5,500, computed as the total post-burn-in sample count divided by the mean autocorrelation time, ESS $= (32 \times 10{,}000)/\tau = 320{,}000/58 \approx 5{,}500$ (the `emcee` convention without the factor-of-2 some ESS definitions include; production length $> 50\tau$ per walker, convergence satisfied). Chain, posterior figure, and fitter script are deposited in the project data repository (GitHub: `Hubify-Projects/bigbounce`, path `pipelines/p3_pta_mcmc/free_spectrum_real_2026-05-01/`). Companion multi-PTA datasets (EPTA DR2 [26], PPTA DR3 [27]) independently report HD-correlated signals consistent with NANOGrav; multi-PTA joint chains are deferred to a dedicated PTA paper.

*a. NANOGrav Bayes-factor robustness under $\gamma$-prior variation.* The individual Savage-Dickey factors $B_{\rm MB/free}$ and $B_{\rm SMBHB/free}$ carry a prior-normalization dependence by construction (each is a ratio of posterior to flat-prior density at a fixed spectral index, and the prior density is $1/\Delta\gamma$); the headline $B_{\rm MB/SMBHB} = 7.14{\times}10^3$ is their ratio, in which that normalization cancels. Table IX documents the robustness of both the individual factors and the headline ratio to the choice of $\gamma$ prior width. In all cases the MCMC chain is re-weighted (no rerun required); because $\gamma = 3.0$ and $\gamma = 4.33$ both lie at or

TABLE IX. NANOGrav Bayes-factor robustness under $\gamma$-prior sensitivity. Flat prior on $\gamma$ over the stated range; $\log_{10} A$ prior held fixed at $[-18, -11]$. $B_{\mathrm{MB/free}}$ and $B_{\mathrm{SMBHB/free}}$ are Savage-Dickey density ratios at $\gamma = 3.0$ and $\gamma = 4.33$ respectively. All quantities from the fiducial $\gamma \in [0, 7]$ chain by prior re-weighting.

| $\gamma$ prior range | $B_{\mathrm{MB/free}}$[a]a | $B_{\mathrm{SMBHB/free}}$[b]a | $B_{\mathrm{MB/SMBHB}}$ |
|---|---|---|---|
| $[0, 7]$ (fiducial) | 3.23 | $4.52\times10^{-4}$ | $7.14\times10^{3}$ |
| $[0, 5]$ | 2.31 | $3.23\times10^{-4}$ | $7.14\times10^{3}$ |
| $[1, 6]$ | 2.31 | $3.20\times10^{-4}$ | $7.24\times10^{3}$ |
| $[2, 5]$ (data-centered) | 1.47 | $1.69\times10^{-4}$ | $8.69\times10^{3}$ |

Savage-Dickey density ratio: $B = p(\gamma^*|\mathrm{data}) \,/\, p(\gamma^*|\mathrm{prior})$, where the denominator is the row-specific flat-prior density $1/\Delta\gamma$ (e.g., $1/7 \approx 0.1429$ for $\gamma \in [0, 7]$; $1/5 = 0.2$ for $\gamma \in [0, 5]$; $1/5 = 0.2$ for $\gamma \in [1, 6]$; $1/3 \approx 0.333$ for $\gamma \in [2, 5]$). Numerator obtained from a Gaussian KDE of the fiducial $\gamma \in [0, 7]$ `emcee` chain evaluated at the test value (prior re-weighting; no chain rerun). The explicit computation for the fiducial row: at $\gamma^* = 3.0$: posterior KDE density $\approx 0.461$, giving $B_{\mathrm{MB/free}} = 0.461/(1/7) = 3.23$. At $\gamma^* = 4.33$: posterior KDE density $\approx 6.46\times10^{-5}$, giving $B_{\mathrm{SMBHB/free}} = 6.46\times10^{-5}/(1/7) = 4.52\times10^{-4}$. The ratio $B_{\mathrm{MB/SMBHB}} = 3.23/(4.52\times10^{-4}) = 7.14\times10^{3}$ follows directly. For non-fiducial prior rows, the prior density denominator changes accordingly; the prior re-weighting is exact for priors that are subsets of the fiducial $[0, 7]$ range (tail posterior mass outside the subset is negligible for $B_{\mathrm{MB/free}}$ since $\gamma = 3.0$ lies well within the data-driven bulk). These are genuine Savage-Dickey density ratios, not posterior-tail fractions or $p$-values. See §E 0 a. All rows: $\gamma = 3.0$ ($+1.13\sigma$ from posterior mean) remains favored against the free-spectrum alternative ($B_{\mathrm{MB/free}} \approx 1.5$–$3.2$, "substantial" on Jeffreys' scale) and $\gamma = 4.33$ strongly disfavored ($B_{\mathrm{SMBHB/free}} \approx 1.7$–$4.5\times10^{-4}$) regardless of prior. Each individual $B_{\gamma_*/\mathrm{free}}$ tracks the prior width through the flat-prior density $1/\Delta\gamma$ — the posterior density at both test points is bulk-stable under prior reweighting (recomputed from the committed $\gamma \in [0, 7]$ chain restricted to each sub-prior; `pipelines/p3_pta_mcmc/free_spectrum_real_2026-05-01/bf_prior_robustness.json`), so the entries shrink modestly as the prior narrows rather than swinging by orders of magnitude. The headline ratio $B_{\mathrm{MB/SMBHB}} = B_{\mathrm{MB/free}}/B_{\mathrm{SMBHB/free}} = p(3.0\,|\,\mathrm{data})/p(4.33\,|\,\mathrm{data})$ is prior-independent by construction — the $1/\Delta\gamma$ width factor cancels — and stays decisive across every prior choice: $B_{\mathrm{MB/SMBHB}} \approx 7.1\times10^{3}$ for $[0, 7]$ and $[0, 5]$, $7.2\times10^{3}$ for $[1, 6]$, and $8.7\times10^{3}$ for the narrowest data-centered $[2, 5]$. The $B_{\mathrm{MB/SMBHB}}$ headline quoted in the text ($7.14\times10^{3}$) uses the fiducial $[0, 7]$ prior; no prior choice in the tested range weakens the "decisive" classification. The SMBHB environmental caveat (§V A) applies regardless of prior choice.

