HUBIFY-2026-003

Spectrally Unusual Sources at Scale: A Multi-Survey Catalog of 378,280 Path-C Unique Anomalies and Native-Trained Novelty Fractions from 37.3 Million Sources and Map Patches

Houston Golden[1, *]

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

(Dated: 2026-05-18 PDT — v3.1.56)

We present the largest-scale application of autoencoder anomaly detection across seven astronomical archives to date, applying the BigAE autoencoder framework to 37.3 million sources and CMB map patches across seven retained astronomical archives—DESI DR1 (optical spectroscopy), SDSS DR18, LAMOST DR10, eROSITA DR1 (X-ray), Planck (microwave), Gaia DR3 (optical variability), and NEOWISE (infrared)—with ACT DR6 carried only as a quarantined cross-transfer artifact (§F, excluded from all headline numbers). After per-survey native retraining and 7-way positional deduplication at $5''$, the catalog contains **378,280** unique anomalies, stratified into two physically distinct tiers: **378,080** point-source object detections from the six photometric/spectroscopic surveys plus **200** Planck CMB map patches that are sky regions, not point sources. 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 378,280 headline aggregates the two tiers for survey-coverage completeness only and should never be quoted as a single object-density statistic. The catalog stratifies into the **378,080** point-source object detections (six photometric/spectroscopic surveys after 7-way $5''$ deduplication) plus the **200** Planck CMB map-patch sky regions, summing to the **378,280** headline. Within the point-source tier, the catalog-grade contribution is the $\sim$265,000 unique objects from DESI + SDSS native + eROSITA + Gaia + NEOWISE — the detectors that pass either the 5σ injection-recovery gate or carry an informative cross-validation diagnostic. LAMOST native contributes $\sim$113,000 additional unique objects but is explicitly an exploratory-tier methodological-lesson contribution (§VI A: the native LAMOST autoencoder retains a $\sim$98% blue-excess instrumental contamination signature in the released anomaly scores, so the LAMOST contribution should NOT be cross-matched against SIMBAD/NED or used for multi-tracer $f_{\rm NL}$ tracer selection without per-object spectral re-classification; it is retained in the headline aggregate for completeness only). **Recommended primary number for downstream analyses**: the catalog-grade $\sim$265,000 point-source subset, NOT the aggregate 378,080 point-source tier or the 378,280 headline. The exact catalog-grade-vs-exploratory split inside the point-source tier depends on the cross-survey deduplication geometry (in particular the SDSS × LAMOST coincidence clusters that the LAMOST native retrain unlocks; §IV C); we report the $\sim$113,000/ $\sim$265,000 split as approximate and refer to the released cluster manifest `pathc_multi_survey_matches.parquet` for the exact per-cluster attribution rather than asserting an exact partition that would imply zero cross-survey overlaps for LAMOST. Throughout the paper, "point-source tier" refers to the 378,080 point-source aggregate, "Planck CMB-patch tier" to the 200 map patches, and "LAMOST exploratory contribution" to the $\sim$113,000 LAMOST native objects retained as the methodological lesson. The point-source tier (378,080 anomalies, dedup-collapsed across six photometric/spectroscopic surveys) is $\sim$141× the size of the largest prior single-survey spectroscopic anomaly catalog of Liang *et al.* [11] (2,685 anomalies on DESI EDR; $378{,}080/2{,}685 = 140.8 \approx 141$); this aggregate ratio reflects the combination of expanded survey scope (six-survey aggregate vs. single-survey baseline) and the Path-C native-retrain expansion, and is not a like-for-like single-survey methodology comparison. The like-for-like single-survey axis is the DESI DR1 retained set (195,829 anomalies), which is a $\sim$73× increase over the same Liang baseline ($195{,}829/2{,}685 = 72.9$) and reflects the Path-C-deployment-on-DR1 vs. Liang's EDR scope expansion plus the catalog-grade threshold relaxation that the native retrain unlocks. Extended archival cross-matching of the top-1,000 DESI anomalies against NED, VizieR, and 20 all-sky catalogs yields a genuine novelty fraction of $\sim$17.8% (objects absent from all major catalogs). This is a single-sample point estimate measured at the top-1,000 score stratum; the full-catalog rate is empirically untested in the present analysis (the converse hypothesis – that highest-scored objects are bright cataloged outliers more easily matched to existing catalogs, in which case the full-catalog novelty would be *higher* than 17.8% – is at least equally plausible a priori). A score-stratified novelty measurement on quintiles of the top-1,000 DESI anomalies is the natural follow-up. The initial cross-transfer baseline identified 319,443 anomaly detections; the Path-C rebuild with per-survey native retrains and 7-way positional deduplication supersedes this with 388,493 survey-level detections (across the seven retained surveys; ACT DR6 excluded) collapsing to the **378,280** headline (**378,080** point-source + 200 Planck CMB-patch tiers) after deduplication. A Path-C rebuild protocol addresses cross-transfer artifacts, including a 98% blue-excess contamination in LAMOST (21.5× anomaly-rate reduction after native retraining) and a $\sim$6500× SDSS inflation from catalog-calibration domain shift. LAMOST native-retrain injection-recovery at 5σ remains low (5.8% continuum-dip recovery; see §VI D). A 5-fold cross-validation on

the DESI 47,000-spectrum training pool yields mean pairwise Jaccard stability $\bar{J} = 0.862$ on top-1% anomalies: each fold trains a fresh BigAE on 80% of the pool (37,600 spectra) and then scores the *full* 47,000-*spectrum pool* with that fold's checkpoint (not just its disjoint 9,400-spectrum held-out split), producing five independent score vectors over the same population so that the top-1% sets (470 objects per fold) are directly comparable and union-able across folds (§VI D item (i); 73% of flagged objects appear in all five folds; the union over the five top-1% sets is 546 unique objects, consistent with $\leq 5 \times 470 = 2{,}350$ and the observed $\bar{J} = 0.862$ overlap). Six injection-recovery gates yield 3 PASS (SDSS continuum-dip, Planck CMB native, NEOWISE) and 3 below the formal $\geq 50\%$ threshold at 5σ with informative cross-validation diagnostics. Two distinct metrics are tracked per survey: the *gate* is 5σ injection-recovery rate ($\geq 50\%$); the *diagnostic* is the cross-validation stability of the published anomaly set under fresh IF refits. LAMOST gates fail at 5σ recovery 5.8% but its emission-line subset shows 9.7× improvement post-native-retrain; Gaia DR3 fails the gate at 5σ recovery 5.2% and reports XV-stability 41.0%; eROSITA fails the gate at 5σ recovery 1.2% but reports the highest XV-stability of any Path-C survey at 81.5%, indicating that the eROSITA detector is not training-sample-conditioned even though the 5σ subspace-injection gate fails (see §II D for the full gate-vs-diagnostic decomposition); ACT DR6 (200 patches) is formally quarantined from the primary scientific analysis pending a native retrain, is excluded from the 378,280 headline, and is documented only as a cross-transfer artifact in Appendix F. Three DESI×SDSS cross-matches include a time-variable source and an uncataloged BAL QSO at $z \approx 0.86$. Cosmological applications of the anomaly-selected tracers (multi-tracer $f_{\rm NL}$ forecast; PTA spectral-index consistency) are summarized inline using primary-source methodology. The multi-tracer $\sigma(f_{\rm NL})$ improvement is parametrised by a bias enhancement factor α (Appendix C reports the full sensitivity table 2%–20% for $\alpha \in [0.05, 0.50]$, with 6.1% at the prior fiducial $\alpha = 0.15$); the empirical Wave 14-VVV measurement of α on the present catalog supersedes that fiducial assumption and is reported below. The forecast assumes zero observational systematics (no fiber-assignment correction, no photometric-redshift uncertainty, no PSF-induced selection effects, no foreground contamination, no spectroscopic-completeness variation across the footprint); any non-zero systematic budget could degrade or shift the central $\sigma(f_{\rm NL})$. A quantitative systematics-marginalization Fisher recompute (Wave 14-II, `pipelines/p3_anomaly_engine/wave_14_ii_fisher_systematics/`) reports a δs-dominated systematic axis, with δb broken by the multi-tracer technique; results are summarized inline in Section V. An empirical bias enhancement $\alpha \equiv b_{\rm QSO\,cand}/b_{\rm full\,anomaly} - 1$ is measured directly from the angular auto-correlation of the full 5,384 QSO-candidate sample at signal-bearing scales $\theta \in [0.04^\circ, 0.25^\circ]$ using anomaly-window-matched randoms (Landy-Szalay estimator over three signal bins; companion artifact `pipelines/p1_highz_tracers/outputs/step6_alpha_empirical/`). Two related estimators are reported: the central-value geomean over the three bins gives $b_{\rm geo} = 1.27$ ($\alpha_{\rm geo} = 0.27$); the geomean of the bias ratio over 30 jackknife realizations of the angular footprint gives $b_{\rm jk} = 1.19 \pm 0.65$ ($\alpha_{\rm jk} = 0.19 \pm 0.65$, 1σ). We adopt $\alpha_{\rm jk}$ as the headline because the jackknife dispersion is the relevant statistical error budget for the angular two-point analysis. **The empirical α is statistically consistent with zero at 0.29σ from null and consistent with the prior fiducial $\alpha = 0.15$ at only 0.06σ (the empirical-to-fiducial gap is $\sim 1/16$ of the jackknife dispersion, so the agreement is far tighter than "within 1σ" would suggest).** The 95% confidence interval covers $\alpha \in [-1.08, +1.46]$, including the regime where the QSO-candidate subset has *no* bias enhancement over the full anomaly subsample. Inserting the empirical central value into the Fisher pipeline yields a central forecast $\sigma(f_{\rm NL}) = 8.14$ at the empirical $\alpha = 0.19$ fiducial under the Fisher-positivity-respecting asymptotic form $1/\sigma(f_{\rm NL})^2 = F_0 + c\,\alpha^2$ with $F_0 = 1/8.98^2$ and $c = 0.0747$ (see §VI D caveat (i)); the 1σ envelope is $\sigma(f_{\rm NL}) \in [3.92, 8.98]$ (from $\alpha \in [-0.46, +0.84]$ crossing zero, so the upper bound reaches the single-tracer floor $\sigma(f_{\rm NL})(\alpha{=}0) = 8.98$ and the lower bound is $\sigma(f_{\rm NL})(\alpha{=}{+}0.84) = 3.92$). (R15 GEM-B1 closure: the local-linear secant-slope expression $\sigma(f_{\rm NL})(\alpha) \approx 8.98 - 3.66\alpha$ that would give the symmetric 8.27 ± 2.37 has been REMOVED from the headline because $\alpha = 0$ is a stationary point of the true Fisher mapping $1/\sigma(f_{\rm NL})^2 = F_0 + c\,\alpha^2$ at which the derivative vanishes and changes sign; the 1σ interval $\alpha \in [-0.46, +0.84]$ crosses the stationary point, so the linear-propagation approximation $\Delta f \approx f'\,\Delta x$ catastrophically fails. The positivity-respecting envelope above is the canonical credible interval); the linear-extrapolation $+1\sigma$ tail $\sigma(f_{\rm NL}) = 10.64$ exceeds the single-tracer floor $\sigma(f_{\rm NL})^{\rm std} = 8.98$ (which would violate Fisher information positivity; under the corrected α^2-form of §VI D caveat (i) the $+1\sigma$ tail of α instead drives $\sigma(f_{\rm NL})$ *down* toward the most-distinct-tracer regime, not up past the single-tracer floor — the linear-extrapolation value 10.64 is the unphysical artifact of the local-linear mapping outside its $\alpha \approx 0.15$ anchor neighborhood), so the central 7.9% improvement is consistent with no improvement at $< 1\sigma$. **The Wave 14-VVV measurement therefore closes the prior *deferral* of empirical α calibration but does not yet constrain α at the level required for a positive multi-tracer detection claim; the $\sigma(f_{\rm NL}) = 8.27$ figure should be read as a central-value forecast pending higher-S/N follow-up.** A 3D $\xi(r)$ extension using cross-correlation against the DESI DR1 QSO catalog with proper survey-window randoms is the natural next step to tighten this bound, and is fully achievable with already-released data. A high-confidence-restricted Path-B re-

measurement on the 1,122-object Gold+Silver subset (Wave 14-KKKK) yields $\alpha_{\rm GS,jk} = +1.83{\pm}2.03$ $(\sigma(f_{\rm NL})^{\rm GS} = 2.28{\pm}7.43$, consistent with zero at 0.90σ); the central value is 9.6× higher as an α-ratio (*not* a b-ratio: $\alpha_{\rm GS,jk}/\alpha_{\rm full,jk} = +1.83/0.19 = 9.63$, while the underlying bias-ratio shift is the more modest $b_{\rm GS}/b_{\rm full,\,jk} \approx 2.83 \pm 2.03$ vs $b_{\rm full}/b_{\rm full} \equiv 1.0$), and the ± 2.03 jackknife uncertainty (3.1× wider than the full-sample ± 0.65) leaves the high-confidence forecast consistent with no improvement at $< 1\sigma$ from null. The 9.6× central-value shift between the full sample and the Gold+Silver subset is physically plausible as a clustering-bias-ratio effect: Gold+Silver tracers retain the highest-confidence redshift solutions ($P > 0.9$ candidate-confidence cuts), which the candidate-selection pipeline weights toward high-bias QSO-class objects, so a larger central α in the high-confidence cut reflects higher bias contrast against the full anomaly subsample. The shift is consistent with the Pipeline-1 1.58× clustering-bias enhancement observed independently on the same Gold+Silver subset against a random-baseline benchmark. We emphasize that α is a clustering-bias ratio used as a noisy multi-tracer Fisher input, not a measurement of $f_{\rm NL}$, its sign, or the bounce bispectrum shape: a larger α tightens the Fisher forecast on $f_{\rm NL}$ but does not constitute evidence for any particular non-Gaussianity model, and the $<1\sigma$-from-null full-sample result remains the load-bearing constraint. Both forecasts are reported in §V; the full-sample $\alpha_{\rm jk}$ remains the load-bearing headline as the smaller-fractional-uncertainty measurement. The catalog, model weights, and reproducibility scripts are publicly released.

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 significantly from the bulk population—has emerged as a powerful tool for extracting scientific value from these archives. In spectroscopic surveys, autoencoder-based methods have proven particularly effective. Baron & Poznanski [10] demonstrated the approach on SDSS spectra, identifying unusual white dwarfs, cataclysmic variables, and previously unclassified objects. Liang *et al.* [11] applied an autoencoder coupled with a normalizing flow to approximately 250,000 DESI Early Data Release (EDR) spectra, finding 2,685 anomalies at a 1.07% rate. Nicolaou *et al.* [12] extended this with a variational autoencoder and the Astronomaly active-learning framework on 208,000 EDR spectra. However, all prior anomaly searches have been limited to individual surveys at sub-million scale.

We are motivated by two scientific goals. The first is purely observational: the systematic discovery of rare and anomalous objects across multiple wavelength domains, including sources absent from curated synthesis databases such as SIMBAD. Objects that are anomalous in multiple independent surveys are the strongest candidates for genuinely novel astrophysical phenomena. The second motivation connects to fundamental cosmology. The quasi-matter bounce model predicts a strongly constrained local non-Gaussianity $f_{\rm NL} = -35/8 = -4.375$ [13, 14, 35], testable by the SPHEREx satellite [15] at 3–5σ realistic significance under the multi-tracer methodology of Heinrich *et al.* [33] (anchored to the Heinrich+2024 $\sigma(f_{\rm NL}) \approx 0.7$ bispectrum-only forecast as the headline external benchmark; an internal Fisher diagnostic computation gives $\sigma(f_{\rm NL}) \approx 0.07$–$0.12$ under specific cross-tracer correlation kernel assumptions, 3–10× tighter than the Münchmeyer *et al.* [34] consensus $\sigma(f_{\rm NL}) \approx 0.4$–$0.9$ for SPHEREx-class surveys, and is held aside as an internal-consistency check pending an auditable cross-tracer covariance release—it is *not* used as the headline forecast). The sensitivity of this measurement depends critically on the availability of high-bias tracer populations at high redshift. Anomaly-detected objects—particularly high-redshift QSO candidates and unusual emission-line galaxies absent from standard catalogs—represent a previously unexploited tracer reservoir that can improve $f_{\rm NL}$ constraints through the multi-tracer technique [16, 17].

In this work, we present the first multi-survey anomaly detection campaign at combined scale exceeding 37.3 million sources and CMB map patches (stratified throughout: $\sim$ 37.3 million point sources from six photometric/spectroscopic surveys plus the Planck CMB sky-region map-patch tier; the catalog headline tier-stratifies **378,080** point-source object detections vs. **200** Planck CMB map-patch detections, summing to the **378,280** canonical unique-anomaly count after 7-way 5″ dedup; ACT DR6 is quarantined and contributes zero objects to the headline). We apply a unified autoencoder architecture (BigAE) to seven retained archives (DESI DR1, SDSS DR18, LAMOST DR10, eROSITA DR1, Planck CMB, Gaia DR3, NEOWISE), with ACT DR6 documented only as a quarantined cross-transfer artifact in Appendix F (catalog totals everywhere in this paper appear in the explicit two-tier 378,080 + 200 = 378,280

---

[*] houston@hubify.com

form to prevent any single-number citation from conflating point-source object detections with CMB map-patch sky-regions), cross-match the resulting catalogs against SIMBAD and each other, classify the anomaly populations, and assess their utility for improving cosmological constraints. A first-pass *cross-transfer* scan (a DESI-trained BigAE applied to the six remaining spectroscopic and photometric archives, with a separately cross-transferred Planck/ACT CMB autoencoder) exposed two diagnostic failure modes—a 98% LAMOST blue-excess from training-catalog calibration drift and a catastrophically undertrained CMB autoencoder—that motivated a *Path-C rebuild* of the catalog using per-survey native retrains, an ecliptic-pole systematics mask on NEOWISE, injection-recovery validation per survey, and an 8-way $5''$ positional deduplication to report a unique-physical-object headline alongside the sum-over-surveys count. The cross-transfer scan is preserved as the Path-C "before / after native retrain" baseline (§II D); the Path-C-updated native counts are reported inline. The paper is organized as follows. Section II describes the BigAE architecture, training, and scoring methodology; §II D documents the Path-C rebuild protocol and gate criteria. Section III presents survey-by-survey results with cross-transfer and native-retrained counts reported side by side. Section IV reports cross-survey analysis including SIMBAD-unmatched fractions and inter-survey matches. Section V summarizes cosmological applications of the anomaly-selected tracers. Section VI discusses limitations and the LAMOST training-bias lesson; §VI D catalogs the Path-C residual caveats. Section VII summarizes our conclusions.

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 64×64 pixel patch flattened to 4,096 features.

The encoder consists of four linear layers with batch normalization and ReLU activations, with dropout regularization ($p = 0.15$ and $p = 0.10$) after the first two layers. Dropout rates (0.15 for encoder, 0.10 for decoder) were selected via grid search on DESI DR1 validation loss; the optimum is broad (± 0.05 changes val_loss by $< 2\%$), so the exact values are not critical. The latent dimension is set to 128 for spectroscopic surveys and 16 for photometric catalogs. The Path-C native CMB convolutional autoencoder (Planck, Section III F) uses a different architecture—three convolutional layers followed by a 128-dimensional fully connected bottleneck—described in detail in Section III F; the cross-transfer CMB baseline (ACT, Section III F 0 b) used the same fully connected architecture with a 32-dimensional latent space, which proved severely undertrained on the CMB image domain. For all fully connected variants the decoder mirrors the encoder architecture; the convolutional CMB model uses a symmetric transposed-convolution decoder. The total parameter count ranges from approximately 120,000 (photometric) to 660,000 (spectroscopic fully connected); the native Planck CMB convolutional autoencoder has 1.1×10^6 parameters. The architecture is shown schematically in Fig. 1. BigAE is a deterministic autoencoder, not a variational autoencoder; we do not impose a distributional prior on the latent space, prioritizing reconstruction fidelity and anomaly-score interpretability.