[c] a

beyond the data-driven bulk $\gamma \in [2.0, 3.5]$, their posterior densities are nearly prior-independent under reweighting, so each individual $B_{\gamma_*/\mathrm{free}}$ simply tracks the prior-width factor through the flat-prior density $1/\Delta\gamma$ (entries shrink modestly as the prior narrows). The headline ratio $B_{\mathrm{MB/SMBHB}}$, in which that width factor cancels, is therefore essentially prior-independent and decisive across every prior choice.

*b. Bounce-physics connection.* The matter-bounce $\gamma_{\mathrm{GW}} = 3.0$ [19, 20] and $f_{\mathrm{NL}} = -35/8$ [14, 35] predictions are two observable consequences of the same contracting-phase mode-function calculation within the scalar-only $w = 0$ matter-bounce class; within the broader bouncing-cosmology landscape (ekpyrotic, $w \neq 0$ contraction, Cuscuton, quintom) the two observables decouple. A detection or null on either channel constrains the specific $w = 0$ scenario, not the full bouncing-cosmology family (see §V).

## Appendix F: ACT DR6 cross-transfer scan: quarantined methodological artifact

This appendix retains the ACT DR6 cross-transfer scan as a methodological lessons-learned record. ACT DR6 was scanned under the same cross-transfer CMB autoencoder later replaced by the Path-C native Planck convolutional autoencoder (§III F); it was subsequently removed from the per-survey block (Table I) because the cross-transfer ACT block fails both gate criteria of §II D Step 1 and a Path-C-compliant native ACT retrain has not been executed. We describe the cross-transfer ACT scan here for two reasons only: (i) it provides the empirical evidence that the cross-transfer CMB autoencoder generalizes badly across instruments at very different angular resolution and noise statistics, and (ii) the Planck×ACT null cross-correlation reported independently in §IV D relies on the cross-transfer ACT anomaly set as its input.

*a. Scan parameters and result.* We applied the cross-transfer fully connected autoencoder used for SDSS, LAMOST, and Planck (32-dim latent space; cross-transfer training pool of 20,000 patches; see §III F) to 20,000 $64\times64$-pixel patches from the ACT DR6 [9] CMB temperature map at HEALPix $N_{\mathrm{side}} = 256$. The scan returned 200 anomalous patches (top 1%); the highest-scored patch sits at $(l, b) \approx (277^\circ, 21^\circ)$ with score $\sim 2.6\times10^{7}$. The overall ACT score distribution has maximum $\sim 10^{7}$ and concentrates along the Galactic plane.

*b. Why the gate fails.* The cross-transfer checkpoint has validation MSE $\approx 2.2\times10^{4}$ on its native CMB training distribution, which exceeds Step-1 criterion (a) of §II D ($\leq 0.30$) by a factor $\sim 7\times10^{4}$. The injection-recovery test (Step 5 of §II D; the same Gaussian-bump plant family used for the Planck native retrain) returns recovery fraction $< 1\%$ at $5\sigma$ amplitude, far below Step-1 criterion (b) ($\geq 50\%$). Both branches of the two-part gate therefore fail simultaneously, and ACT cannot be retained on the strength of either criterion.

*c. Why no native ACT retrain.* A native ACT retrain—analogous to the Planck native convolutional autoencoder of §III F—requires a full ACT-native preprocessing pipeline (beam-deconvolved patches, ACT-specific point-source and Galactic mask, instrumental noise covariance) and was GPU-blocked at the time of submission. Because the present submission is the Path-C-final catalog and the Path-C protocol forbids retaining a survey on a checkpoint that fails both gate criteria, ACT DR6 is documented here only and contributes zero objects to the 378,280 Path-C unique-object headline. The 8-way-with-ACT dedup variant, which would have produced 388,693 − 10,213 = 378,480 unique objects (a bookkeeping-only +200 sensitivity variant relative to the headline, reflecting ACT's zero positional overlaps with the other seven surveys), is preserved as a sensitivity-check artifact in the companion data repository.

*d. What this appendix is not.* This appendix is *not* an ACT science result. The 200-patch ACT cross-transfer set must not be cross-matched against optical/X-ray catalogs as if it were a science-grade anomaly catalog, must not be used as a tracer of CMB fluctuation statistics, and must not be interpreted as evidence of any astrophysical signal at the highlighted galactic-plane position. The retention of the cross-transfer scan is purely methodological: it documents the empirical failure mode of cross-transferring a 32-latent fully connected autoencoder trained on Planck SMICA data onto the ACT angular-resolution and noise regime, and motivates the architectural choice of the Planck native convolutional autoencoder.

---

[1] DESI Collaboration, "Data Release 1 of the Dark Energy Spectroscopic Instrument," Astron. J. (accepted 2025), arXiv:2503.14745.

[2] LAMOST Collaboration, "LAMOST Data Release 10 (v2.0)," `https://www.lamost.org/dr10/` (2023); survey description: X.-Q. Cui *et al.*, *Research in Astronomy and Astrophysics* **12**, 1197 (2012).

[3] A. Almeida *et al.* (SDSS Collaboration), "The Eighteenth Data Release of the Sloan Digital Sky Survey: Targeting and Spectroscopy," Astrophys. J. Suppl. Ser. **267**, 44 (2023).

[4] A. Merloni *et al.*, "The SRG/eROSITA All-Sky Survey: The first X-ray all-sky survey in the 21st century," Astron. Astrophys. **682**, A34 (2024).

[5] Gaia Collaboration, "Gaia Data Release 3," Astron. Astrophys. **674**, A1 (2023).

[6] A. Mainzer *et al.*, "NEOWISE Reactivation Mission Year Ten," *Planetary Science Journal*, 2024.

[7] Planck Collaboration, "Planck 2018 results. I. Overview and the cosmological legacy of Planck," Astron. Astrophys. **641**, A1 (2020).

[8] Planck Collaboration, "Planck 2018 results. IX. Constraints on primordial non-Gaussianity," Astron. Astrophys. **641**, A9 (2020).

[9] F. J. Qu *et al.* (ACT Collaboration), "The Atacama Cosmology Telescope: A Measurement of the DR6 CMB Lensing Power Spectrum and Its Implications for Structure Growth," Astrophys. J. **962**, 112 (2024), arXiv:2304.05202.

[10] D. Baron and D. Poznanski, "The weirdest SDSS galaxies: results from an outlier detection algorithm," Mon. Not. Roy. Astron. Soc. **465**, 4530 (2017).

[11] Y. Liang *et al.*, "Outlier Detection in the DESI Bright Galaxy Survey," Astrophys. J. Lett. **956**, L6 (2023), arXiv:2307.07664.

[12] C. Nicolaou *et al.*, "Identifying Anomalous DESI Galaxy Spectra with a Variational Autoencoder," Mon. Not. Roy. Astron. Soc. **547**, Issue 2 (2026), arXiv:2506.17376.

[13] D. Wands, "Local non-Gaussianity from inflation," Class. Quant. Grav. **27**, 124002 (2010).

[14] Y.-F. Cai, W. Xue, R. Brandenberger, and X. Zhang, "Non-Gaussianity in a matter bounce," J. Cosmol. Astropart. Phys. **0905**, 011 (2009), arXiv:0903.0631.

[15] O. Doré *et al.* (SPHEREx Collaboration), "Cosmology with the SPHEREx All-Sky Spectral Survey," arXiv:1412.4872 (2014).

[16] U. Seljak, "Extracting Primordial Non-Gaussianity without Cosmic Variance," Phys. Rev. Lett. **102**, 021302 (2009).

[17] N. Hamaus, U. Seljak, and V. Desjacques, "Optimal constraints on local primordial non-Gaussianity from the two-point statistics of large-scale structure," Phys. Rev. D **86**, 103513 (2012).

[18] G. Agazie *et al.* (NANOGrav Collaboration), "The NANOGrav 15 yr Data Set: Evidence for a Gravitational-wave Background," Astrophys. J. Lett. **951**, L8 (2023).