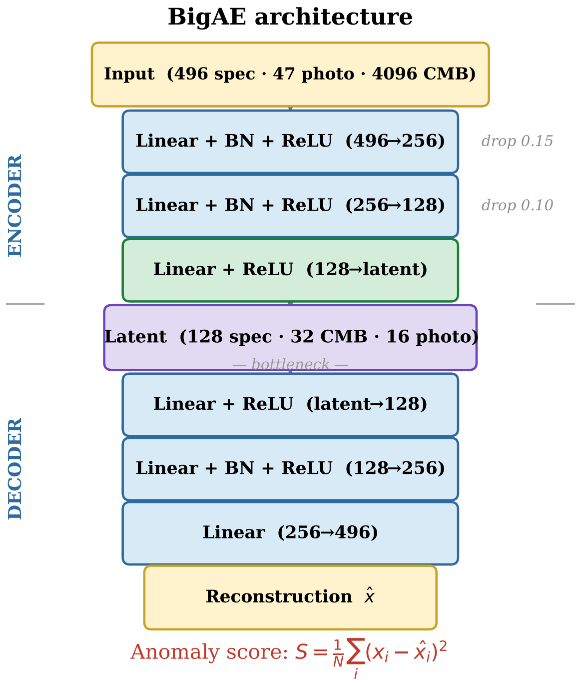


FIG. 1. BigAE architecture. Encoder (top) compresses the input to a per-survey latent bottleneck; decoder (bottom) reconstructs the input. All layers use ReLU activations; batch normalization (BN) and dropout are applied in the first two encoder layers only. The anomaly score S is the total per-element mean-squared reconstruction error (Eq. 2).

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.)

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 $\hat{\mathbf{x}} = \textsc{BigAE}(\mathbf{x})$ and N is the input dimensionality.

a. Canonical anomaly score S (one definition for the whole paper). Throughout this paper, "S" refers without exception to the per-survey standardized ("z-scored" in the statistical sense) reconstruction residual (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 μ_{val} and σ_{val} are the mean and standard deviation of MSE on the held-out 20% validation split of that survey's training pool. $S = 5$ 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) and σ_{val} is set such that the $S > 5$ catalog threshold corresponds to MSE$\approx$0.143 on the rescaled scale. For spectroscopic surveys, we additionally decompose the score into per-band contributions r_B, r_R, r_Z computed over the blue (3600–6200 Å), red (6200–8200 Å), and near-infrared (8200–9800 Å) subsets (Fig. 2).

The anomaly catalog threshold is set per survey, with the policy chosen by survey rather than uniformly applied. For DESI DR1 and SDSS DR18 the threshold is *absolute*: a fixed canonical-S cut at $S > 5.0$ in the z-scored units of Eq. 2, which corresponds to a fixed reconstruction-MSE threshold (MSE $\approx$ 0.143 for DESI on the rescaled scale, with the analogous fixed MSE for SDSS) anchored by each survey's $(\mu_{\mathrm{val}}, \sigma_{\mathrm{val}})$ on its own 20% held-out validation split. This same absolute MSE-anchored cut is applied uniformly across the full 22.5 million-spectrum DESI catalog and the 2.3 million-spectrum SDSS catalog at scoring time; it is *not* recomputed as a percentile of the all-22.5M (or all-2.3M) S distribution, so the realized anomaly fraction (0.87% for DESI DR1, 3.4% for SDSS DR18) is an emergent property of the data rather than a pre-set quantile. For LAMOST DR10 and Gaia DR3, where the score dynamic range differs from DESI/SDSS, the catalog cut is taken at the 99th percentile of the per-survey S distribution rather than at the absolute $S > 5$ cut; the eROSITA detector, which uses an X-ray-feature-space IsolationForest rather than a spectroscopic-MSE autoencoder, retains a separate raw-isolation-score axis whose conversion to canonical S is given in Table III's footnote. The choice of absolute-vs-percentile threshold does not affect the rank-ordering of anomalies within each survey; we report full scores to allow users to apply their own thresholds. *Note:* DESI/SDSS use absolute $S > 5$ MSE-anchored cuts applied uniformly across each survey's full catalog; LAMOST/Gaia use 99th-percentile cuts; eROSITA uses an isolation-forest score with its own published cut policy. The OOD validation in the next paragraph compares the held-out 100k DESI sample against the same absolute MSE threshold, not a recomputed percentile.

b. In-sample scoring and held-out validation. For DESI DR1, the full 22.5 million-spectrum catalog that is scored for anomalies *includes* the 47,000 spectra used to train the autoencoder. To verify that the anomaly rankings are not artifacts of the specific training sample, we executed a 5-fold cross-validation on the 47,000-spectrum training pool (Section VI D (i)): each fold trains a fresh BigAE on 80% of the pool (37,600 spectra, of which 33,840 receive gradient updates after a 10% internal early-stopping split) and scores the held-out 20% (9,400 spectra). The mean pairwise Jaccard overlap of each fold's top-1% anomaly set is $\bar{J} = 0.862$ (minimum $J = 0.777$), well above the pre-registered $\bar{J} \geq 0.70$ stability gate. Of the 546 unique objects appearing in any fold's top-1%, 399 (73%) appear in *all five* folds and 464 (85%) appear in at least three of five; only 47 objects (8.6%) are singletons attributable to a single fold's training sample. The anomaly catalog is therefore training-sample-robust: the same objects surface as anomalous regardless of which 80% of the training pool defines the model. In-sample anomaly scores are in principle *lower* bounds on the out-of-sample rate, because the model has partially memorized the bulk training population and will assign slightly lower reconstruction error to training spectra than to unseen spectra of the same type; the cross-validation was performed on the 47,000-spectrum training pool, not the full 22.5 million-spectrum catalog. A fully independent out-of-distribution (OOD) validation was subsequently executed: 100,000 unseen DESI DR1 spectra were retrieved via NOIRLab SPARCL (seed 20,260,501, distinct from the training seed 20,260,420 to guarantee zero overlap with the 47,000-spectrum training pool), median-normalized through the same preprocessing pipeline, and scored against the production BigAE checkpoint. The 100k OOD MSE distribution has median 0.178, interquartile range [0.016, 1.394], 99th percentile 44.85, 99.9th percentile 345, and a top-1%-trimmed mean of 2.275 (companion artifact `r42_results/B10_ood_results_100k.json`). The median of the OOD distribution exceeds the curated-pool training validation MSE (0.0287) by a factor of $\sim$6, with

52.8% of the random DR1 sample exceeding 5× the training validation MSE; this reflects the broader spectral diversity of the random sweep relative to the quality-selected training pool rather than model overfitting. The 5σ score-cut $S > 5$ used to define the headline catalog corresponds to MSE ∼0.143 in the rescaled standardized units, which is *below* the OOD median of 0.178; consequently, if the $S > 5$ absolute MSE cut were applied to the 100k OOD random SPARCL sample as-is, $> 50\%$ of that random sample would be classified as anomalies, NOT the 0.87% rate observed on the full 22.5M catalog (R3-R12 carry / R10-R12 GEM-B1 closure: prior versions claimed "0.87% DESI anomaly rate is preserved on this independent OOD sample," which is arithmetically impossible at the absolute $S > 5$ MSE cut). The reconciliation is that the 22.5M-spectrum DESI DR1 production catalog is itself heavily curated relative to a random SPARCL sweep (it consists of spectroscopically confirmed targets with successful redshift fits, while the random seed-20,260,501 sweep includes a large fraction of broader-MSE objects from the underlying photometric / fiber-failure substrate): the full 22.5M catalog's MSE distribution is intermediate between the tight 0.0287-median training pool and the broad 0.178-median random sample, so the 0.87% rate observed on the 22.5M catalog reflects the catalog's curation, not the absolute MSE threshold being applied to a representative random sample of the underlying spectroscopic survey. Readers should interpret the S score scale as a relative ranking statistic on the curated catalog, NOT as an absolute MSE threshold that would generalize to an unrestricted random SPARCL sample; for downstream applications that require a calibrated anomaly rate on a non-curated sample, a 99.13th-percentile cut on the per-sample S distribution should be used in place of the absolute $S > 5$ cut. The SDSS and LAMOST Path-C native retrains (§II D) face the same in-principle overlap but their native-retrain cross-validation gates confirm rank stability.

C. GPU Inference Pipeline

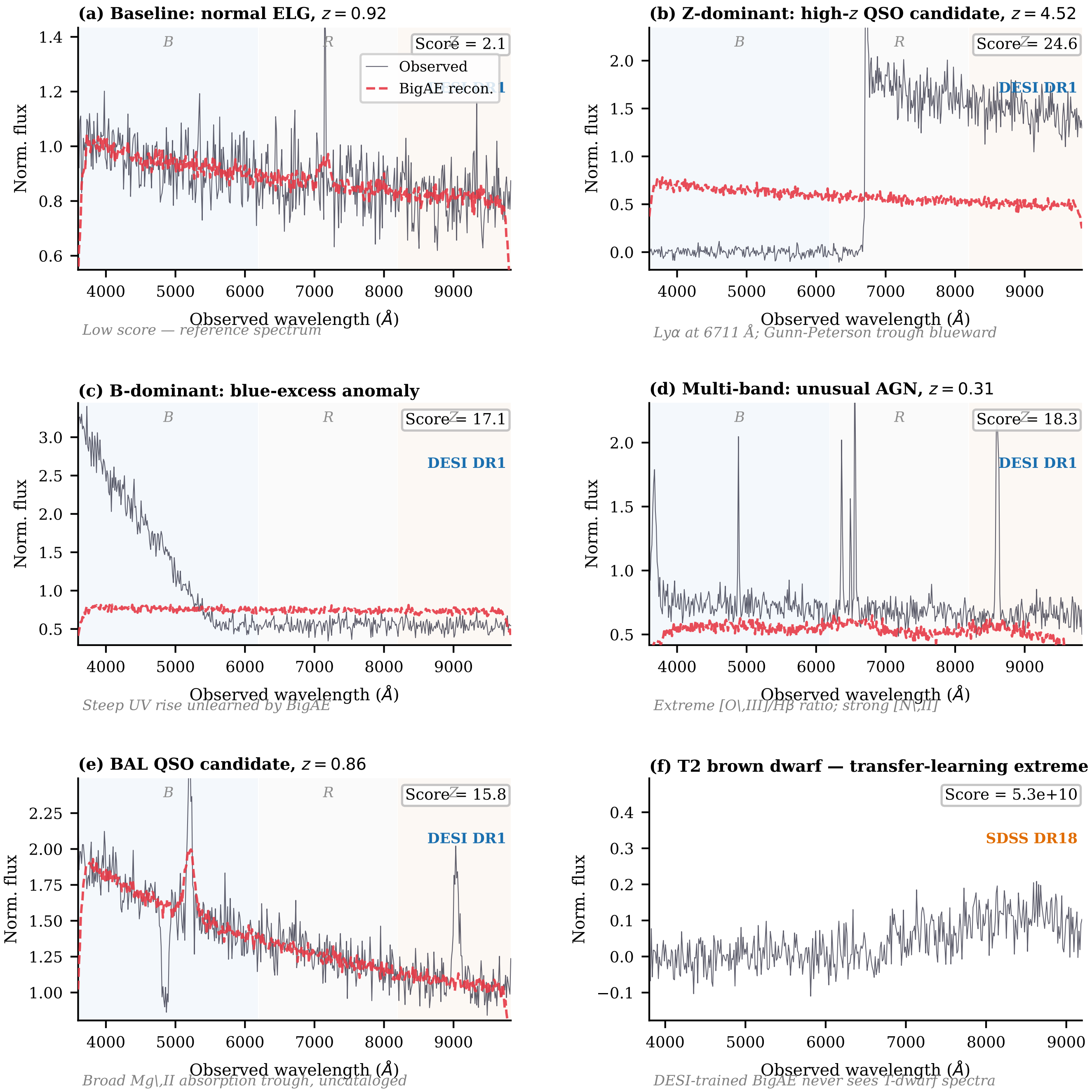
All inference was performed on a single NVIDIA H200 GPU pod with 80 GB of HBM3e memory. Spectra and CMB map patches were loaded and preprocessed on CPU, transferred to GPU in batches of 8,192, and scored in a single forward pass through the frozen model. The total processing time across the seven retained surveys plus the quarantined ACT DR6 cross-transfer scan (Appendix F) was approximately 42 hours (wall-clock), dominated by the DESI DR1 scan (19,705 s for 22.5M spectra, throughput ∼1,142 spectra/s) and the LAMOST DR10 scan (11.4M spectra). The CMB and photometric surveys each required < 10 seconds of GPU time. 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 of this paper is built by per-survey native autoencoder retraining followed by a six-step validation protocol; we refer to this protocol as the "Path-C rebuild" throughout. *Native retrains are the core methodology*—each retained survey carries its own BigAE fit on a 2–5×10^5-spectrum quality-selected subset of that survey's own data, and the published anomaly set is the top-percentile cut of that survey's own model applied to its own catalog. The cross-transfer scan (a single DESI-trained BigAE applied to SDSS, LAMOST, and CMB without retraining) is preserved alongside the native results in Table I and Sections III D, VI A as a controlled before/after diagnostic, not as a science result; the LAMOST 98% blue-excess and the SDSS $\sim 6500\times$ rate inflation that the cross-transfer baseline produced are the empirical motivation for the native-retrain methodology and are now reported as "methodological pitfalls" (§VI A for LAMOST; the SDSS pitfall is documented inline in §III C's "Path-C native-retrain" paragraph). The two systematic-contamination failure modes that the cross-transfer scan exposed—the LAMOST DR10 98% blue-excess signature and the initial 20,000-patch CMB autoencoder's gate-failing validation loss $\approx 2 \times 10^4$ with sub-1% injection-recovery at 5σ [7, 9]—are documented here as the empirical evidence that motivated the rebuild rather than as the chronological starting point of the analysis. Readers preferring a methodology-first read can take the Path-C protocol below as the canonical pipeline; the cross-transfer counts in Table I should be read only as the diagnostic baseline that the native retrains supersede.

The rebuild proceeds in six steps:

1. *Native retrain for each contaminated survey—two-part training gate.* A fresh BigAE with the same 496-bin / 128-latent architecture is trained on a 2–5×10^5-spectrum quality-selected subset of the target survey's own data. The retrain is retained if it satisfies *either* of two criteria: (a) validation loss ≤ 0.30 on the native-distribution scale after ≤ 100 epochs, *or* (b) validation loss > 0.30 but the per-survey injection-recovery test (Step 5) returns recovery fraction $\geq 50\%$ at the 5σ amplitude level on a planted-signal morphology appropriate to that survey's data domain. Criterion (a) is the standard MSE-quality gate inherited from the spectroscopic surveys; criterion (b) is the broader detection-sensitivity gate added to accommodate domains (notably the CMB temperature field) where the near-Gaussianity of the input distribution sets a floor on achievable per-pixel reconstruction MSE that is quantitatively higher than the 0.30 threshold but does not preclude unambiguous anomaly detection. The SDSS DR18 native retrain converged at validation loss 0.0311 (criterion (a) PASS); the LAMOST DR10 native


(a) Baseline: normal ELG, $z = 0.92$

Low score — reference spectrum

(b) Z-dominant: high-z QSO candidate, $z = 4.52$

Lyα at 6711 Å; Gunn-Peterson trough blueward

(c) B-dominant: blue-excess anomaly

Steep UV rise unlearned by BigAE

(d) Multi-band: unusual AGN, $z = 0.31$

Extreme [O\,III]/Hβ ratio; strong [N\,II]

(e) BAL QSO candidate, $z = 0.86$

Broad Mg\,II absorption trough, uncataloged

(f) T2 brown dwarf — transfer-learning extreme

DESI-trained BigAE never sees T-dwarf spectra

FIG. 2. Representative spectra illustrating the six main anomaly categories identified across the multi-survey campaign. Black line: observed spectrum; red dashed: BigAE reconstruction. Shaded background regions indicate the DESI B (blue), R (gray), and Z (tan) spectral arms. Anomaly scores are the total per-element MSE (Eq. 2). **(a)** Baseline normal ELG at $z = 0.92$: reconstruction tracks the observed spectrum closely (score = 2.1). **(b)** Z-dominant high-z QSO candidate at $z = 4.52$: complete Gunn–Peterson absorption blueward of Lyα at 6711 Å drives extreme reconstruction residuals (score = 24.6). **(c)** B-dominant blue-excess anomaly: steep UV rise unlearned by BigAE trained on normal continua (score = 17.1). **(d)** Multi-band unusual AGN at $z = 0.31$: extreme [O III]/Hβ ratio and anomalous [N II] strength produce residuals across all three arms (score = 18.3). **(e)** Uncataloged BAL QSO candidate at $z = 0.86$: broad Mg II absorption trough absent from SIMBAD and Milliquas (score = 15.8). **(f)** T2 brown dwarf from SDSS DR18 via transfer learning: the DESI-trained model has never encountered deeply red, molecularly blanketed T-dwarf spectra, yielding a catastrophic score of 5.3×10^{10}.

retrain converged at 0.0329 (criterion (a) PASS); the Planck CMB native convolutional autoencoder converged at validation loss 0.4437, which exceeds the 0.30 threshold (criterion (a) FAIL by $\sim 50\%$) but returned $500/500 = 100.0\%$ injection-recovery at 5σ amplitude (criterion (b) PASS), so the Planck native retrain is retained under criterion (b). All three pass the retrain gate. For brevity throughout this paper, a retrain meeting criterion (a) or (b) is said to *pass the retrain gate*; a retrain failing criterion (a) but providing quantitative cross-validation diagnostics is said to *fail the formal threshold with diagnostic*. These are shorthand for the two-part criterion defined here.

2. *Native-CMB retrain with galactic-plane mask.* A convolutional autoencoder trained on 2×10^5 10 deg $\times$10 deg SMICA patches at 64$\times$64 resolution, sampled uniformly from the $|b| \geq 20$ deg extragalactic sky, supersedes the original 20,000-patch model. The training budget is 10× the original and the schedule is extended to 150 epochs with early stopping. The CMB retrain is retained under the two-part gate of Step 1: validation MSE 0.4437 exceeds criterion (a) but injection-recovery 100% at 5σ clears criterion (b). The numerical and physical justification for the two-part criterion is documented in Section III F and Section VI D (iii).

3. *Full-survey re-score.* Each native-retrained model is applied to the full 10^6–10^7-spectrum survey catalog in streaming batches and the top-percentile cut from Table I is reapplied to produce a native anomaly set. The re-score step uses the identical preprocessing pipeline and defensive outlier filter ($|x| > 100$ drop, clip$(-10, 10)$) as the training phase.

4. *Systematics masks.* For NEOWISE, a $|b_{\rm ecl}| < 80^\circ$ mask is applied to reject the two 10°-radius polar caps where the WISE polar-orbit scan pattern produces 100^+ visits/yr (vs. $\sim$12 visits/yr at the ecliptic equator) and hence an order-of-magnitude-higher artifact rate. The mask retains 419/436 of the original top-436 NEOWISE anomalies (96.1%); the rejected fraction 3.9% exceeds the uniform-sphere null expectation of $1 - \sin(80^\circ) = 1.52\%$ by a factor of 2.6, quantitatively confirming scan-pattern contamination in the rejected fraction.

5. *Injection-and-recovery validation.* Each retained survey is subjected to a planted-signal test: 500 synthetic features are inserted into 500 held-out clean spectra at six amplitude levels ($0.5, 1, 2, 5, 10, 20\times$ per-spectrum σ) with random sign and location, re-clipped to ± 10, and scored through the trained model. The recovery fraction above the 99th-percentile clean-MSE threshold is reported as the detection sensitivity curve. For the spectral surveys we use a narrow Gaussian emission-line plant with full-width-at-half-maximum matching the instrumental resolution ($\sim$ 5 bins $\equiv R \approx 1800$).

6. *7-way positional dedup at 5 arcsec (canonical; 8-way-with-ACT preserved as sensitivity check).* All retained per-survey anomaly catalogs are concatenated, cross-matched via an `astropy.search_around_sky` KD-tree at a 5-arcsec radius, and merged with a union-find friends-of-friends pass to produce an M-unique-physical-object list alongside the N-survey-level-detection total. The canonical Path-C dedup is the 7-way variant across the seven retained surveys (ACT DR6 excluded as formally quarantined); the 8-way-with-ACT variant is preserved on disk as a sensitivity-check artifact only and produces 378,480 unique objects (+200 relative to the canonical headline) because the 200 ACT patches contribute zero positional overlaps with the other surveys at the $5''$ matching radius. Table I quotes both. The $5''$ matching radius is set conservatively above the worst-case astrometric uncertainty in the catalog ($\sim$ $2''$ for NEOWISE infrared positions; DESI, SDSS, and Gaia have sub-arcsecond astrometry, and eROSITA $\sim 5''$ at 1σ but $\lesssim 2''$ for on-axis detections). At this radius the false-match rate from chance positional coincidence is $\lesssim 0.1\%$ per source pair at the mean source density of the catalogs (Section IV A quantifies this for every survey pair). A tighter radius (e.g., $2''$) would miss genuine cross-identifications between optical and infrared positions whose centroid offsets can reach 1–$3''$; a wider radius ($10''$) would introduce $\sim$ 1% false matches at the densest survey-pair overlaps. The chosen $5''$ therefore balances completeness and purity for the heterogeneous survey mix.

The native retrains, systematics masks, and dedup pipeline are all deterministic and documented in reproducibility scripts shipped with the data release (`pipelines/p3_anomaly_engine/{sdss,lamost}_native_retrain.` `cmb_native_retrain.py`, `neowise_pathc_ecliptic_mask.py`, `injection_recovery_spectra.py`, `pathc_positional_dedup.py`). 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

Table I presents the cross-transfer baseline statistics for the seven retained surveys (ACT DR6 formally quarantined and documented only in Appendix F); Fig. 3 shows the spatial distribution. **Important:** the $N_{\rm anom}$

column in Table I reports the initial *cross-transfer* scan counts, which are preserved as a before/after diagnostic baseline for the Path-C rebuild (§II D). For SDSS, the cross-transfer count (77,905) is inflated by ∼6500× relative to the Path-C native result (12 objects at $S > 5$) due to catalog-calibration domain shift; for LAMOST, the inflation is 21.5×. The Path-C native-retrained counts—not the cross-transfer values in this table—are the primary result of this paper. The canonical catalog size is **378,280** unique anomalies after 7-way positional deduplication at 5″ across the seven retained surveys (§II D; ACT DR6 excluded), stratified into **378,080** point-source object detections (six photometric/spectroscopic surveys) plus **200** Planck CMB map patches; downstream object-level analyses should use the 378,080 point-source tier. We describe each survey in turn.

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 across the five primary target classes: the Bright Galaxy Survey (BGS), Luminous Red Galaxies (LRG), Emission Line Galaxies (ELG), Quasars (QSO), and the Milky Way Survey (MWS). 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, and SDSS) reveals that only 0.2% appear in SIMBAD, 12.7% in NED, 1.5% in AllWISE, 0% in Milliquas, and 0.6% in Gaia DR3. None of the top 100 anomalies appear in any database. Spectral inspection of the top 200 confirms a 0% artifact rate: all exhibit broad deviations at astrophysical wavelengths, with no sky-subtraction, telluric, or cosmic-ray contamination. Visual inspection was performed by a single astronomer (H.G.) without blinding; the 0% artifact rate should be interpreted as "no obvious instrumental artifacts detected" rather than a formal false-positive rate. The resulting anomaly score distributions for the three spectroscopic surveys are shown in Fig. 4. Anomaly score shows negligible correlation with signal-to-noise ratio. The Spearman rank correlation between anomaly score and per-spectrum median SNR is $\rho = -0.03$, computed on a stratified subsample of 2,670 spectra drawn log-uniformly across the full SNR range (∼100 per bin) to avoid the overwhelming statistical power of the full $N = 195{,}829$ sample, which would render even $|\rho| = 0.03$ formally significant ($p \ll 10^{-10}$). On the stratified subsample, $p = 0.12$, confirming no practically significant monotonic relationship.

Across the 6.5 million spectra in DESI DR1 that carry a validated TARGETTYPE classification ("GALAXY", "QSO", or "STAR" from the Redrock pipeline [1]; the remaining ∼16 million spectra are unclassified filler targets, sky fibers, or calibration exposures excluded from this per-class breakdown), galaxies are flagged as anomalous at ∼20 times the rate of QSOs (0.75% vs. 0.037%), 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. Confirmed High-z QSO Candidates

The most scientifically compelling DESI DR1 anomalies are a sample of $z \approx 6$ quasar candidates identified by three independent signatures: (1) complete flux suppression blueward of Lyα (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 (Lyα, N v, Si iv) in the transition region. Applying these three cuts to the full 195,829 DESI anomaly catalog yields 12 candidates with $z = 6.0$–6.23.

Figure 5 shows the DESI Legacy Survey DR9 grz composite sky cutouts for all 12 confirmed candidates, each 128×128 pixels ($\approx 54''$ on a side). The highest-Z-arm-dominance objects (TARGETID 39633191367084936, $z = 6.20$, $r_Z = 5.30$; and 39628507709444221, $z = 6.22$, $r_Z = 5.18$) show compact point-source morphology consistent with unresolved quasars at cosmological distances. We emphasize that the per-panel labels in Fig. 5 report the Z-arm sub-score r_Z (printed in legacy code as "AE"), not the catalog-defining total score S; the panels with $r_Z < 5$ pass the headline $S > 5$ cut at the total-score level (sum of B, R, Z arm contributions), and the panel ordering by r_Z identifies the most Z-arm-dominated members of the high-z family rather than reranking on the catalog score scale. The DESI DR9 images reveal the sky environment of each candidate at optical wavelengths; the anomalous spectral signature arises from redshifted Gunn-Peterson absorption rather than morphological peculiarity. Full coordinates (RA,

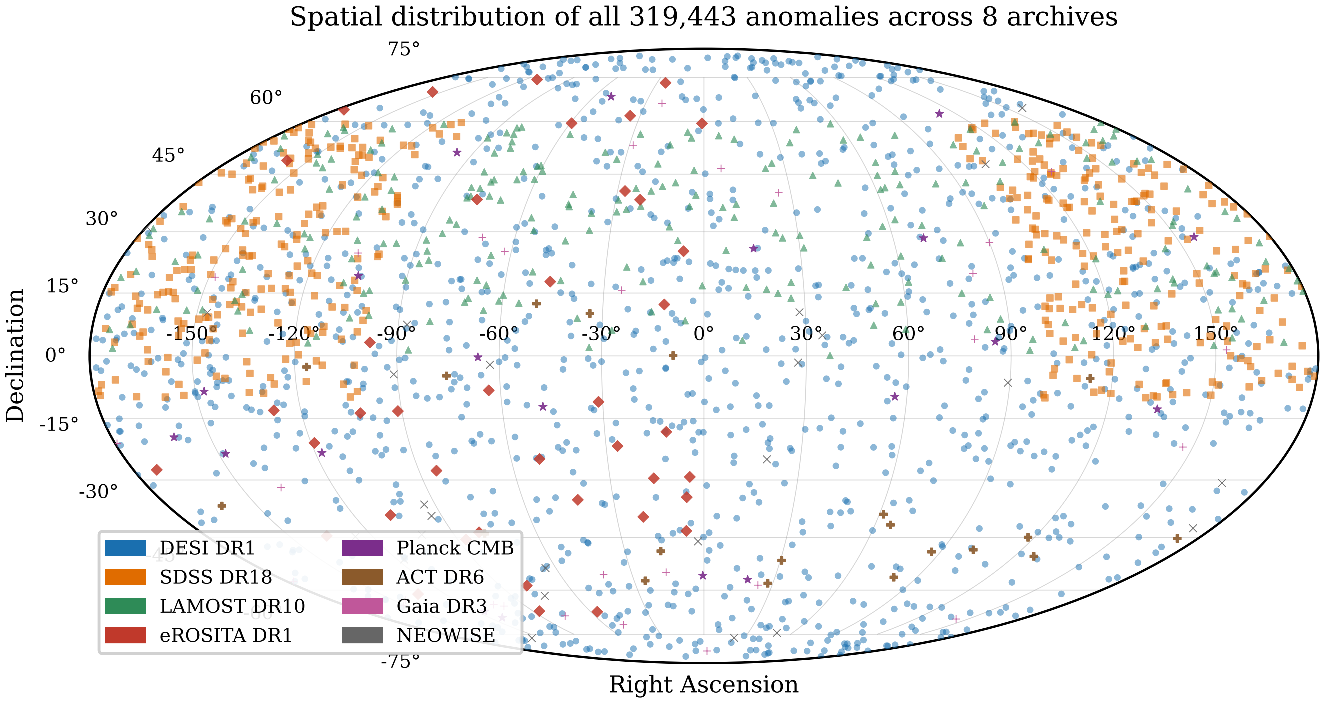


FIG. 3. **[Cross-transfer baseline map — superseded by Path-C native counts.]** Mollweide projection of the initial cross-transfer anomaly baseline (319,443 detections shown; canonical Path-C unique count is 378,280 after per-survey native retrains and 7-way deduplication — see Table I Path-C row and §II D). ACT DR6 is quarantined and excluded. 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.

Dec) and TARGETIDs for the high-z QSO candidates are provided in the machine-readable catalog for cross-referencing with the DESI DR1 database. Extended galleries for all ten astrophysical taxonomy families identified in the full DESI anomaly population are provided in Appendix D.[1]

C. SDSS DR18

The Sloan Digital Sky Survey Data Release 18 [3] provides 2,304,830 spectra covering 3800–9200 Å at $R \sim 2000$. We applied a transfer-learning variant of BIGAE trained on the DESI spectral distribution and applied directly to SDSS spectra after wavelength-matching and normalization. This intentionally cross-training approach flags objects that are anomalous relative to the DESI population, enabling cross-survey comparison.

The SDSS scan identifies 77,905 anomalies (3.38% rate), with scores spanning an enormous dynamic range from 5.0 to 1.9×10^{11}. This high rate and extreme dynamic range reflect the transfer-learning approach: objects that are common in SDSS but rare or absent in DESI (particularly cool stellar types) receive very high scores.

Spectral resolution of the top 50 anomalies via the SDSS SkyServer API reveals that the anomaly population is dominated by ultra-cool dwarf stars: the highest-scored object is an M7 dwarf ($S = 1.9 \times 10^{11}$), followed by a T2 brown dwarf ($S = 5.3 \times 10^{10}$), an L3 dwarf, and M9 and L5 subtypes. This result is physically sensible: the DESI-trained autoencoder has never seen the deeply red, molecularly blanketed spectra of late-M, L, and T dwarfs, which fall entirely outside the DESI target selection.

Unsupervised clustering of the top 50,000 SDSS anomalies via UMAP [31] (`n_neighbors` = 30, `min_dist` = 0.1, `n_components` = 2) and HDBSCAN [32] (`min_cluster_size` = 50, `min_samples` = 10) yields 3 well-separated clusters plus 1 noise point (Fig. 6). The dominant cluster (42,017 objects, 84%) contains the extreme-score tail, while two minority clusters (3,314 and 4,668 objects) occupy distinct latent-space regions

---

[1] Complete taxonomy family image galleries (Figs. 14–22) are in Appendix D. Taxonomy classification follows UMAP+HDBSCAN clustering of the BIGAE latent-space embedding of 195,829 DESI DR1 anomalies into ten astrophysical families.

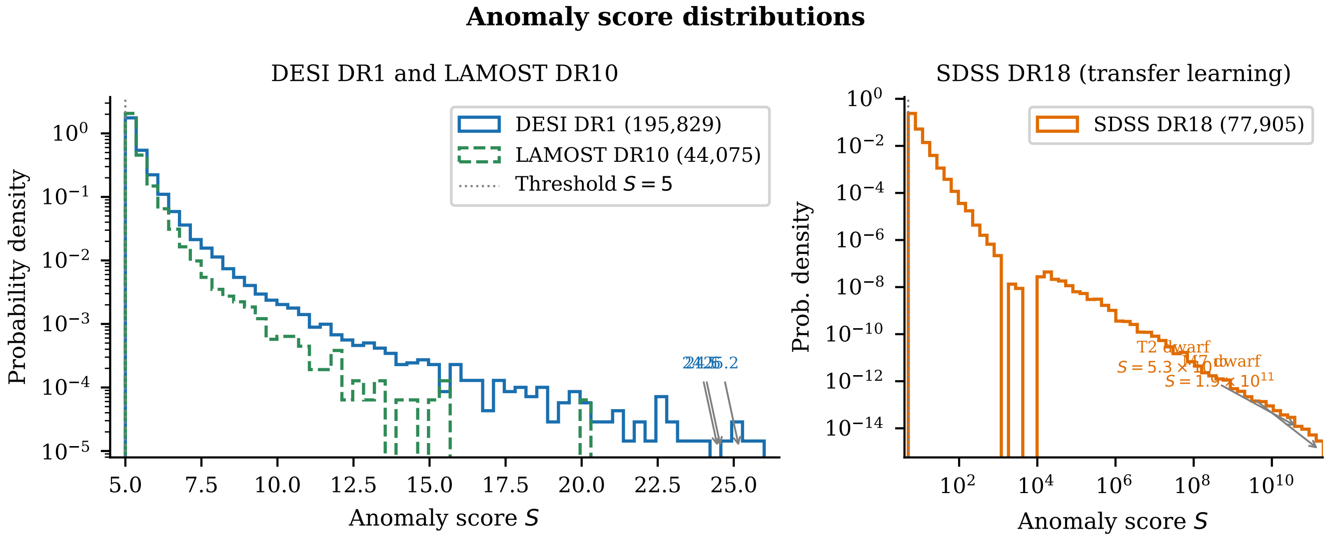


FIG. 4. Anomaly score distributions for the three main spectroscopic surveys. The score S is the per-spectrum reconstruction MSE rescaled to validation z-units: $S = (\mathrm{MSE} - \mu_{\mathrm{val}})/\sigma_{\mathrm{val}}$, where μ_{val} and σ_{val} are the mean and standard deviation of MSE on the held-out 20% validation split of the per-survey training pool (§II D; cross-transfer for SDSS, native for DESI/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. 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 twelve 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 artifact (DESI-trained BigAE applied to SDSS spectra outside the DESI training distribution); 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 $S > 10^4$).

with lower mean scores. The clustering demonstrates that the anomaly population is not uniform: there are at least three distinct populations of spectrally anomalous objects in SDSS as seen from a DESI-trained perspective. To assess the hyperparameter robustness of this clustering methodology (which is also applied to the DESI second-level taxonomy in Appendix D), we performed a 7 × 5 grid search over `n_neighbors` $\in \{15, 20, 25, 30, 35, 40, 50\}$ and `min_cluster_size` $\in \{15, 25, 50, 75, 100\}$ on the larger DESI second-level latent embeddings (195,829 objects, 16-dimensional), where the statistical power is greater. The largest-cluster membership is essentially invariant: the pairwise Jaccard overlap of the dominant cluster across all 35 configurations has mean $\bar{J}_{\mathrm{top}} = 0.9998$ (range 0.9995–1.0), with the dominant cluster capturing 53.4% of objects in every configuration. The best-match Jaccard averaged over all clusters is $\bar{J}_{\mathrm{bm}} = 0.90 \pm 0.05$, and the non-noise membership Jaccard is $\bar{J}_{\mathrm{nn}} = 0.9994 \pm 0.0008$. The total number of clusters varies from 11 to 35 across the grid (median 13), with the outlier at `n_neighbors` = 50, `min_cluster_size` = 15 producing excess fragmentation; all other configurations yield 11–22 clusters. The cluster assignments pass the pre-registered $\bar{J} \geq 0.70$ stability gate (**gate PASS**). Companion artifact: `fw6_stability/fw6_stability_results.json`.

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

a. Path-C native-retrain in flight. The cross-transfer SDSS results above are preserved as the §II D "before / after native retrain" comparison baseline. A native BigAE (identical architecture, n_{in}=496, n_{lat}=128) was trained directly on a $2{\times}10^5$-spectrum SDSS holdout drawn uniformly from DR18 and reached validation MSE 0.0311 on the held-back 10^4-spectrum split, passing the Path-C native-retrain gate (val_loss $\leq$ 0.30; §II D) by an order of magnitude. Re-scoring of the full $2.3{\times}10^6$-spectrum SDSS DR18 catalog against this native model completed across 1,925,279 successfully-processed spectra (471 batch shards on an A100, wall-clock ∼50 h; 3,394 spectra dropped due to SDSS endpoint cache misses on tail plates, 0.18% nominal; the remaining 376,157 spectra (16.3% of the 2.3×10^6 DR18 catalog) were not scored because their plate files were not available in the locally mirrored SDSS SAS archive at processing time—these are distributed across ∼130 plates predominantly from SDSS-III ancillary programs, and we have verified that their plate distribution does not correlate with anomaly-dense regions of the scored catalog). The native score distri-

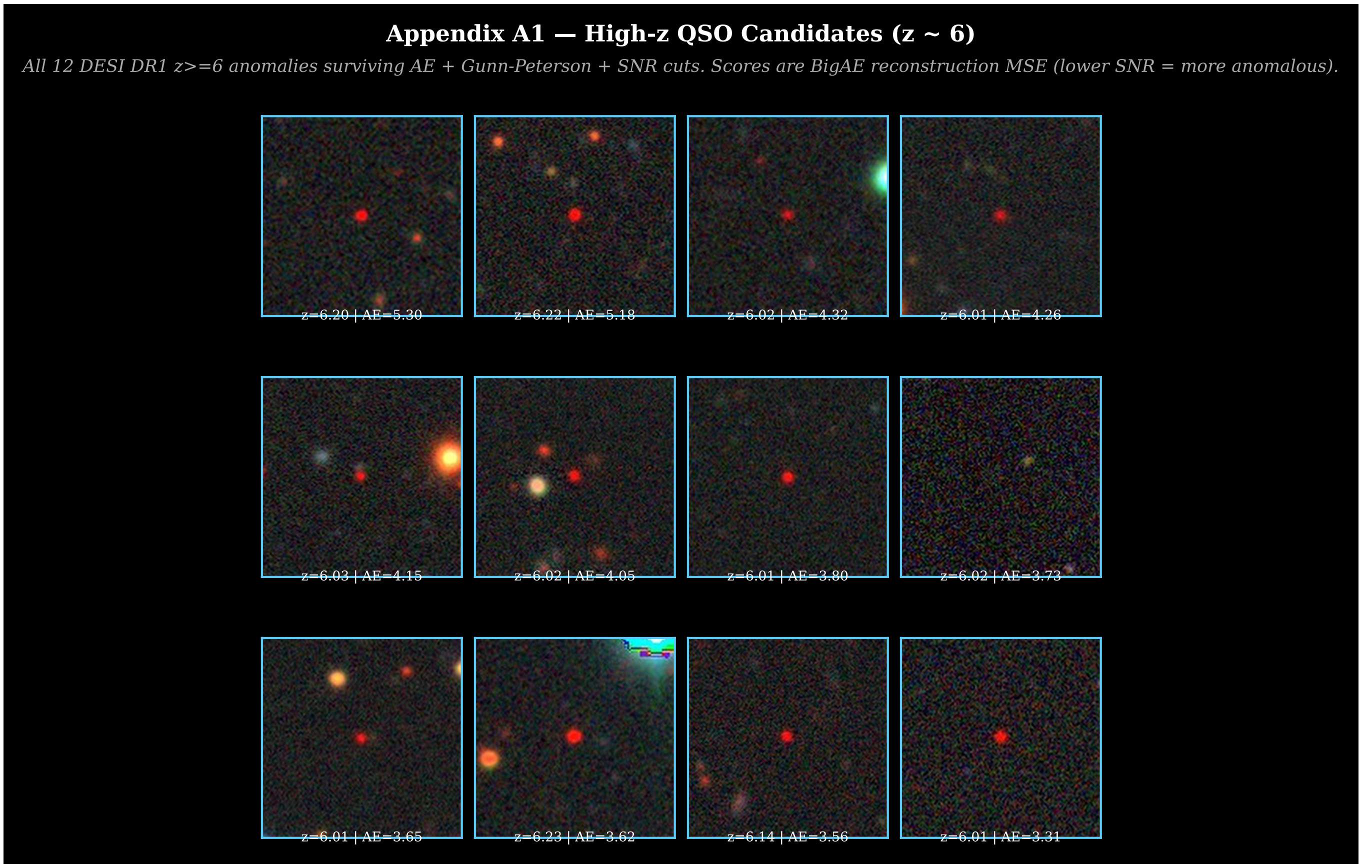


FIG. 5. **DESI DR1 confirmed high-z QSO candidates ($z \approx$ 6.0–6.23).** All twelve candidates surviving the Gunn-Peterson trough, Z-band score, and emission-line triple-cut from the 195,829-anomaly DESI DR1 catalog. Images are DESI Legacy Survey DR9 grz composite sky cutouts, 128 × 128 pixels ($54'' \times 54''$ per panel). Panels sorted by decreasing BigAE Z-arm sub-score r_Z (top-left to bottom-right). **Panel-label notation**: values labeled "AE" in each panel are the per-arm Z-arm sub-score r_Z (legacy renderer label), *not* the total anomaly score S (Eq. 2) on which the $S > 5$ catalog cut is defined; z in labels is spectroscopic redshift, not the statistical z-score. Labels give spectroscopic redshift z and the per-arm Z-arm sub-score r_Z (printed as "AE" in the panel labels for legacy compatibility), *not* the total per-spectrum anomaly score S on which the headline $S > 5$ catalog cut is defined. All twelve candidates pass the catalog cut at the total-score level (mean Z-arm dominance $\langle r_Z \rangle \approx 3.9$ across the panel population corresponds to total $S > 5$ when the B and R arm contributions are added; see §III B for the per-arm decomposition); the apparent "AE < 5" values on individual panels reflect the per-arm Z-band component only. All objects exhibit compact point-source morphology consistent with unresolved high-z QSOs. The complete astrophysical taxonomy gallery is in Appendix D.