[19] J. Quintin, Y. F. Cai, and R. H. Brandenberger, "Matter creation in a nonsingular bouncing cosmology," Phys. Rev. D **90**, 063507 (2014).

[20] Y.-F. Cai, "Exploring bouncing cosmologies with cosmological surveys," Sci. China Phys. Mech. Astron. **57**, 1414 (2014).

[21] A. Sesana, F. Shankar, M. Bernardi, and R. K. Sheth, "Selection bias in dynamically measured supermassive black hole samples," Mon. Not. Roy. Astron. Soc. **463**, L6 (2016).

[22] S. Burke-Spolaor *et al.*, "The astrophysics of nanohertz gravitational waves," Astron. Astrophys. Rev. **27**, 5 (2019).

[23] R. Trotta, "Bayes in the sky: Bayesian inference and model selection in cosmology," Contemp. Phys. **49**, 71 (2008), arXiv:0803.4089.

[24] L. Verde, P. Protopapas, and R. Jimenez, "Planck and the local universe: Quantifying the tension," Phys. Dark Univ. **2**, 166 (2013), arXiv:1306.6766.

[25] R. W. Hellings and G. S. Downs, "Upper limits on the isotropic gravitational radiation background from pulsar timing analysis," Astrophys. J. Lett. **265**, L39 (1983).

[26] J. Antoniadis *et al.* (EPTA Collaboration), "The second data release from the European Pulsar Timing Array: III. Search for gravitational wave signals," Astron. Astrophys. **678**, A50 (2023), arXiv:2306.16214.

[27] D. J. Reardon *et al.* (PPTA Collaboration), "Search for an isotropic gravitational-wave background with the Parkes Pulsar Timing Array," Astrophys. J. Lett. **951**, L6 (2023), arXiv:2306.16215.

[28] A. Afzal *et al.* (NANOGrav Collaboration), "The NANOGrav 15-year data set: Search for signals from new physics," Astrophys. J. Lett. **951**, L11 (2023), arXiv:2306.16219.

[29] E. S. Phinney, "A practical theorem on gravitational wave backgrounds," arXiv:astro-ph/0108028 (2001).

[30] M. Wenger *et al.*, "The SIMBAD astronomical database," Astron. Astrophys. Suppl. Ser. **143**, 9 (2000).

[31] L. McInnes, J. Healy, and J. Melville, "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction," arXiv:1802.03426 (2018).

[32] L. McInnes, J. Healy, and S. Astels, "hdbscan: Hierarchical density based clustering," J. Open Source Softw. **2**, 205 (2017).

[33] C. Heinrich, O. Doré, and E. Krause, "Measuring $f_{\rm NL}$ with the SPHEREx multitracer redshift-space bispectrum," Phys. Rev. D **109**, 123511 (2024), arXiv:2311.13082 [astro-ph.CO].

[34] M. Münchmeyer, M. S. Madhavacheril, S. Ferraro, M. C. Johnson, and K. M. Smith, "Constraining local non-Gaussianities with kinetic Sunyaev-Zel'dovich tomography," Phys. Rev. D **100**, 083508 (2019), arXiv:1810.13424.

[35] E. Wilson-Ewing, "The Matter Bounce Scenario in Loop Quantum Cosmology," JCAP **1303**, 026 (2013), arXiv:1211.6269.

[36] L. Lentati *et al.*, "Hyper-efficient model-independent Bayesian method for the analysis of pulsar timing data," Phys. Rev. D **87**, 104021 (2013), arXiv:1210.3578.

[37] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, "`emcee`: The MCMC Hammer," Publ. Astron. Soc. Pac. **125**, 306–312 (2013), arXiv:1202.3665 [astro-ph.IM].

[38] J. Yoo, A. L. Fitzpatrick, and M. Zaldarriaga, "A New Perspective on Galaxy Clustering as a Cosmological

Probe: General Relativistic Effects," Phys. Rev. D **80**, 083514 (2009), arXiv:0907.0707 [astro-ph.CO].

[39] C. Bonvin and R. Durrer, "What galaxy surveys really measure," Phys. Rev. D **84**, 063505 (2011), arXiv:1105.5280 [astro-ph.CO].

[40] A. Challinor and A. Lewis, "Linear power spectrum of observed source number counts," Phys. Rev. D **84**, 043516 (2011), arXiv:1105.5292 [astro-ph.CO].

[41] E. Di Dio, F. Montanari, J. Lesgourgues, and R. Durrer, "The CLASSgal code for relativistic cosmological large scale structure," J. Cosmol. Astropart. Phys. **11**, 044 (2013), arXiv:1307.1459 [astro-ph.CO].