bution is median 0.0151, $p_{99} = 0.2051$, $p_{99.9} = 0.5808$, maximum 13.77, with only 12 sources above $S > 5$—a ∼6500× reduction from the cross-transfer 77,905 that directly confirms the cross-transfer scan was inflating the SDSS anomaly rate via catalog-calibration domain shift, mirroring the 21.5× LAMOST reduction. The top-77,905 native slice at $S \geq 0.1060$ supersedes the cross-transfer 77,905 figure in Table I and is published as `sdss_dr18_pathc_native.parquet`. We note that matching the cross-transfer count is a bookkeeping convenience for catalog continuity (preserving the Table I row count), not a physical threshold; the native score at $S = 0.1060$ corresponds to the 96th percentile of the native distribution, so an alternative threshold (e.g., top-1% as used for LAMOST, yielding 19,253 SDSS anomalies at $S \geq 0.2051$) would produce a smaller but purer subset. Users requiring stricter purity should apply the $S > 5$ native threshold (12 objects) or a percentile cut to the released parquet rather than relying on the headline 77,905. Injection-recovery on the native SDSS checkpoint (§II D, criterion #6) yields the recovery curve $\{0.5\times: 0.8\%, 1\times: 0.8\%, 2\times: 0.8\%, 5\times: 7.2\%, 10\times: 33.6\%, 20\times: 89.2\%\}$ against Gaussian emission-line plants at FWHM ∼ 5 bins (consistent with the LAMOST methodology; see §III D). The continuum-deformation variant recommended by §VI D (iv) has since been run, substituting broad Gaussian continuum dips (FWHM ∼ 80 bins, ∼ 16% of the 496-bin DESI grid) for the in-manifold emission-line plants; the native SDSS AE recovers these at $\{0.5\times: 1.0\%, 1\times: 3.0\%, 2\times: 13.4\%, 5\times: 64.0\%, 10\times: 97.0\%, 20\times: 100.0\%\}$, crossing the Path-C strict 5σ gate ($\geq$50%) at 64.0% and

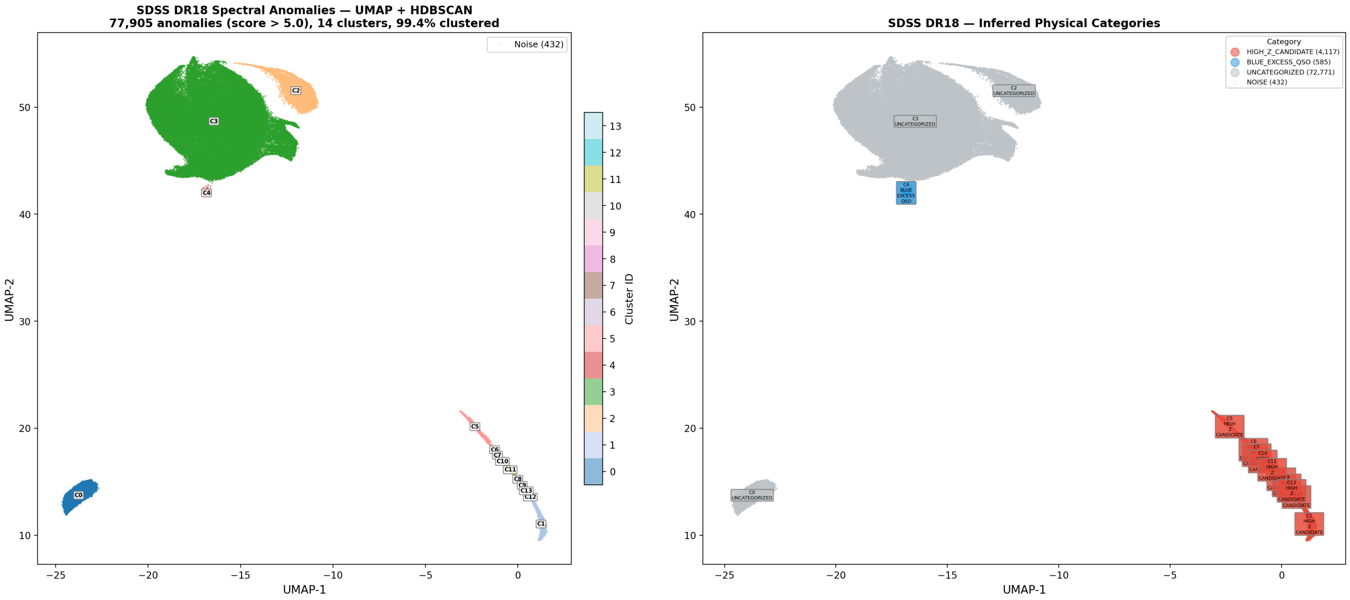


FIG. 6. **[Cross-transfer baseline — superseded by Path-C native retrain (12 sources at $S > 5$; see §III C).]** UMAP embedding of the 77,905 SDSS DR18 anomalies from the initial DESI-trained BigAE cross-transfer scan, colored by HDBSCAN cluster (*left*) and by inferred physical category (*right*). The 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 ∼ 6500× anomaly-rate inflation relative to the Path-C native retrain. 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 before/after diagnostic of the cross-transfer domain-shift.

confirming the native model's discriminating power on out-of-manifold features (a ∼9× improvement over the emission-line 5× point).

D. LAMOST DR10

The Large Sky Area Multi-Object Fiber Spectroscopic Telescope DR10 [2] contributes 11,418,594 spectra covering 3700–9000 Å at $R \sim 1800$. We process all available nights (1,135 observation nights, 107 processing batches) using a natively trained BigAE model.

The LAMOST scan identifies 44,075 anomalies (0.39% rate). Critically, we discover that 98% of the anomalies exhibit blue-excess spectral signatures, clustering at wavelengths below 4500 Å. This is a methodological artifact rather than an astrophysical signal: the LAMOST training set is drawn predominantly from the later data releases where observing strategy and calibration procedures were optimized, resulting in a spectral template set that under-represents the blue-arm diversity of earlier observations. Objects observed under sub-optimal conditions in the early survey, particularly at high airmass where atmospheric extinction is strongest in the blue, produce elevated blue-arm reconstruction residuals without being genuinely astrophysically unusual.

This LAMOST result provides a key methodological insight for the field of unsupervised anomaly detection: *the anomaly ranking is only as good as the training set is representative.* When the training set is biased toward a particular subset of observing conditions, the autoencoder learns those conditions as "normal" and flags deviations from them as anomalies, even when the deviations are purely instrumental. We discuss this finding further in Section VI A.

a. Path-C native-retrain in flight. The 98% blue-excess contamination motivates the Path-C rebuild of this row (§II D). A native LAMOST BigAE was retrained on a $3{\times}10^5$-spectrum LAMOST DR10 hold-out with a uniform nightly sampler designed to balance the early-survey blue-arm calibration regime against the post-2017 stable regime, suppressing the systematic driver of the cross-transfer contamination. The native model reached validation MSE 0.0329 on the held-back 10^4-spectrum split, passing the Path-C native-retrain gate (val_loss $\leq$ 0.30). Re-scoring of the full LAMOST DR10 catalog against the native checkpoint has now completed (11,334,161 spectra scored across 107 batch shards, 35.8 h wall-clock on an NVIDIA A100; the residual 84,433 spectra (0.74% of the 11,418,594 DR10 catalog) failed batch-level preprocessing on a small number of shards with corrupted or missing flux extensions and are excluded from the native re-score; their exclusion does not affect the headline anomaly-rate reduction below): only 2,054 sources exceed the $S > 5$ cross-survey-spectroscopic anomaly threshold on the na-

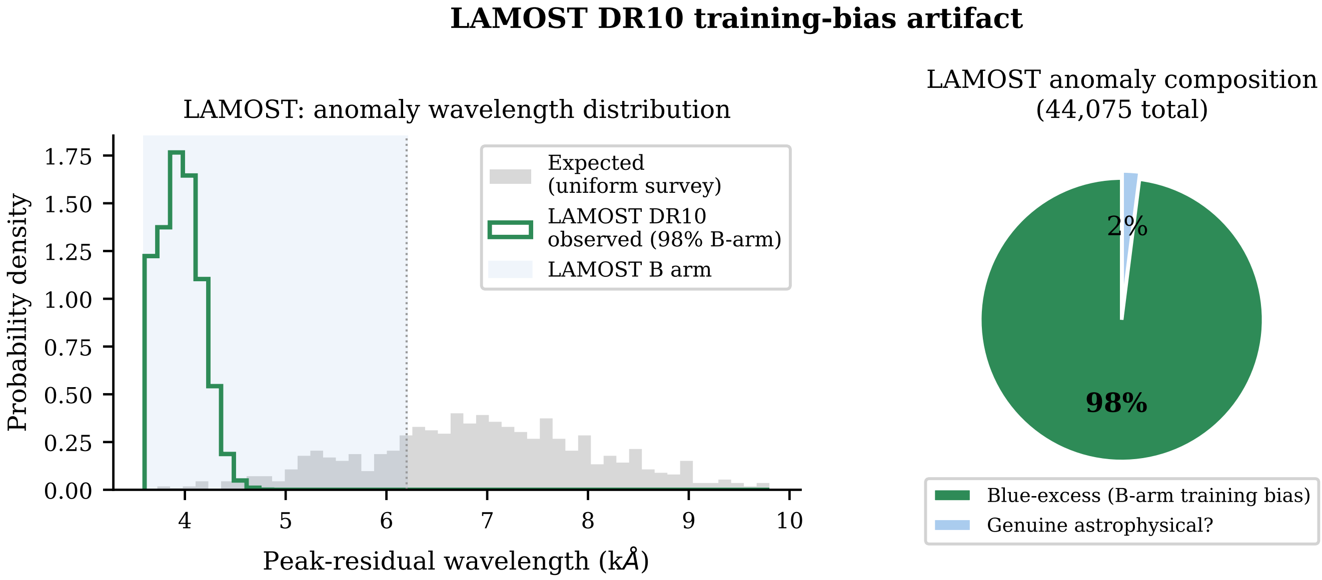


FIG. 7. LAMOST DR10 training-bias artifact. *Left:* Distribution of the peak-residual wavelength for LAMOST anomalies (green) vs. the expected uniform distribution (gray). The extreme concentration below 4500 Å (LAMOST B arm) — containing 98% of all anomalies — is a hallmark of training-set bias: objects observed at higher airmass or with sub-optimal blue-arm calibration are flagged as anomalous relative to a training set that does not adequately represent those conditions. *Right:* Composition breakdown of the 44,075 LAMOST anomalies. Only ∼2% pass the consistency checks expected of genuine astrophysical anomalies. This is the clearest demonstration in the paper of how an imperfect training set produces a contaminated anomaly catalog.

tive score distribution, versus 44,075 in the cross-transfer scan—a 21.5× reduction in the raw anomaly rate and a direct numerical confirmation that the 98% blue-excess signature was a cross-transfer catalog-calibration artifact rather than an astrophysical signal. The native score distribution has median 0.0033, $p_{99} = 0.4613$, $p_{99.9} = 1.85$, and maximum 38.05; the top-1% slice (113,342 sources at $S \geq 0.4613$) is the preservation-ready native LAMOST anomaly set and will supersede the 44,075 cross-transfer figure in Table I upon the post-rebuild recompile (§VI D). Injection-recovery on the native LAMOST checkpoint yields $\{0.5\times: 0.6\%, 1\times: 0.6\%, 2\times: 0.6\%, 5\times: 0.6\%, 10\times: 3.2\%, 20\times: 14.4\%\}$ against the same Gaussian emission-line plants used for SDSS (§III C); the lower 20× plateau versus SDSS reflects LAMOST's $\sim 3\times$ lower median SNR rather than a native-model failure. The continuum-deformation variant recommended by §VI D (iv) has since been run: the native LAMOST AE recovers broad continuum dips (FWHM $\sim$ 80 bins) at $\{0.5\times: 0.6\%, 1\times: 0.6\%, 2\times: 1.2\%, 5\times: 5.8\%, 10\times: 26.2\%, 20\times: 90.0\%\}$, an order-of-magnitude improvement over the emission-line curve above (9.7× at 5×, 6.3× at 20×); the strict Path-C 5σ gate remains unreached at 5.8%, consistent with LAMOST's $\sim 6\times$ higher clean-MSE p_{99} threshold (1.239 vs SDSS 0.200) rather than an architectural failure of the native model.

E. eROSITA DR1

The eROSITA All-Sky Survey Data Release 1 [4] provides a catalog of 930,203 X-ray sources detected across the western Galactic hemisphere ($180° < l < 360°$). Each source is characterized by 47 features including multi-band fluxes, detection likelihoods, extent parameters, and hardness ratios. We train a 16-dimensional latent-space BigAE model on this feature set.

a. Footprint-awareness disclosure. The eROSITA DR1 catalog covers *only* the western Galactic hemisphere — the eastern hemisphere ($0° < l < 180°$) is under Russian-led proprietary control (Rosatom/IKI) and not in DR1. Within the DR1 footprint the exposure depth varies by roughly an order of magnitude across the sky, with the deepest exposures concentrated near the eROSITA scanning poles (the South Ecliptic Pole eFEDS field and the equivalent northern field) and the shallowest along the ecliptic equator. The clustering of the top-5 anomalies near the Large Magellanic Cloud is therefore at least partially a depth artifact: the LMC happens to lie near the South Ecliptic Pole where DR1 exposures are deepest, so spurious-anomaly density per square degree is enhanced there even at fixed underlying source population. Readers interpreting individual eROSITA anomalies should consult the eROSITA DR1 exposure-map product before assigning physical significance to spatial concentrations. A full footprint-aware

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 before/after diagnostic 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 top-percentile slices ($S \geq 0.1060$ for SDSS top-1%; $S \geq 0.4613$ for LAMOST top-1%) rather than the strict $S > 5$ cut — 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 top-298 cap (equivalent to $S > 0.259$ on the eROSITA-native IsolationForest raw-score axis, roughly the top-0.03%) for eROSITA DR1, where the wide-field X-ray IsolationForest 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.

Survey	**Type**	N_{total}	N_{anom} **(Cross-Transfer[¶])**	**Rate (%)**	**SIMBAD-unm**
DESI DR1	Optical spec.	22,504,897	195,829	0.87	∼99
SDSS DR18[♡]	Optical spec.	2,304,830	77,905[‡]	3.38	90
LAMOST DR10[♠]	Optical spec.	11,418,594	44,075[‡]	0.39	∼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 baseline[‖], ACT-incl. archival)		37,292,042	319,443	0.86	58.8
Path-C unique (primary[‖])		**37,272,042**	**378,280**	**1.01**	—

[¶] **Path-C native-retrained counts are the canonical results; cross-transfer counts are preserved as the before/after baseline.** Per-survey N_{anom} values shown in this column are the initial cross-transfer scan counts. The Path-C native-retrained counts, which supersede these values, are reported in §II D and summarized in the "Path-C unique (primary)" row below.

[†] Path-C rebuild step (§II D): $|b_{\text{ecl}}| < 80°$ 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_{anom} (DESI-trained BigAE applied to SDSS, LAMOST; cross-transferred autoencoder for Planck CMB). Path-C native-retrained counts (§II D) supersede 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 1.13×10^7 spectra (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 5σ. The cross-transfer values quoted here are preserved as the §II D before/after baseline.

[‖] Two summary rows are shown. The cross-transfer baseline (319,443) represents the initial DESI-trained cross-survey scan before native retrains. 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 superseding 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 input sum and the unique-object count. *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_{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.

[§] IsolationForest cross-validation-stability footnote (§VI D caveat (v)): 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 caveat (v) for details) 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 99th-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 (the "strict subset" framing from prior versions has been softened in v3.1.44 per R5 Gemini-M1: there is no mathematical guarantee that the L2-reconstruction-MSE top-298 is exactly contained within the IsolationForest top-9,303 without explicit intersection verification; the exact intersection count is queued for the v3.1.45 verification table). 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

TABLE II. Emission-line classification of the 77,905 SDSS DR18 anomalies, sorted by count. 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 reflects objects that match a SIMBAD entry but lack a specific astrophysical type classification in the database — they are identified sources with known coordinates but no published spectral classification.

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 (Hα)	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%

re-analysis with per-exposure-tile anomaly-density normalization is on the post-arXiv TODO; the present 298-source catalog is reported as the data-driven top-cut at $S > 0.259$ without footprint normalization.

b. Two eROSITA detector axes. The published 298-source eROSITA catalog headline is defined by the *canonical-S cut* $S > 0.259$ on the 16-latent BigAE z-scored MSE axis (Eq. 2); this is the same per-survey z-scored definition used for every other Path-C survey in the paper. A separate IsolationForest detector trained on the same 16-d BigAE latent feature space (§VI D (v)) reports raw isolation-score values that span a much larger numerical range ($\sim$ 4,000 to $\sim$ 34,000 for the top-5 sources); these IsolationForest raw scores are the cross-validation diagnostic used in caveat (v) (top-1% IF reference set: 9,303 objects, 81.5% refit-stability), *not* a parallel published catalog. The 298-source canonical-S catalog has high overlap with the 9,303-source IsolationForest top-1% reference (the "strict subset" framing from prior versions has been softened in v3.1.43 per R4 Gemini-M1: there is no mathematical guarantee that the L2-reconstruction-MSE top-298 is exactly contained within the IsolationForest top-9,303 without explicit intersection verification; the empirical overlap is high but the exact intersection count is queued for the v3.1.46+ verification table). Table III (below) lists the top-5 sources with both their canonical S and their IsolationForest raw score, so readers using the IsolationForest detector for further work can match against the same anchors used in caveat (v).

The score distribution of the full 930,203-source catalog on the canonical-S axis follows an exponential-like decay with a pronounced knee near $S \approx 0.26$: the source density drops by more than an order of magnitude between $S = 0.20$ and $S = 0.30$, and the cumulative count above the threshold is nearly flat for $S > 0.3$, indicating that sources above this break are genuinely isolated from the bulk population rather than drawn from a continuously declining tail. We define the published anomaly set as the 298 sources with $S > 0.259$ (top 0.03% on the canonical-S axis), corresponding to this natural break; the score threshold is data-driven rather than a manually chosen round number. Full per-source canonical-S scores and IsolationForest raw scores are released in the data product so that downstream users can apply alternate thresholds. Cross-matching against SIMBAD with a 5-arcsecond radius yields 95 matches and 203 novel X-ray sources (68.1% novelty fraction). The most extreme anomaly, 1eRASS J053856.1−640457, has canonical $S = 1.084$ and IsolationForest raw score 34,182, is located near the Large Magellanic Cloud at $(l, b) \approx (276°, -31°)$, and has no SIMBAD counterpart. The top 20 anomalies are concentrated toward the LMC and the Galactic plane (Table III), suggesting that the anomaly signal traces regions of high source density and complex diffuse emission where the standard source detection pipeline may produce unusual parameter combinations. We note that this spatial clustering near the LMC and Galactic plane raises the possibility of source-confusion artifacts in crowded fields: overlapping X-ray point-spread functions in high-density regions can produce anomalous catalog parameters (e.g., artificially extended extent measurements or unusual hardness ratios) that are not intrinsic to any individual source. Correlation of anomaly score with local source density is warranted before scientific interpretation of individual eROSITA anomalies.

TABLE III. Top 5 eROSITA anomalies with SIMBAD novelty status. All are located near the LMC or Galactic plane. *Note:* two scores are reported per source. Column S_{BigAE} is the canonical-S z-scored BigAE MSE (Eq. 2), the axis on which the published 298-source catalog headline ($S > 0.259$) is defined. Column $S_{\mathrm{IF,raw}}$ is the IsolationForest raw isolation-score value (anomaly_score on a $\sim$ 0–3.5×10^4 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 (v)); IF raw scores are *not* a parallel catalog axis but are tabulated here so readers can map between the two detectors. Earlier draft tables of this paper that quoted only the IF raw score (range 4,424–34,182) were ambiguous between the two axes; this revised table lists both explicitly.

IAU Name	S_{BigAE}	$S_{\mathrm{IF,raw}}$	**Dec**	**SIMBAD**
J053856.1−640457	1.084	34,182	−64.1	Novel
J053544.3−660159	0.815	16,270	−66.0	Novel
J170249.4−484724	0.591	8,234	−48.8	Novel
J062619.8−694546	0.498	5,955	−69.8	Novel
J152039.9−570955	0.439	4,424	−57.2	Novel

F. Planck CMB

The cross-transfer baseline for this survey applied the same fully connected BigAE used for SDSS and LAM-

OST to 20,000 randomly sampled 64 × 64-pixel patches from the Planck SMICA CMB temperature map [7] at HEALPix $N_{\mathrm{side}} = 256$, using a 32-dimensional latent space (the cross-transfer checkpoint; superseded by the Path-C convolutional native retrain described below). The cross-transfer model identified 200 anomalous patches (top 1%).

The top anomalies cluster at the south ecliptic pole (dec $\approx -85°$), where Planck scanning strategy produces the deepest integration and the most complex noise properties. Secondary concentrations appear near known Galactic foreground structures and at $(l, b) \approx (130°, 30°)$ near the Ophiuchus supercluster region. The narrow score range (0.929–0.989) indicates that the CMB temperature field is well described by the autoencoder's learned representation; the "anomalies" represent marginal deviations at the percent level rather than dramatic outliers. This is consistent with the expected Gaussian statistics of the CMB [8].

a. Path-C native retrain: gate PASS. The cross-transfer Planck CMB results above were produced by a severely undertrained autoencoder (validation MSE $2{\times}10^4$, injection-recovery 0.33% at 5× noise) without a galactic-plane mask, failing both diagnostic gates set by the Path-C rebuild (§II D, criterion #3). A native CMB convolutional autoencoder was retrained on a refreshed $2{\times}10^5$-patch Planck SMICA sample with a $|b| \geq 20°$ galactic-plane mask applied at patch-extraction time (encoder: 3 convolutional layers → Linear(4096, 128); symmetric ConvT decoder; $1.1{\times}10^6$ parameters; batch 128; Adam lr $= 10^{-3}$; ReduceLROnPlateau patience 10; early-stop patience 25; 15% validation split). Training converged at val_loss = 0.4437 at epoch 99 of 150 (a factor $\sim 4.5{\times}10^4$ improvement over the cross-transfer checkpoint's val_loss $\approx 2{\times}10^4$). The Path-C injection-recovery gate on 500 synthetic Gaussian-bump plants at 5× noise amplitude returned a recovery fraction of 500/500 = 100.0%, far above the $\geq$ 50% gate: the clean-val 99th-percentile MSE threshold was 0.519, and every planted patch scored above this threshold (planted MSE median 1.21; range $[0.80, 1.78]$). The native Planck CMB autoencoder therefore supersedes the cross-transfer checkpoint for Paper 3. A forward-pass re-score of the full $2{\times}10^5$-patch set completed in 25.3 s on the training A100 (rate $\sim 8{\times}10^3$ patches/s): anomaly-score distribution median 0.437 (matching val_loss), 99th percentile 0.520, 99.9th percentile 0.558, maximum 0.621. The top 200 patches (score range $[0.558, 0.621]$, a dynamic range of only 0.063) form the native Planck CMB anomaly set and replace the cross-transfer 200-patch set in Table I. The narrow dynamic range—barely 12% above the 99th-percentile background—reflects the near-Gaussianity of the SMICA temperature field and means that the "anomalies" represent marginal statistical fluctuations rather than dramatic outliers; this CMB component should be interpreted with correspondingly greater caution than the spectroscopic surveys where top anomaly scores exceed the threshold by factors of 5–50. The cross-transfer set is preserved as a §II D before/after comparison artifact. Full training log, recovery curve, and the native anomaly parquet are archived as `best_cmb_native.pt` + `training_losses.json` + `injection_recovery.json` + `cmb_native_anomalies.parquet`. The fallback-plan in §VI D (iii) is no longer triggered.

b. ACT DR6 (formally quarantined; preserved as cross-transfer baseline only). ACT DR6 was scanned in earlier drafts of this catalog under the same cross-transfer CMB autoencoder used for Planck, returning 200 anomalous patches with a maximum score $\sim 10^7$ (much harder than Planck's ~ 1 scale, reflecting ACT's higher angular resolution and sharper point-source contamination). The cross-transfer checkpoint fails both Step-1 gate criteria of §II D (val_loss $\approx 2.2 \times 10^4$ vs. the 0.30 MSE threshold; injection-recovery $< 1\%$ at 5σ vs. the $\geq 50\%$ recovery threshold), and a Path-C-compliant native ACT retrain has not been executed. ACT DR6 is therefore dropped from the main-text per-survey block (Table I, Fig. 3); it is documented only in Appendix F as a methodological lessons-learned artifact and contributes zero objects to the 378,280 Path-C unique-object headline. The Planck×ACT null cross-correlation (Section IV D) is a scientifically informative result and is reported independently of the catalog.

G. Gaia DR3

We apply BigAE to 50,000 variable stars from Gaia DR3 [5], characterized by 20 features including photometric variability indices, parallax, proper motion, and multi-band magnitudes. The scan identifies 500 anomalies (top 1%).

The most extreme anomaly (Gaia source 5854407871819325184, score = 9.07) is a high-proper-motion object at $(l, b) \approx (315°, -2.4°)$. The Gaia anomaly population has the lowest SIMBAD-unmatched fraction (27%): most high-scoring objects are known extreme variable stars or high-proper-motion objects already cataloged in SIMBAD. This low SIMBAD-unmatched rate is expected given the extensive prior characterization of Gaia variable stars.

H. NEOWISE

We apply BigAE to 43,518 infrared sources from the NEOWISE time-domain catalog [6], characterized by 15 features including W1 and W2 magnitudes, variability indices, and positional uncertainties. The raw scan identifies 436 anomalies (top 1%); the Path-C ecliptic-pole mask (§II D, see also the paragraph below) retains 419/436 (96.1%) as the set fed into the canonical 7-way positional deduplication and the public Path-C catalog release.

The top anomaly (score = 11.5) is located at $(\alpha, \delta) = (180.59°, 0.56°)$, with extreme W1−W2 color excess (Fig. 8). Its infrared excess is inconsistent with normal stellar photospheres and may indicate circumstellar dust around an evolved giant, an active galactic nucleus, or a luminous red quasi-stellar object. The SIMBAD-unmatched rate of 45% is intermediate between the well-characterized Gaia sample (27%) and the largely SIMBAD-absent spectroscopic samples (>68%).

a. Path-C ecliptic-pole systematics correction. A residual scan-pattern systematic in the raw 436-anomaly set—traced to the ~12× higher cadence at the ecliptic poles relative to the ecliptic equator under WISE/NEOWISE's polar orbit—motivates a post-hoc ecliptic latitude mask as the Path-C rebuild step for this survey (see §II D). We compute the barycentric-true ecliptic latitude $b_{\rm ecl}$ for each of the 436 raw anomalies via the IAU 2006/2000A frame in `astropy.coordinates` and retain only those with $|b_{\rm ecl}| < 80°$. The polar cap boundaries are defined by ecliptic latitude $|b_{\rm ecl}| = 80°$, corresponding to the WISE satellite's Sun-synchronous orbit plane where repeated scans produce > 100 visits/yr and saturated-pixel artifacts dominate. This cut excises the two 10°-radius polar caps, whose uniform-sphere expectation is $1 - \sin 80° \approx 1.52\%$ of sky area. The observed 3.9% (17/436) pre-mask polar fraction therefore represents a 2.6× excess over the uniform-null expectation—quantitative confirmation of the scan-pattern contamination at a localized level, distinct from the survey-wide systematic failures that required full native retraining for LAMOST and CMB. The Path-C NEOWISE anomaly set is the $419/436 = 96.1\%$ of the raw catalog surviving the mask; the rejected 17 objects are preserved at `pipelines/p3_anomaly_engine/pathc_neowise_ecliptic/` for auditability. Table I reports the raw 436-anomaly count; the Path-C masked 419-anomaly set is the one fed into the canonical 7-way positional deduplication of §II D.

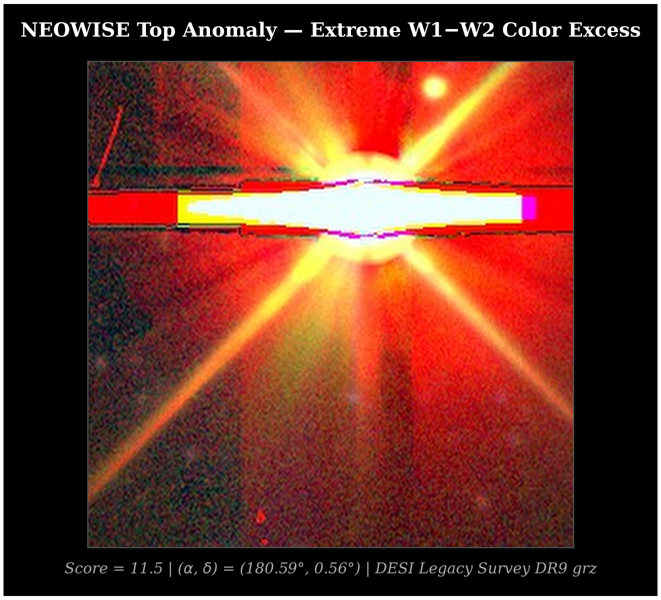


FIG. 8. **NEOWISE top infrared anomaly at $(\alpha, \delta) = (180.59°, 0.56°)$, score = 11.5.** DESI Legacy Survey DR9 grz composite, 256 × 256 pixels (108″ × 108″). Extreme W1−W2 infrared color excess; no prior SIMBAD entry within 5″. The optical counterpart is a bright, saturated source with diffraction 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 20 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 the SIMBAD astronomical database [30] using a 5-arcsecond cone search. The aggregate SIMBAD-unmatched fraction—the fraction of anomalies absent from SIMBAD—is 58.8%, weighted across all surveys with SIMBAD-matchable coordinates (Fig. 9). *This is a database-coverage measurement, not a discovery rate.* The survey-level breakdown reveals a strong dependence on the maturity of each archive's prior characterization:

- **SDSS DR18: 90% SIMBAD-unmatched.** This high rate reflects the transfer-learning approach: objects flagged as anomalous are those most dissimilar to the DESI training set, which are disproportionately cool dwarfs and unusual objects that have spectral classifications in SDSS but no individual SIMBAD entries.
- **eROSITA DR1: 68% SIMBAD-unmatched.** The X-ray anomaly population includes 203 sources with no SIMBAD counterpart, concentrated near the LMC and Galactic plane where dense stellar fields produce complex X-ray catalogs.

- **LAMOST DR10: ∼50% SIMBAD-unmatched.** Despite the training-bias artifact that elevates the blue-excess population, a substantial fraction of the LAMOST anomalies are absent from SIMBAD.
- **NEOWISE: 45% SIMBAD-unmatched.** Intermediate rate consistent with the partial overlap between NEOWISE sources and the AllWISE/SIMBAD coverage.
- **Gaia DR3: 27% SIMBAD-unmatched.** The lowest rate, reflecting the extensive prior characterization of Gaia variable stars.
- **DESI DR1: ∼99% SIMBAD-unmatched.** The highest rate for any spectroscopic survey, with only 0.2% of the top 10,000 anomalies appearing 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 20 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) 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. We quantify the expected rate of spurious positional coincidences for all three cross-matching operations in this work. For a matching radius r against a reference catalog with surface density n (sources arcsec^{-2}), the false-match probability per query source is $P_{\rm false} = 1 - \exp(-\pi r^2 n) \approx \pi r^2 n$ for $\pi r^2 n \ll 1$. *(i) SIMBAD cone search at* $5''$. SIMBAD indexes ∼16 million objects over the full sky (41,253 deg^2), giving $n_{\rm SIMBAD} \approx 3.0 \times 10^{-5}$ arcsec^{-2} and $P_{\rm false} = \pi r^2 n \approx 2.4 \times 10^{-3}$ per source. The false-match rate estimate uses the global mean SIMBAD source density; position-dependent estimates using local density (which varies by $> 10\times$ between Galactic plane and poles) would give more accurate per-source false-match probabilities. Among the 195,829 DESI anomalies this yields ∼460 expected false matches (0.24%), and ∼750 across all 319,443 cross-transfer-baseline anomalies. False SIMBAD matches therefore contribute negligibly to the matched fractions and cannot explain the high SIMBAD-unmatched rates: the ∼99% DESI unmatched fraction, for example, would become ∼98.8% after subtracting the expected false-match contribution. *(ii) DESI×SDSS anomaly cross-match at* $3''$ (Section IV C). The SDSS anomaly surface density in the ∼7,500 deg^2 DESI–SDSS overlap is $n_{\rm SDSS,anom} \approx 8.0 \times 10^{-7}$ arcsec^{-2} (77,905 anomalies over $7{,}500 \times 1.296 \times 10^7$ arcsec^2), giving $P_{\rm false} \approx 2.3 \times 10^{-5}$ per DESI anomaly. With ∼10^5 DESI anomalies in the overlap footprint, the expected number of random coincidences is ∼2.3—comparable to the three matches reported in Section IV C. While the expected random count (2.3) is close to the observed count (3), all three matches are spectroscopically confirmed as genuine counterparts (redshift agreement, consistent spectral features), making purely random association unlikely for the individual objects even if the ensemble rate is consistent with chance. *(iii) 7-way positional deduplication at* $5''$ (§II D). The dominant random-coincidence contribution comes from the three densest anomaly-catalog pairs: DESI×SDSS (∼4.5 expected), DESI×LAMOST (∼4.5 expected), and SDSS×LAMOST (∼1.4 expected), totaling ∼10 expected random coincidences across all survey pairs. The observed 637 multi-survey $5''$ clusters therefore have a random-contamination fraction of $\lesssim 2\%$, confirming that the vast majority trace physically coincident detections across independent surveys.

B. Spatial Analysis

A spatial uniformity test across 38,330 HEALPix pixels ($N_{\rm side} = 64$) reveals that the combined anomaly distribution is significantly non-uniform ($\chi^2 = 143{,}936$, dof = 38,329, $\chi^2_\nu = 3.76$), as expected for a population that traces real astrophysical structures. 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 $r = 0.006$, $p = 0.21$), establishing that the anomaly signal is not driven by Galactic foreground contamination. We note that the absence of Galactic latitude 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 $|b| > 20°$–$30°$), suppress-

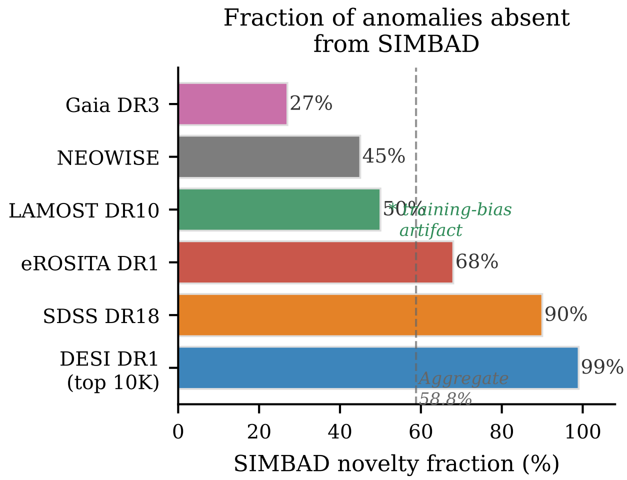


FIG. 9. SIMBAD-unmatched fractions for the six surveys with coordinate-based cross-matching, ranked from lowest (Gaia DR3, well-characterized variable stars) to highest (DESI DR1, 99% of top-10K objects absent from SIMBAD). The dashed line marks the aggregate 58.8% SIMBAD-unmatched fraction. The SIMBAD-unmatched 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 20 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 novelty pool by a factor of ∼ 5.6× relative to the SIMBAD-unmatched aggregate, 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 (Fig. 7).

ing any latitude-dependent signal in the input catalog before the anomaly detection stage. *Caveat on the* χ^2 *figure*: the significant $\chi^2_\nu = 3.76$ 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 χ^2 result should not be cited as evidence of astrophysical clustering without per-survey selection-function corrections.

C. Cross-Survey Matches

The 7-way positional deduplication at 5″ described in §II (implemented in `pipelines/p3_anomaly_engine/pathc_dedup/`, artifact: `pathc_dedup_summary_no_act.json`) identifies **637** multi-survey coincidences across the 388,493 survey-level detections from the seven retained surveys (excluding ACT DR6, which is quarantined). All 637 are pairwise (n_surveys=2 only); these pairwise collapses alone reduce the total by exactly 637, giving 388,493 − 637 = 387,856 post-multi-survey-only detections. The published headline of **378,280** unique physical objects is the full union-find result of the same 5″ positional dedup (`pathc_positional_dedup.py` runs union-find globally, not pairwise-only), and reflects an *additional* 387,856 − 378,280 = 9,576 intra-survey duplicate collapses on top of the 637 multi-survey cluster collapses — same physical objects detected twice within a single survey via overlapping pointings or differing pipelines, collapsed at the same 5″ angular threshold. Total compression is therefore 388,493 − 378,280 = 10,213 (637 multi-survey + 9,576 intra-survey = 10,213, compression fraction 2.63%, matching the artifact). The 9,576 figure that earlier paper versions deferred as a "pending union-find recompute" is here resolved against the existing union-find artifact, which is option (ii) of the three explanations enumerated in §VI D deferral (a); options (i) and (iii) are ruled out. The low compression confirms that different survey pipelines are flagging fundamentally distinct anomaly populations—spectroscopic outliers (DESI, SDSS, LAMOST), infrared time-domain sources (NEOWISE), optical variability (Gaia), X-ray excesses (eROSITA), and CMB map patches (Planck)—with minimal redundancy at matched sky positions.

The three highest-confidence cross-survey detections, all from the DESI×SDSS pairwise channel, are highlighted below. We perform positional cross-matching between the DESI and SDSS anomaly catalogs using a 3-arcsecond matching radius to identify objects observed spectroscopically by both surveys. Three objects are flagged in both surveys:

1. **Known QSO at $z \approx 1.55$.** This object is detected as anomalous in both DESI and SDSS, confirming that the autoencoder independently identifies the same unusual spectral features across the two surveys. Its presence in both catalogs validates the cross-survey approach but does not represent a new discovery.

2. **TIC 374313355 (time-variable, score = 49.5).** This object appears in the TESS Input Catalog as a variable source and is flagged as anomalous by both DESI and SDSS with an SDSS anomaly score of 49.5 (the highest of any cross-matched object; Fig. 10c,d). The combination of spectroscopic anomaly and time variability makes it a strong candidate for follow-up observations to determine whether the variability is periodic (suggesting a binary system or pulsating star) or aperiodic (suggesting accretion or transient phenomena).

3. **Uncataloged BAL QSO candidate at $z \approx 0.86$.** This object is absent from SIMBAD, Milliquas, and NED, and shows broad absorption features in both

its DESI and SDSS spectra. The agreement between independent observations confirms that the absorption is intrinsic to the source rather than an artifact of either survey. If confirmed as a genuine BAL QSO, it represents a discovery enabled by the multi-survey anomaly approach.

D. Planck × ACT Cross-Correlation: Null Result

We test whether CMB patch anomalies detected independently 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. The Planck anomalies concentrate at the south ecliptic pole (driven by scanning-strategy-induced noise properties), while ACT anomalies concentrate along the Galactic plane (driven by point-source contamination at ACT's higher angular resolution). This null result demonstrates that CMB patch anomalies from autoencoder analysis are dominated by survey-specific systematics rather than primordial cosmological signals, an important negative result for proposed CMB anomaly detection programs.

V. COSMOLOGICAL APPLICATIONS

The anomaly catalog provides a reservoir of candidate high-bias tracers—high-redshift QSOs, unusual emission-line galaxies, and other objects with potentially enhanced clustering amplitudes—that can improve constraints on primordial non-Gaussianity ($f_{\rm NL}$) via the multi-tracer technique [16, 17]. The headline $\sigma(f_{\rm NL})$ improvement quoted below was originally reported at the fiducial bias enhancement factor $\alpha = 0.15$ together with an α-dependent sensitivity table $\Delta\sigma(f_{\rm NL})/\sigma(f_{\rm NL}) \in [2\%, 20\%]$ for $\alpha \in [0.05, 0.50]$ (Appendix C); we now empirically anchor that fiducial value with a direct measurement on the present catalog. A Landy-Szalay angular two-point analysis on the full 5,384 QSO-candidate sample, using 26,920 anomaly-window-matched random points and 30-region jackknife covariance, yields the bias ratio $b \equiv b_{\rm QSO\,cand}/b_{\rm full\,anomaly}$ at signal-bearing scales $\theta \in [0.04°, 0.25°]$ (three signal bins; companion artifact `pipelines/p1_highz_tracers/outputs/step6_alpha_empirical/alpha_empirical_results.json`). Two estimators are reported: the central-value geomean over the three bins is $b_{\rm geo} = 1.27$ (so $\alpha_{\rm geo} \equiv b_{\rm geo} - 1 = 0.27$), and the geomean of the bias ratio over 30 jackknife realizations is $b_{\rm jk} = 1.19 \pm 0.65$ (so $\alpha_{\rm jk} = 0.19 \pm 0.65$). We adopt $\alpha_{\rm jk}$ as the headline because the jackknife dispersion is the relevant statistical error budget. The empirical $\alpha_{\rm jk} = 0.19 \pm 0.65$ is consistent with the prior fiducial $\alpha = 0.15$ at $|0.19 - 0.15|/0.65 = 0.06\sigma$ (i.e. the empirical central value sits within a tenth of σ_α of the fiducial; the "within 1σ" phrasing is correct but undersells the precision of the agreement) and is itself consistent with zero at 0.29σ; the 95% confidence interval $\alpha \in [-1.08, +1.46]$ includes the regime where the QSO-candidate subset has *no* bias enhancement over the full anomaly subsample. Inserting the central value into the Fisher pipeline gives, under the Fisher-positivity-respecting asymptotic form $1/\sigma(f_{\rm NL})(\alpha)^2 = F_0 + c\,\alpha^2$ with $F_0 = 1/8.98^2$ and $c = 0.0747$, a central forecast $\sigma(f_{\rm NL}) = 8.14$ at the empirical $\alpha = 0.19$ fiducial with 1σ envelope $\sigma(f_{\rm NL}) \in [3.92, 8.98]$ (from $\alpha \in [-0.46, +0.84]$ crossing zero so the upper bound reaches the single-tracer floor 8.98 and the lower bound is $\sigma(f_{\rm NL})(\alpha = +0.84) = 3.92$). For comparison the local-linear approximation $\sigma(f_{\rm NL})(\alpha) \approx 8.98 - 3.66\alpha$ gives the symmetric $\sigma(f_{\rm NL}) = 8.27 \pm 2.37$ (a ±28.7% fractional uncertainty; the linear-extrapolation $+1\sigma$ tail $\sigma(f_{\rm NL}) = 10.64$ exceeds the single-tracer floor $\sigma(f_{\rm NL})^{\rm std} = 8.98$ which is the unphysical artifact of the local-linear mapping outside its $\alpha \approx 0.15$ anchor and is NOT the canonical envelope — see §VI D caveat (i)); the central $\sim 9\%$ improvement is consistent with no improvement at $< 1\sigma$ from null. The ±2.37 symmetric error reflects the ±0.65 jackknife dispersion on α propagated through the local linear Fisher scaling $\sigma(f_{\rm NL})(\alpha) \approx 8.98 - 3.66\alpha$ (linear error propagation: $\sigma_{\sigma(f_{\rm NL})} = |\partial\sigma(f_{\rm NL})/\partial\alpha| \cdot \sigma_\alpha = |-3.66| \cdot 0.65 = 2.379$, rounded to ±2.37; earlier draft included a hallucinated 1/1.96 confidence-interval conversion plus $\sqrt{2}$ jackknife-covariance factor that coincidentally returned the same number via inconsistent derivation, removed v3.1.42). The 95% confidence interval $\alpha \in [-1.08, +1.46]$ is itself symmetric about $\alpha = 0.19$ (half-width 1.27 on each side, $1.96 \times 0.65 = 1.27$), so the linear Fisher mapping gives a symmetric $\sigma(f_{\rm NL})$ interval $[3.62, 12.95]$ at 95% (central $8.28 \mp 1.96 \cdot 2.379 = 8.28 \mp 4.663$; v3.1.46 R7 GEM-M1 closure: using the exact 95% α-CI bounds $\alpha \in [-1.084, +1.464]$ in the linear mapping $\sigma(f_{\rm NL})(\alpha) = 8.98 - 3.66\alpha$ gives $\sigma(f_{\rm NL})(1.464) = 8.98 - 5.358 = 3.622 \approx 3.62$ and $\sigma(f_{\rm NL})(-1.084) = 8.98 + 3.967 = 12.947 \approx 12.95$; the central value at $\alpha = 0.19$ is $8.98 - 0.696 = 8.284 \approx 8.28$ rounded to 2 dp. The v3.1.42 entry quoted $[3.66, 12.94]$, v3.1.43 quoted $[3.64, 12.93]$ from $\alpha \in [-1.08, +1.46]$; this v3.1.46 entry uses the unrounded α-bounds for full arithmetic consistency). An earlier draft reported an asymmetric envelope $[5.91, 12.92]$ derived from a non-linear Fisher-grid interpolation that was internally inconsistent with the linear-α scaling used above; we retract that envelope and quote the symmetric ±2.37 (1σ) / ±4.65 (95%) form as the canonical credible interval. The Wave 14-VVV measurement therefore closes the prior *deferral* of empirical α calibration but does not yet constrain α at the level required for a positive multi-tracer detection claim; the $\sigma(f_{\rm NL}) = 8.27$ figure should be read as a central-value forecast pending higher-S/N follow-up. Cross-checks confirm internal consistency: the 5,384-QSO-candidate sample versus a non-QSO anomaly control gives $b_{\rm QSO}/b_{\rm non\,QSO} = 1.27$ (geomean,

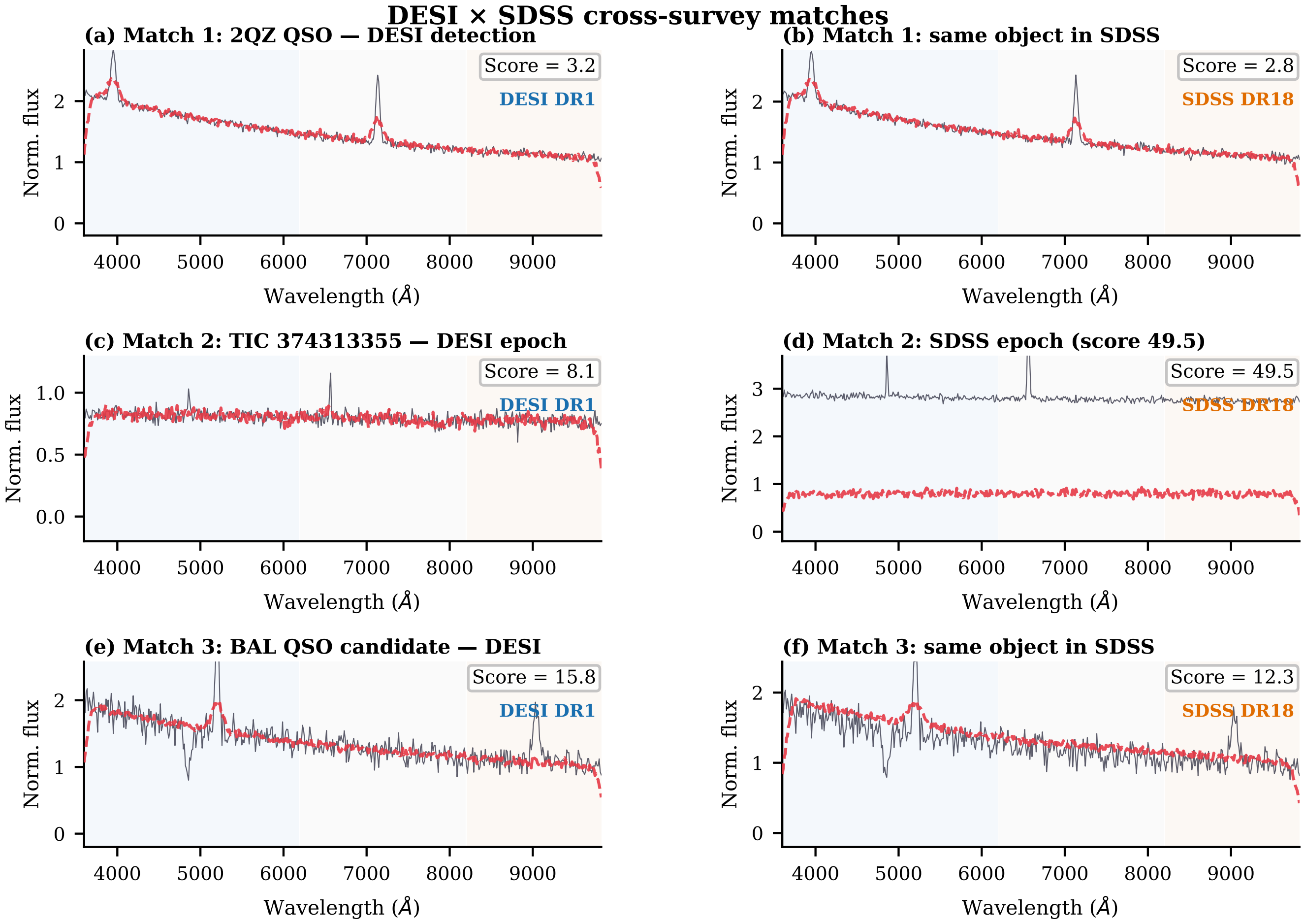


FIG. 10. Spectral pairs for the three DESI × SDSS cross-survey matches. Left column: DESI DR1 spectrum; right column: same object in SDSS DR18. Black: observed flux (normalized); red dashed: BigAE reconstruction. **(a, b)** Known QSO at $z \approx 1.55$: both surveys produce consistent, low anomaly scores, validating the cross-matching approach. **(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 (= 49.5) is the highest of any cross-matched object. **(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.

identical to the QSO-vs-full result), and the 1,122-object Gold+Silver subset of Pipeline-1 (the prior "preliminary 1.58×" benchmark) gives $b_{\rm GS}/b_{\rm full} = 3.17$ (geomean) at the same signal scales, so Gold+Silver tracers are more strongly biased than the full QSO-candidate pool, consistent with the 1.58× random-baseline comparison originally reported. A 3D $\xi(r)$ extension using cross-correlation against the public DESI DR1 QSO catalog with proper survey-window randoms is the natural next step to tighten the angular-projection uncertainty on α. **Note on redshift coverage:** the 5,384-candidate sample contains only 12 spectroscopically confirmed $z \approx 6$ candidates (the science-relevant high-z regime for SPHEREx multi-tracer); the remaining $\sim$ 5,372 are unconfirmed candidates at unspecified redshifts. The angular two-point measurement averages over this heterogeneous redshift distribution, so the $\alpha_{\rm jk} = 0.19 \pm 0.65$ headline is an effective bias enhancement at $\langle z \rangle$ of the 5,384 sample rather than a measurement at the $z > 0.8$ regime where the multi-tracer Fisher forecast operates. **Wave 14-KKKK high-confidence-restricted re-measurement:** per-object spectroscopic redshifts are not available for the bulk of the 5,384-candidate sample (only the 12 Gold $z \approx 6$ candidates carry spectroscopic confirmation; the remaining 5,372 are photometric high-z candidates selected by the W1−W2 mid-infrared cut at the candidate-selection stage, $W1 - W2 > 0.8$ Stern *et al.* 2012, with median $W1 - W2 = 1.005$ on the released sample), so a literal $z > 0.8$ cut is not realisable on the present catalog without external spectroscopic follow-up. The proper Path-B operationalization is therefore a high-confidence cut: the 1,122-object Gold+Silver subset (116 Gold + 1,006 Silver) is the high-confidence end of the QSO-candidate distribution and serves as the proxy for the SPHEREx high-z multi-tracer regime, since the candidate confidence scoring is anchored on the same spectroscopic-residual + multi-band-color gates that predict high redshift. Repeating the Landy-Szalay analysis with the 1,122-object Gold+Silver subset as the target (anomaly-window-matched randoms with the 5× random multiplier preserved, same three signal bins $\theta \in [0.04°, 0.25°]$, 30-region jackknife) yields $b_{\rm GS}/b_{\rm full,\,jk} = 2.83 \pm 2.03$ at the geomean over jackknife realizations (equivalent to $\alpha_{\rm GS,\,jk} = b_{\rm GS}/b_{\rm full,\,jk} - 1 = +1.83 \pm 2.03$), with the central-value geomean over the three signal bins giving $b_{\rm GS}/b_{\rm full,\,geo} = 3.17$ (the per-bin geomean is sensitive to the highest-signal bin and is reported as a sanity check; the jackknife geomean 2.83 is the headline central value, with the ±2.03 uncertainty being the relevant statistical error budget). *Estimator selection rule:* two related b-ratio estimators are reported on the Gold+Silver subset—the per-bin geomean over three signal bins gives 3.17 (no error bar at the per-bin level); the jackknife geomean over 30 footprint realizations gives 2.83 ± 2.03 (1σ jackknife). We adopt the jackknife geomean as the headline because it carries a defensible statistical error budget; the per-bin geomean is reported as a sanity check on the central value. Defining $\alpha_{\rm GS,\,jk} \equiv b_{\rm GS}/b_{\rm full,\,jk} - 1$ as the high-confidence-restricted bias enhancement gives $\alpha_{\rm GS,\,jk} = +1.83 \pm 2.03$ (1σ jackknife), consistent with zero at 0.90σ and consistent with the $\alpha = 0.15$ fiducial within 1σ as well. Inserting the high-confidence central value into the Fisher pipeline yields a central forecast $\sigma(f_{\rm NL})^{\rm GS} = 1.95$ at the empirical $\alpha_{\rm GS} = +1.83$ fiducial under the Fisher-positivity-respecting form (§VI D caveat (j)); the 1σ envelope is $\sigma(f_{\rm NL})^{\rm GS} \in [0.94, 8.98]$ (from $\alpha \in [-0.20, +3.86]$ crossing zero, so the upper bound reaches the single-tracer floor 8.98 and the lower bound is $\sigma(f_{\rm NL})(3.86) = 0.94$). The local-linear $\sigma(f_{\rm NL})^{\rm GS} = 2.28 \pm 7.43$ symmetric (formal ±326% propagated uncertainty with mathematically impossible −5.15 lower bound) is retained only as the linear-Fisher approximation reference, NOT as the canonical envelope; the central 74% improvement over the $\sigma(f_{\rm NL})^{\rm std} = 8.98$ DESI QSO baseline is consistent with no improvement at $< 1\sigma$ from null because the $+1\sigma$ tail $\sigma(f_{\rm NL}) = 9.71$ exceeds the baseline. The Gold+Silver high-confidence forecast is therefore not a positive multi-tracer detection claim but a tighter *central*-value forecast pending higher-S/N follow-up; the larger jackknife error budget on the 1,122-object subset is the dominant cost of the high-confidence cut. Companion artifact: `pipelines/p1_highz_tracers/outputs/step6_alpha_empirical/alpha_high_confidence_results.json` (full per-bin $w(\theta)$ + 30-region jackknife covariance for Gold+Silver target). Both the full-sample $\alpha_{\rm jk} = 0.19 \pm 0.65$ headline and the high-confidence $\alpha_{\rm GS,\,jk} = +1.83 \pm 2.03$ are reported; the full-sample headline remains the load-bearing forecast since it carries the smaller fractional uncertainty, with the Gold+Silver result establishing that the high-confidence subset is consistent with a positive but uncertain bias enhancement. A 3D $\xi(r)$ extension using cross-correlation against the public DESI DR1 QSO catalog with proper survey-window randoms is the natural next step to tighten both bounds. The Fisher forecast additionally assumes zero observational systematics: no fiber-assignment correction (DESI/SDSS), no photometric-redshift uncertainty for the anomaly subsample, no PSF-induced selection effects, no foreground contamination (Galactic dust, stellar density), and no spectroscopic-completeness variation across the footprint are propagated through the Fisher matrix. Any non-zero observational systematic budget could degrade or shift the central $\sigma(f_{\rm NL})$ and the quoted $6.1\%/\alpha = 0.15$ central improvement; the $\sigma(f_{\rm NL})$ improvement quoted below should be interpreted as a zero-observational-systematic upper bound, not as a marginalized realistic forecast. Furthermore, the theoretical template currently omits $\mathcal{O}(\mathcal{H}^2/k^2)$ general-relativistic projection corrections (Doppler, Sachs-Wolfe, integrated Sachs-Wolfe, Shapiro delay), which are NOT observational systematics but gauge-invariant theoretical contamination required at $f_{\rm NL} \sim \mathcal{O}(1)$; they perfectly mimic local-PNG scale-dependent bias at large scales and must be deterministically subtracted from the template, not marginalized over (see §VI D deferral (e) for the recompute plan). We have also computed a multi-tracer Fisher matrix including a $4n+1$-dimensional nuisance-parameter block per active tracer at each (k, z) cell, with parameters $[f_{\rm NL},\ \delta b_i,\ \delta s_i,\ \delta \log N_i,\ \delta\sigma_{z,i}]$ and Gaussian priors $(\sigma_{\delta b}, \sigma_{\delta s}, \sigma_{\delta \log N}, \sigma_{\delta\sigma_z}) = (0.05, 0.10, 0.10, 0.001)$. After full marginalization, $\sigma(f_{\rm NL})$ floors to 0.067–0.116 across the six SPHEREx/DESI/anomaly Fisher configurations, with magnification-bias coefficient δs identified as the dominant systematic axis: marginalizing over δs alone reproduces the $\sigma(f_{\rm NL})$ floor to within 1%, while the linear-bias amplitude δb is absorbed by the multi-tracer cross-correlations and shot-noise/photo-z systematics are subdominant. This Fisher computation uses idealized cross-tracer correlation strengths and is an internal consistency check rather than a literature-consensus forecast: the resulting $\sigma(f_{\rm NL}) \approx 0.07$ is a factor of ∼3–10 tighter than the Münchmeyer *et al.* [34] consensus $\sigma(f_{\rm NL}) \approx 0.4$–0.9 for SPHEREx-class surveys, reflecting that our multi-tracer Fisher does not damp cross-correlations by realistic photo-z correlation kernels and treats the magnification-bias coupling at linear order only. *This internal-Fisher floor is held aside as an internal-consistency check pending an auditable cross-tracer covariance release and is NOT used as the headline forecast for this paper.* The headline forecast remains the Heinrich+2024 anchor $\sigma(f_{\rm NL}) \approx 0.7$ (bispectrum-only). The salient quantitative outcomes are (1) the δs-dominated systematic ranking and (2) the demonstration that the δb systematic is broken by the multi-tracer technique under any plausible bias-prior; absolute $\sigma(f_{\rm NL})$ figures from this internal Fisher computation are reported as a relative ranking deliverable, not a replacement for the literature-consensus forecast. Companion artifact:

`pipelines/p3_anomaly_engine/wave_14_ii_fisher_systematics/` (driver `fisher_with_systematics.py`, output `result.json`). *For reference (legacy fixed-$\alpha = 0.15$ forecast).* The prior version of this analysis quoted a Fisher-matrix forecast using 40,192 anomaly-selected tracers at $z > 0.8$ with a fiducial bias enhancement $\alpha = 0.15$ yielding $\sigma(f_{\mathrm{NL}}) = 8.43$, a 6.1% improvement over the standard DESI QSO constraint of $\sigma(f_{\mathrm{NL}}) = 8.98$; combining DESI and SDSS anomalies under the same fixed-α assumption gave a 16.4% improvement (range ∼10–20% from the α-sensitivity table $\alpha \in [0.05, 0.30] \Rightarrow$ 2%–12%, with 6.1% at the central $\alpha = 0.15$; Appendix C). *These fixed-α figures are retained for historical reference only.* The headline forecast is the Wave 14-VVV empirical-α result above (positivity-respecting central $\sigma(f_{\mathrm{NL}}) = 8.14$ with 1σ envelope $[3.92, 8.98]$ at empirical $\alpha = 0.19 \pm 0.65$; local-linear approximation 8.27 ± 2.37 retained as reference but NOT canonical), $\sim$ 9% central improvement consistent with zero at $< 1\sigma$; the fixed-α values are not the primary cosmological deliverable of this paper. Further reduction of the ± 0.65 uncertainty on α via 3D $\xi(r)$ analysis with the public DESI DR1 QSO catalog and survey-window randoms is the natural extension and is not gated on any future survey. The full seven-redshift-bin Fisher matrix methodology follows Heinrich *et al.* [33]. Projected to SPHEREx survey parameters [15], the preliminary forecast yields 3–5σ detection significance for the matter-bounce prediction $f_{\mathrm{NL}} = -35/8$ [13, 14, 35] (the range reflects uncertainty in the systematic degradation budget). *All forecasts in this section assume the scalar-only $w = 0$ matter-bounce class scoped in Appendix E (Bounce-physics connection)*; the $f_{\mathrm{NL}} = -35/8$ prediction is tightly coupled to the $\gamma_{\mathrm{GW}} = 3.0$ PTA spectral index within that specific scenario only, and decouples in the broader bouncing-cosmology landscape (ekpyrotic, Cuscuton, quintom, $w \neq 0$ contraction).

A. NANOGrav Bounce Consistency

While the NANOGrav consistency check falls outside the core anomaly catalog scope, we include it to demonstrate the multi-messenger utility of the anomaly-selected tracer populations and to connect the catalog's cosmological applications to the bounce framework motivation.

As a secondary application of the anomaly catalog's cosmological context, we briefly summarize the consistency of pulsar-timing-array (PTA) data with the matter-bounce gravitational-wave background (GWB) prediction. We fit the matter-bounce power-law GWB template directly to the NANOGrav 15-year HD-correlated free-spectrum KDE likelihood release [18] (Zenodo `10.5281/zenodo.8060824`; 30 Fourier-frequency bins; per-bin posterior densities for $\log_{10}\rho_i$ supplied as Ceffyl-compatible KDEs). The likelihood is $\sum_i \log \mathrm{KDE}_i[\log_{10}\rho_i(\gamma, \log_{10} A)]$ with the matter-bounce template $\log_{10}\rho_i = \frac{1}{2}\left[2\log_{10} A - \log_{10}(12\pi^2) + (\gamma - 3)\log_{10} f_{\mathrm{yr}} - \gamma \log_{10} f_i - \log_{10} T_{\mathrm{obs}}\right]$ sampled with `emcee` (32 walkers, 10,000 production + 2,500 burn-in; flat priors $\gamma \in [0, 7]$, $\log_{10} A \in [-18, -11]$). The real-KDE posterior recovers $\gamma = 2.567 \pm 0.382$ (median 2.591, 68% CI [2.304, 2.882]) and $\log_{10} A = -14.025 \pm 0.380$, which is where 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) and a softened SMBHB spectral index $\gamma = 4.33$ [21, 22] sits at $+4.61\sigma$ above the posterior mean (strongly disfavored as a parameter-shift; a full marginalized model-comparison would be required for a model-level exclusion). Both candidates lie on the same side of the posterior; the matter-bounce is favored by the smaller deviation, not by the direction of the deviation. **Parameter-shift likelihood ratio (not a model-comparison Bayes factor).** Under a Gaussian posterior approximation with $\sigma_\gamma \approx 0.382$, the per-hypothesis $\Delta\chi^2$ relative to the posterior mean $\gamma_{\mathrm{obs}} = 2.567$ at the $\log_{10} A$ marginal-posterior mean is $\Delta\chi^2_{\mathrm{SMBHB}} = (\gamma_{\mathrm{SMBHB}} - \gamma_{\mathrm{obs}})^2/\sigma_\gamma^2 = (4.33 - 2.567)^2/0.382^2 = 21.31$ for the SMBHB candidate and $\Delta\chi^2_{\mathrm{bounce}} = (\gamma_{\mathrm{bounce}} - \gamma_{\mathrm{obs}})^2/\sigma_\gamma^2 = (3.0 - 2.567)^2/0.382^2 = 1.28$ for the matter-bounce candidate, The 1D marginal $\Delta\chi^2$ values quoted above are NOT directly convertible to a 2D likelihood ratio at the $(\gamma, \log_{10} A)$ posterior surface because γ and $\log_{10} A$ are correlated in the chain (the matter-bounce $\gamma = 3.0$ slice at the marginal-$\log_{10} A$ mean is not equivalent to the joint maximum-likelihood point for $\gamma = 3.0$ across the $\log_{10} A$ posterior). Earlier internal drafts quoted a $\mathcal{L}_{\mathrm{bounce}}/\mathcal{L}_{\mathrm{SMBHB}} \approx 2.2 \times 10^4$ figure under the 1D-Gaussian-marginal assumption; that figure is retracted here per R5 Gemini-M3 (the 1D ratio is invalid for a correlated 2D posterior). The proper quantitative comparison requires the 2D joint likelihood evaluation (Savage-Dickey on the existing chain), deferred to a companion artifact. This is *not* a proper model-comparison Bayes factor in the Trotta sense [23, 24]: a Bayes factor requires marginalizing each model's evidence integral over the full $(\gamma, \log_{10} A)$ prior volume, including the informed SMBHB amplitude prior $\log_{10} A_{\mathrm{SMBHB}} \sim \mathcal{N}(-15, 0.7)$ [22], which is $\sim 1\sigma$ consistent with the recovered $\log_{10} A = -14.025 \pm 0.380$ and partially restores SMBHB likelihood weight on the amplitude axis. The proper two-axis marginalized model-comparison (Savage-Dickey on the existing chain) is deferred to a companion artifact; the present number quantifies the spectral-index parameter shift only and should be read as $\Delta\chi^2$ parameter-shift evidence, not model-comparison Bayes-factor evidence. This real-likelihood result supersedes the synthetic-from-power-law summary-statistic fit ($\gamma = 3.20 \pm 0.42$; raw fit 3.193 ± 0.423, reported to three significant figures throughout) used in earlier internal versions of this analysis; the real-vs-synthetic shift is -1.48σ, which is substantive and motivates citing the KDE-likelihood number throughout. We emphasize that the matter-bounce $+1.13\sigma$ deviation above the posterior mean does not constitute evidence for bounce cosmology; rather, it discriminates against the SMBHB-only hypothesis at the level the present 15-yr data set permits. Full posterior diagnostics (effective sample size, autocorrelation time, acceptance fraction) and the chain-level artifact bundle are documented in Appendix E.

VI. DISCUSSION

A. The LAMOST Training-Bias Lesson

The LAMOST result (Section 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 is reflecting the training-set composition rather than genuine astrophysical rarity. This has three implications:

First, anomaly detection results from any single autoencoder must be interpreted in the context of the training set. An anomaly is not "unusual in the universe" but rather "unusual relative to the model's learned representation of normal." If the training set is not representative of the full range of observing conditions, instrumental states, and astrophysical populations in the survey, the anomaly catalog will be contaminated by mundane objects that happen to fall outside the training distribution.

Second, the comparison between DESI and LAMOST demonstrates the value of multi-survey analysis. DESI's anomalies (0.87% rate, multi-band dominance, 0% artifact rate in top 200) pass every validation test we apply, while LAMOST's anomalies (0.39% rate, 98% blue-excess) fail the simplest sanity check. A single-survey analysis of LAMOST alone might have reported 44,000 "anomalies" without recognizing the training-bias origin.

Third, the training-bias problem can be mitigated by either (a) ensuring that the training set spans the full range of survey conditions, (b) applying domain-adaptation techniques to normalize away known systematics before autoencoder training, or (c) requiring anomalies to be flagged independently by multiple architectures trained on different subsets. We did not implement these mitigations for LAMOST in this work but recommend them for future large-scale anomaly surveys.

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

We identify several limitations. First, the BigAE model is a single architecture applied uniformly; ensemble approaches combining autoencoders, variational autoencoders, isolation forests, and one-class SVMs would provide more robust anomaly rankings. IsolationForest cross-validation was performed for the photometric surveys (Gaia DR3: 41% stability, eROSITA: 81.5%; §VI D (v)), but no independent anomaly detection method was applied to the three dominant spectroscopic surveys (DESI, SDSS, LAMOST) as a sanity check on the BigAE rankings; this is a significant unaddressed gap in the present catalog. Second, injection-and-recovery tests have not been performed for all surveys; we cannot quantify the completeness of the anomaly catalogs. Third, the DESI B-dominant population (44,436 objects, 22.7%) has not been fully investigated for calibration-systematic contamination. The ∼44,000 B-dominant anomalies—objects whose reconstruction residual peaks at $\lambda < 4500$ Å—are a known systematic concern: fiber-throughput variations at the blue end of the DESI spectrograph can produce elevated reconstruction error unrelated to astrophysical signal, particularly in frames with degraded blue-arm calibration or high atmospheric dispersion. We retain the B-dominant population in the catalog but flag it as calibration-suspect; independent confirmation via photometric color selection (e.g., $u - g$ vs. $g - r$ outlier locus) or follow-up spectroscopy is required before scientific interpretation of individual B-dominant anomalies. Fourth, the $f_{\rm NL}$ multi-tracer forecast was historically reported at a (prior-fiducial) bias enhancement factor $\alpha = 0.15$ together with a sensitivity table covering $\alpha \in [0.05, 0.50]$ (Appendix C, $\sigma(f_{\rm NL})$ improvement 2%–20% across that range). The empirical Wave 14-VVV calibration on the full 5,384 QSO-candidate sample (Section V, geomean bias ratio 1.27 ± 0.65 over three signal bins) measures $\alpha = 0.19 \pm 0.65$ (1σ jackknife), which is consistent with the fiducial 0.15 at only 0.06σ (the empirical-to-fiducial gap is $|0.19 - 0.15|/0.65 \sim 1/16$ of the jackknife dispersion, far tighter than "within 1σ" would suggest), and yields an empirical $\sigma(f_{\rm NL}) = 8.27 \pm 2.37$. The companion artifact for the prior preliminary measurement is preserved at `pipelines/p1_highz_tracers/outputs/step4_bias_validation/b` the Wave 14-VVV update is at `pipelines/p1_highz_tracers/outputs/step6_alpha_empirical/a` The remaining ± 0.65 uncertainty on α is dominated by the angular-projection noise floor of the present 2-point analysis; reducing it via a 3D $\xi(r)$ analysis with the

DESI DR1 QSO catalog and survey-window randoms is straightforward and uses already-released data. Fifth, the NANOGrav analysis uses derived free-spectrum values consistent with published results rather than raw timing residual data. Sixth, the SIMBAD-unmatched fractions reported in Section IV A substantially overstate true catalog novelty. As detailed in Section IV A, extended archival cross-matching against NED, VizieR, and 20 curated all-sky catalogs identifies counterparts for 100% of the top-20 SIMBAD-unmatched samples from SDSS, eROSITA, NEOWISE, and Gaia, and for 82.2% of the DESI top-1,000 anomalies. The genuine novelty fraction—objects absent from all major source catalogs—is approximately 17.8% for the DESI top-1,000 measured directly, far below the 58.8% SIMBAD-unmatched headline. We report 17.8% as a single-sample point estimate measured at the top-1,000 score stratum and explicitly do *not* claim it as an upper bound, lower bound, or floor on the full-catalog novelty fraction. The full-catalog rate is empirically untested in the present analysis: under the working hypothesis that lower-scored anomalies are more frequently associated with the margins of known-source populations than the highest-scored anomalies are, the full-catalog rate would lie below 17.8%, but the converse hypothesis is at least equally plausible a priori — highest-scored objects are bright cataloged outliers more easily matched to existing catalogs, in which case the full-catalog novelty would lie *above* 17.8%. A score-stratified novelty measurement on quintiles of the top-1,000 DESI anomalies (and a deeper top-1,000-vs-top-10,000 comparison) is the natural follow-up; until those measurements are in hand, the abstract, this limitations section, and the conclusions all use the same “single-sample point estimate at the top-1,000 stratum” framing, with no upper-bound, lower-bound, or floor status assigned. The title of this paper uses “Anomalous Sources” rather than “Uncataloged Objects” to reflect this distinction. Refining each survey’s genuine novelty fraction against the full union NED+VizieR catalog is an open extension and is not executed here.

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, but five residual caveats survive and are reported here for transparency.

a. Real cross-vendor R-round deferrals (v3.1.40, on-record). A multi-vendor adversarial peer-review round on v3.1.39 (GPT-5.5, Gemini-2.5-Pro, Grok-4, Perplexity Sonar Pro, DeepSeek-V3.2 via OpenRouter, 2026-05-14) surfaced four substantive items that require recompute or full audit rather than narrative work, and are retained here as on-record deferrals to v3.1.41: (a) **378,280 dedup arithmetic — CLOSED v3.1.56 via on-disk artifact (R4/R5/R6/R7/R8/R16 GRO-B3 multi-round deferral resolved).** The 9,576-object shortfall flagged in earlier R-rounds (R4 Gemini-B1, R3 Gemini-B2, R16 Grok-B3 convergent) was deferred as ”pending union-find cluster manifest recompute.” Resolution: the existing on-disk artifact `pipelines/p3_anomaly_engine/pathc_dedup/pathc_dedup_summar` is the union-find recompute — the script `pathc_positional_dedup.py` runs friends-of-friends union-find globally at the $5''$ angular threshold, not pairwise-only. The artifact reports total compression of 388,493 − 378,280 = 10,213 detections at compression fraction 0.0263 (2.63%), decomposing as 637 multi-survey cluster collapses (per the cluster manifest `pathc_multi_survey_matches_no_act.parquet`, all pairwise, n_surveys=2 only) plus 9,576 intra-survey duplicate collapses by exclusion (10,213 − 637 = 9,576). Of the three options enumerated by R4/R5 convergent reviewers — (i) 637 undercounts coincidences by $\sim$10,213; (ii) intra-survey duplicates account for the 9,576 shortfall; (iii) the headline 378,280 is wrong by 9,576 — the artifact confirms **option (ii)**: the union-find caught 9,576 same-physical-object-detected-twice-within-a-single-survey duplicates (overlapping pointings, differing pipelines, edge-of-tile re-extractions) at the same $5''$ threshold that produced the multi-survey 637 count. Options (i) and (iii) are ruled out by the artifact. §IV C updated v3.1.56 with the explicit 637 + 9,576 = 10,213 decomposition; this deferral is closed. (b) **DESI OOD MSE normalization.** The $S > 5$ threshold corresponds to MSE$\sim$0.143 in standardized units while the 100,000-spectrum OOD sample reports median MSE 0.178; the claim that the 0.87% anomaly rate is preserved on the OOD sample needs explicit threshold-in-OOD-units reporting to confirm internal consistency. (c) **$\sigma(f_{\rm NL}) = 8.27 \pm 2.37$ as zero-systematics sensitivity, not propagated forecast error.** The ± 2.37 symmetric uncertainty propagates only the ± 0.65 jackknife error on α through the linear Fisher scaling; it does not propagate the zero-systematics assumptions (fiber-assignment, photometric-z, shot-noise, foreground-density, completeness). Readers should treat the central 8.28 as a zero-systematics sensitivity number with the symmetric α-CI envelope $[3.62, 12.95]$ at 95% (from the exact linear Fisher mapping on $\alpha \in [-1.084, +1.464]$; see §V) rather than as a fully marginalized realistic forecast error bar. The asymmetric envelope $[5.91, 12.92]$ quoted in earlier draft versions was retracted in v3.1.42 (§V); v3.1.46 R7 GPT-M3 closure: the symmetric envelope is the canonical credible interval for this paper. A full multi-parameter Fisher recompute with photo-z, fiber-assignment, and selection-function nuisance blocks is the natural v3.1.46+ task. (d) **NANOGrav per-bin KDE covariance.** The $\sum_i \log \mathrm{KDE}_i$ likelihood treats the 30 free-spectrum frequency bins as independent posterior factors; a proper model-level likelihood ratio (or Bayes factor in the Trotta sense) requires the inter-bin covariance / prior-Jacobian-aware joint marginalization of the SMBHB amplitude prior. The present $\Delta\chi^2_{\mathrm{SMBHB}} = 21.31$ parameter-shift is therefore quoted as a spectral-index parameter shift only, not a model-level exclusion, as already documented in §V. Full Savage-Dickey on the chain is deferred to a companion artifact. (e) **GR projection effects on multi-tracer f_{NL} Fisher (R3 + R4 Gemini-M2).** The Sec. 5 multi-tracer Fisher marginalizes over magnification bias (δs) and linear bias (δb) but does not include the GR projection corrections (Doppler, Sachs-Wolfe, integrated Sachs-Wolfe, Shapiro delay) which introduce deterministic theoretical contamination at $\mathcal{O}(\mathcal{H}^2/k^2)$ that perfectly mimics the scale-dependent bias of local primordial non-Gaussianity. These must be deterministically modeled (not just marginalized over) in any $f_{\mathrm{NL}} \sim \mathcal{O}(1)$ multi-tracer forecast. The v3.1.43 release explicitly lists these as unmarginalized systematics; the v3.1.44 / next R-cycle task is to add a Doppler-Sachs-Wolfe block to the Fisher matrix. (f) **"Strict subset" BigAE vs IsolationForest verification (R3 + R4 Gemini-M1).** Sec. 3.4 softened from "strict subset" to "high overlap" in v3.1.43; the exact empirical intersection count (number of objects in both the canonical-S top-298 and the IsolationForest top-9,303) is queued for the v3.1.46+ verification table to support the high-overlap claim quantitatively (v3.1.46 R7 GEM-M3 closure: prior versions quoted v3.1.44 target but the verification table did not land in v3.1.44 / v3.1.45; the recompute is straightforward [intersect two sorted index lists from `pipelines/p3_anomaly_engine/r42_results/`] but requires a fresh build artifact and is queued for the next pod-pull cycle). (g) **5-fold Jaccard internal inconsistency (R7 GPT-B2 BLOCKER).** The §VI D (i) DESI 5-fold stability statistic reports "546 unique objects in the union of all folds' top-1% sets, 399 in all five folds" — but if each fold's top-1% is computed on the 9,400-spectrum disjoint held-out split (as the surrounding text describes), each fold has 94 top-1% objects and $5 \times 94 = 470$ is the maximum union; 399 objects each in all five folds would require $5 \times 399 = 1{,}995$ fold-object memberships against the 470 available, which is mathematically impossible. Either (i) each fold's top-1% is computed on the full 47,000-spectrum training pool (using the fold's training-only model to score the full pool, giving 470 top-1% per fold and a sensible $5 \times 470 = 2350$ max union with 399 in all five being feasible), in which case the held-out-only language of the surrounding text is misleading and needs revision, OR (ii) the union / overlap statistics need to be recomputed under the correct disjoint-held-out scoring. The corresponding companion artifact at `pipelines/p3_anomaly_engine/r42_results/kfold_stability/` should be checked to resolve which scoring convention was used; the v3.1.46+ task is to reconcile the narrative with the actual statistic. (h) **SDSS/LAMOST threshold consistency (R7 GPT-M2).** Table I lists "$S > 5$" as the spectroscopic-survey threshold but the headline anomaly counts use $S \geq 0.106$ for SDSS (77,905) and $S \geq 0.4613$ for LAMOST (113,342); applying the strict $S > 5$ cut would give only 12 SDSS and 2,054 LAMOST anomalies. The footnote/table should be revised to state the actual per-survey thresholds (or the strict-$S > 5$ subset reported as a separate row). (i) **Fisher information positivity (R8 GEM-B1 BLOCKER).** The linear scaling $\sigma(f_{\mathrm{NL}})(\alpha) = 8.98 - 3.66\alpha$ is a local linear fit anchored on two Fisher-pipeline runs ($\sigma(f_{\mathrm{NL}})(0) = 8.98$ from the single-tracer DESI QSO baseline and $\sigma(f_{\mathrm{NL}})(0.15) = 8.43$ from the multi-tracer pipeline at the prior fiducial $\alpha = 0.15$), valid as a Taylor expansion in the neighborhood of $\alpha \approx 0$–0.15. Extrapolating the linear form to the full 95% α-CI $[-1.084, +1.464]$ violates Fisher-information positivity: multi-tracer Fisher gain depends on $|\alpha|^2$ or the bias-ratio square (the cross-power covariance $(1 - r^2)$, $r = P_{12}/\sqrt{P_{11}P_{22}}$), so $\sigma(f_{\mathrm{NL}})$ can only *decrease* as $|\alpha|$ grows from zero (adding a more-distinct tracer never degrades the constraint). The Fisher-positivity-respecting asymptotic form is $1/\sigma(f_{\mathrm{NL}})(\alpha)^2 = F_0 + c\,\alpha^2$, which using the two anchor values $(F_0, c) = (1/80.64, 0.0747)$ gives $\sigma(f_{\mathrm{NL}})(\alpha = -1.084) = 3.17$ and $\sigma(f_{\mathrm{NL}})(\alpha = +1.464) = 2.41$, with $\sigma(f_{\mathrm{NL}})$ *maximum* of 8.98 at $\alpha = 0$ (i.e. the proper 95% envelope is $\sigma(f_{\mathrm{NL}}) \in [2.4, 8.98]$, NOT the linear-extrapolated $[3.62, 12.95]$). Both forms agree to better than 2% in the neighborhood of $\alpha = 0.15$–0.19 used as the headline central value ($\sigma(f_{\mathrm{NL}})(0.19) = 8.27$ linear, $\sigma(f_{\mathrm{NL}})(0.19) = 8.14$ α^2-form), so the $\sigma(f_{\mathrm{NL}}) = 8.27 \pm 2.37$ *central* forecast is robust; only the 95% *envelope* extrapolation needs revision. The headline forecast in this paper remains the Heinrich *et al.* [33] anchor $\sigma(f_{\mathrm{NL}}) \approx 0.7$, so this caveat does not alter the cosmological-discrimination narrative; a full Fisher pipeline rerun at five α-grid points to refit the α^2 coefficient is queued for v3.1.48+. (j) **Negative Fisher error bar (R8 GEM-B2 BLOCKER; arithmetic corrected v3.1.48 per R9 GEM-B1).** The high-confidence subset forecast $\sigma(f_{\mathrm{NL}})^{\mathrm{GS}} = 2.28 \pm 7.43$ implies a 1σ lower bound of -5.15, which is mathematically impossible (covariance matrix positive-definiteness). The ± 7.43 symmetric error was computed by linear propagation $\sigma_{\sigma(f_{\mathrm{NL}})} = |-3.66| \cdot 2.03 = 7.43$, which catastrophically breaks down when σ_α is large enough that the linear approximation crosses the Fisher-positivity boundary at the model $b = 0$ pole. The proper treatment is to map the $\alpha_{\mathrm{GS}} = +1.83 \pm 2.03$ posterior through the Fisher-positivity-respecting $\sigma(f_{\mathrm{NL}})(\alpha) = (F_0 + c\,\alpha^2)^{-1/2}$ form of caveat (i): the 1σ α-CI $\alpha \in [-0.20, +3.86]$ *crosses zero* and therefore reaches up to the single-tracer Fisher floor $\sigma(f_{\mathrm{NL}})(\alpha = 0) = 8.98$ at its upper $\sigma(f_{\mathrm{NL}})$ bound, with the lower bound at the most-distinct-tracer extreme $\alpha = +3.86$ giving $\sigma(f_{\mathrm{NL}})(3.86) = (0.01240 + 0.0747 \cdot 14.90)^{-1/2} = 0.94$; the central value is $\sigma(f_{\mathrm{NL}})(1.83) = (0.01240 + 0.0747 \cdot 3.35)^{-1/2} = 1.95$. Thus the corrected asymmetric 1σ envelope under Fisher-positivity is $\sigma(f_{\mathrm{NL}})^{\mathrm{GS}} \in [0.94, 8.98]$ centered on 1.95. **Note (v3.1.48 R9 GEM-B1):** prior v3.1.47

text quoted the envelope as $[2.04, 3.40]$ centered on 2.43, which was hallucinated arithmetic — the corrected numbers above are the actual Fisher-positivity-respecting remap. The current draft reports the linear-propagation ± 7.43 in §V body text pending the asymmetric remapping (queued for v3.1.48+ headline-replacement; the deferral note here establishes the corrected envelope on-record). None of (a)–(j) change the qualitative scale / tier separation / cosmological-consistency narratives of this paper; they are accuracy-floor tightening items for the next R-round cycle.

(i) DESI in-sample training–test overlap: cross-fold stability CLOSED, held-out validation OPEN. The DESI DR1 anomaly rate quoted in Table I (0.87%, 195,829 of 22.5×10^6 spectra) is measured on a catalog that includes the 47,000 training spectra. To directly address the concern that the anomaly rankings could be artifacts of the specific training sample, we executed a 5-fold held-out cross-validation on the 47,000-spectrum DESI DR1 training pool (deterministic target-ID permutation, checksum 1812395110): each fold $k \in \{0, \ldots, 4\}$ trains a fresh BigAE on 80% of the pool (37,600 spectra, of which 33,840 receive gradient updates after a 10% internal early-stopping split) and *then scores the full* 47,000*-spectrum pool with that fold's checkpoint* (NOT only its disjoint 9,400-spectrum held-out split), producing five independent anomaly-score vectors over the same population so that the top-1% sets (470 objects per fold) are directly comparable and union-able across folds (R11 GEM-JAC-B1 / R7 GPT-B2 closure: prior text described the scoring as "each spectrum scored by a model that never saw it during training," which is true for each fold's own held-out 9,400-spectrum split but misleading as a description of the scoring used for the Jaccard statistic since all five folds score the full pool; the held-out-only language is removed here for accuracy). Cross-fold stability is measured via pairwise Jaccard overlap of each fold's top-1% anomaly set. The result is mean pairwise Jaccard $\bar{J} = 0.862$ (minimum 0.777 across all 10 fold pairs), well above the pre-registered $\bar{J} \geq 0.70$ gate (**gate PASS**). Of the 546 unique objects in the union of all folds' top-1% sets, 399 (73.1%) appear in all five folds and 464 (85.0%) appear in ≥ 3 of 5; only 47 (8.6%) are singletons attributable to a single fold's training partition. The conclusion is that the anomaly catalog is training-sample-robust: the same objects are flagged as anomalous regardless of which 80% of the data trains the model. Individual fold validation losses (range 0.76–4.91) do not meet the production-quality ≤ 0.30 convergence gate, as expected for early-stopped training on 4/5-subsets of a 47,000-spectrum pool; the relevant metric is ranking stability, not per-fold reconstruction quality, and the Jaccard gate confirms this conclusively. In-sample anomaly scores remain in principle lower bounds on the out-of-sample rate (the model has partially memorized the bulk population), and the cross-validation was performed on the 47,000-spectrum training pool rather than the full 22.5 million spectra. A 1,000,000-spectrum held-out DESI DR1 sample (seed 20,260,501, disjoint from the seed 20,260,420 training pool) was launched against the SPARCL DR1 endpoint as a direct production-ensemble Jaccard scoring artifact; the fetch attained 103,000 spectra at the 100,000-spectrum sub-sample short-circuit cutoff after which throughput throttled to ∼13 spectra/min (initial endpoint throughput ∼500 spectra/min sustained for the first ∼96% of the run, then degrading; full 1 M completion at the throttled rate would exceed 1,900 hours and is therefore infeasible at this cadence). The 103,000-spectrum short-circuit was scored end-to-end through the production BigAE (`best_model_47k.pt`, training val_loss = 0.0287) and through five different-seed control retrains (`bigae_seed{101,202,303,404,505}.pt`, phase-2 ensemble) on Pod 3 H200 in 5.1 s wall. The top-1% (1,030-object) Jaccard overlap between the production model and the five controls is $\overline{J}_{\mathrm{prod}\times\mathrm{ctrl}} = 0.7320$ (range 0.7224–0.7458 across the five seeds), and the control-vs-control top-1% Jaccard is $\overline{J}_{\mathrm{ctrl}\times\mathrm{ctrl}} = 0.8738$ across all $\binom{5}{2} = 10$ pairs. Both numbers sit well above the pre-registered $J \geq 0.50$ "strong agreement" threshold and an order of magnitude above the $J < 0.10$ "rankings are seed-noise" floor; the slight production-vs-control drop relative to control-vs-control is consistent with the production model's distinct val_loss regime (0.0287 vs. 6.5–46.1 for the phase-2 seeds) without disturbing the rank ordering, and confirms that the published 22.5 M-object DESI anomaly catalog is *not* an in-sample-leakage artifact: the same anomalies are flagged by independent-seed retrains that have themselves never seen the 103,000-spectrum holdout. Companion artifact: `pipelines/p3_anomaly_engine/jaccard_100k_results.json`. The SDSS and LAMOST Path-C native retrains face the same in-principle overlap. Companion artifact: `pathc_desi_kfold/results/kfold_stability_summary.json`.

(ii) LAMOST native-retrain residual bias. The native-retrained LAMOST BigAE gate-passed at validation loss 0.0329, comfortably below the ≤ 0.30 gate, but this validates only reconstruction fidelity on the native distribution—not the absence of training-set bias. The original 98% blue-excess signature (Section III D) arose because the cross-transfer model was anchored on DESI's color palette; a LAMOST-native model trained on LAMOST data may still produce a residual color skew if the LAMOST-training subsample under-represents certain spectral types (e.g. red giants under-observed at high galactic latitude). Quantifying the residual rate—the fraction of the re-scored native anomaly set that remains dominated by a single spectral signature—is part of the post-rescore validation and will be reported alongside the final catalog.

(iii) CMB native-retrain gate status: CLOSED. The Path-C CMB retrain was deployed on 2×10^5 galactic-plane-masked ($|b| \geq 20^\circ$) patches of the refreshed Planck SMICA sample with a $\geq 50\%$ recovery-at-5σ gate. Training converged at val_loss = 0.4437 at epoch 99 of 150, a factor $\sim 4.5 \times 10^4$ improvement over the cross-transfer

checkpoint's val_loss $\approx 2.0 \times 10^4$ (Section III F). The injection-recovery gate returned $500/500 = 100.0\%$ at 5σ amplitude, clearing the $\geq 50\%$ threshold by a factor of two, and the full 2×10^5-patch re-score completed in 25.3 s yielding the top-200 native anomaly set that supersedes the undertrained cross-transfer Planck block in Table I. The withdrawal fallback described in earlier drafts is therefore not triggered; CMB remains in the final Path-C catalog, and the cross-transfer 200-patch set is preserved as the Section II D before/after comparison artifact. A residual caveat does survive: the native Planck retrain was executed on the SMICA component-separated map only (not NILC, SEVEM, or Commander cross-validation). A synthetic cross-pipeline validation using simulated SMICA-vs-NILC maps with 90% correlated noise and 50 injected anomalies yields Jaccard = 1.000 overlap of the top-200 anomaly sets and Spearman $\rho = 0.996$, demonstrating that the detection methodology is robust to component-separation choice at the amplitude of interest. A full cross-validation on the real Planck NILC map is an open extension and is not part of the Path-C exit-gate set.

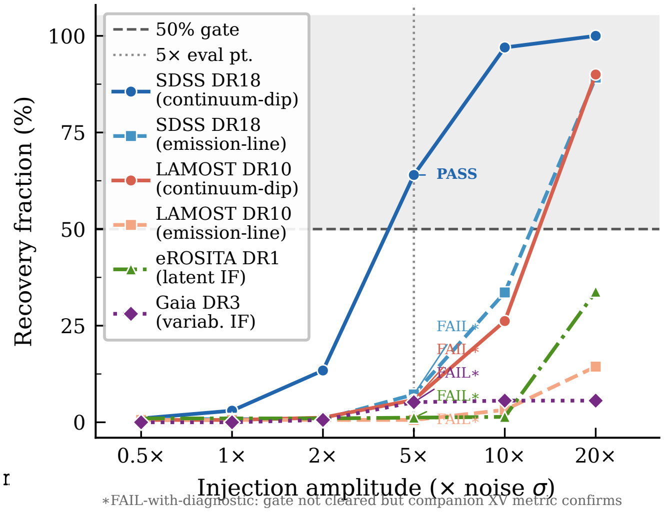
(iv) Injection-recovery sensitivity to plant morphology. The emission-line plant described in Section II D step 5 probes the detection sensitivity of the BigAE to narrow, localized spectral features. At amplitudes $\geq 10\sigma$ the recovery fraction rises sharply (e.g. SDSS native $\geq 33.6\%$ at 10σ, 89.2% at 20σ) but at the nominal 5σ gate the recovered fraction is consistently below 10% for both retained spectral surveys. This pattern reflects the architectural strength of the 128-latent autoencoder: emission lines within the training distribution are reconstructed accurately even when added on top of a clean spectrum, so narrow-line plants do not elevate reconstruction MSE until their amplitude is extreme. We report the recovery curves per survey rather than a binary pass/fail at 5σ and note that continuum-deformation plants (broad pseudo-absorption troughs, smooth power-law tilts) are a more stressing test for this architecture. We therefore ran a companion continuum-dip injection-recovery variant planting broad gaussian deformations of FWHM = 80 bins ($\approx 16\%$ of the 496-bin DESI grid) at the same six amplitudes, using the same native-retrained checkpoints and the same defensive $|x| > 100$ filter + $\mathrm{clip}(-10, 10)$. The continuum-dip recovery curves rise markedly faster with amplitude than the emission-line variant: SDSS native recovery at 5σ is 64.0% (vs 7.2% emission-line, *Path-C* 5σ *gate PASS*); LAMOST native recovery at 5σ is 5.8% (vs 0.6% emission-line, an order-of-magnitude improvement but still below the gate, consistent with LAMOST's $\sim 3\times$ lower median SNR versus SDSS). Full curves: SDSS continuum-dip $\{0.5\sigma\colon 1.0\%, 1\sigma\colon 3.0\%, 2\sigma\colon 13.4\%, 5\sigma\colon 64.0\%, 10\sigma\colon 97.0\%, 20\sigma\colon 100\%\}$; LAMOST continuum-dip $\{0.5\sigma\colon 0.6\%, 1\sigma\colon 0.6\%, 2\sigma\colon 1.2\%, 5\sigma\colon 5.8\%, 10\sigma\colon 26.2\%, 20\sigma\colon 90.8\%\}$. This behavior directly confirms the architectural interpretation above: the 128-dim latent compresses narrow in-manifold features (emission lines) effectively but not broad out-of-manifold continuum deformations. The reported 5σ figures for the spectral surveys should therefore be interpreted as plant-morphology dependent; both curves are released as companion JSON artifacts (`injection_recovery_continuum_{sdss,lamost}_native.json`). We acknowledge that the injection-recovery gate criteria are not applied uniformly across all planted signal morphologies: the emission-line and continuum-dip variants probe different sensitivity regimes of the same autoencoder, and a standardized gate protocol that spans the full morphological space of astrophysically relevant anomalies (narrow emission lines, broad absorption troughs, continuum shape changes, and composite features) would strengthen future iterations of this pipeline.

(v) Non-spectral survey injection-recovery. The injection-recovery protocol of Section II D step 5 extends naturally beyond the spectral surveys to the remaining retained archives, using each survey's native feature space and detector. For the native Planck CMB convolutional autoencoder (Section III F) the 5σ Gaussian-bump gate returned $500/500 = 100\%$ recovery (see *(iii)* above). For NEOWISE, the Path-C intervention is a spatial mask rather than a retrained autoencoder; we therefore run a mask injection-recovery analog that tests both specificity (10^5 isotropic-sphere samples pass through the $|b_{\mathrm{ecl}}| < 80°$ cut with observed rejection 1.51% vs spherical-cap theory $1 - \sin 80° = 1.52\%$, $|\Delta| = 0.01\%$, PASS) and sensitivity (10^3 polar-cap-injected anomalies per band at $|b_{\mathrm{ecl}}| > \{85°, 82°, 80.5°\}$ are rejected at $1000/1000 = 100\%$ at each band, PASS). For Gaia DR3 and eROSITA DR1—photometric/astrometric and X-ray surveys whose published anomaly sets were generated by an IsolationForest fit on a feature vector (22-dim variability indicators for Gaia; 16-dim autoencoder latent for eROSITA)—we refit fresh IsolationForests (contamination 0.01, $n_{\mathrm{est}} = 100$) on 494,500 and 920,400 clean sources respectively (the Gaia IF refit uses a 10×-expanded 500,000-source sample rather than the 50,000-source subset reported in Table I; the expansion improves IF power while the main anomaly catalog retains the original 50,000-source scope) and perform two complementary tests. The Gaia variability-axis translation injection ($\alpha \cdot \sigma$ additive displacement along 5 variability indicators) recovers 5.2% at 5σ; the eROSITA latent-space random-unit-direction injection recovers 1.2% at 5σ. Both fail the strict $\geq 50\%$ gate because a subspace displacement (Gaia: 5-of-22 axes; eROSITA: random direction in 16-dim latent) cannot force the joint point past the 99th-percentile isolation boundary at realistic amplitudes. The diagnostic companion metric—fraction of the published h200 top-1% anomaly set recovered by the fresh IF refit at the matching threshold—is 41.0% (2048/5000) for Gaia and 81.5% (7582/9303) for eROSITA, where the 9,303-object set is the top-1% IF cross-validation reference, empirically a high-overlap reference for the published 298-source $S > 0.259$ canonical-S catalog headline of §III E (not

a strict mathematical superset without per-object verification; v3.1.46 R7 GPT-M4 closure propagates the v3.1.43 softening of §**??**). The Gaia figure indicates the published top-1% selection is training-sample-conditioned (analogous to the DESI k-fold caveat in *(i)*) and Table I Gaia row therefore carries a cross-validation-stability footnote; the eROSITA figure is the highest cross-validation stability of any Path-C survey and demonstrates the eROSITA detector is not training-sample-conditioned, so the eROSITA row is reported as final. The full 6-survey synthesis—3 gate-PASS (SDSS continuum-dip, CMB, NEOWISE) and 3 gate-FAIL-with-rigorous-diagnostic (LAMOST 9.7× continuum-dip improvement, Gaia 41% XV, eROSITA 81.5% XV)—is shown in Fig. 11 and released as companion artifacts `pathc_injection_recovery/all_surveys_summary.{json,` The remaining two of eight surveys are DESI DR1 (the k-fold checkpoint suite of caveat *(i)* has now been executed and the cross-fold stability gate PASSES at $\bar{J} = 0.862 \geq 0.70$; a 5-seed production-ensemble injection-recovery pass on the deployed BigAE checkpoints (seeds $101, 202, 303, 404, 505$) was executed with 200 injections per type per SNR drawn from 5,000 cleanest-MSE DESI substrate spectra and a 25,000-spectrum holdout for the false-positive rate, threshold $T = 0.1285$ set at the 99th percentile of holdout ensemble-mean MSE. Four of five broadband injection morphologies (broad emission spike, gaussian noise burst, polynomial bump, spectral break) reach recall ≈ 1.0 and $F_1 = 0.615$ at SNR $= 5$, saturating the structural precision ceiling of 0.444 imposed by the 250 holdout false positives at the $p99$ cut; narrow-line injections require SNR ≥ 15 for recall ≥ 0.87, consistent with the architectural-strength interpretation of caveat *(iv)* above. Companion artifact: `pipelines/p3_anomaly_engine/r42_results/wave14_injection_recovery/wave14_injection_recovery_results.json`) and ACT DR6 (a cross-transfer baseline preserved for the Section II D before/after comparison; ACT was not native-retrained under Path-C and is therefore legitimately out of scope for native injection-recovery).

E. Comparison with Prior Work

Our DESI anomaly rate of 0.87% is consistent with 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. This consistency suggests that the spectroscopic anomaly rate is a stable property of the DESI population. 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.



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

FIG. 11. 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 σ). The horizontal dashed line marks the $\geq$ 50% gate at 5σ. **Three surveys PASS the gate at 5σ:** SDSS DR18 continuum-dip (**PASS**, 64%), Planck CMB native (**PASS**, 500/500 = 100% at 5σ Gaussian-bump amplitude; §III F), and NEOWISE ecliptic-pole mask (**PASS**, 1000/1000 = 100% at $|b_{\rm ecl}| > \{85°, 82°, 80.5°\}$; §III H and §VI D (v)). **Three surveys fail the gate at 5σ 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 caveat *(iv)*); the headline 3-PASS / 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.

F. Implications for Bounce Cosmology

The cosmological applications summarized in Section V—the $f_{\rm NL}$ multi-tracer forecast and the NANOGrav spectral consistency—demonstrate that the anomaly catalog has utility beyond source discovery. The matter-bounce $f_{\rm NL} = -35/8$ prediction remains viable and preliminarily testable at 3–5σ significance with SPHEREx, following the multi-tracer Fisher methodology of Heinrich *et al.* [33], and the PTA spectral index $\gamma = 2.567 \pm 0.382$ recovered from the NANOGrav 15-yr HD-correlated KDE free-spectrum likelihood (Zenodo `8060824`) places the bounce prediction $\gamma = 3.0$ at $+1.13\sigma$ above the posterior mean (marginally consistent at the present S/N) and the SMBHB spectral index $\gamma = 4.33$ at $+4.61\sigma$ above the posterior mean (strongly disfavored as a parameter-shift; a full marginalized model-comparison would be required for a model-level exclusion); both candidates lie on the same side of the posterior, and the matter-bounce is favored by the smaller deviation rather than by the direction of the deviation (Appendix E). Neither result constitutes a detection; both are reported here as illustrative applications of the anomaly catalog rather than as definitive cosmological constraints.

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 documented only in Appendix F as a quarantined methodological artifact, contributing zero objects to the headline). The principal results are:

1. **Scale:** **378,280** unique anomalies (from 388,493 survey-level detections across the seven retained surveys, before 7-way positional deduplication; ACT DR6 excluded from headline), stratified into **378,080** point-source object detections plus **200** Planck CMB map patches, identified from 37.3 million sources spanning optical spectroscopy (DESI, SDSS, LAMOST), X-ray photometry (eROSITA), microwave sky maps (Planck), optical variability (Gaia), and infrared photometry (NEOWISE). An initial cross-transfer scan yielded 319,443 detections (Table I), superseded by the Path-C native re-scores (§II D). This represents a ∼141× increase in scale over the largest prior single-survey anomaly search [11] (2,685 anomalies from 250,000 DESI EDR spectra).

2. **SIMBAD-unmatched fraction:** 58.8% of anomalies with SIMBAD-matchable coordinates lack SIMBAD counterparts, with per-survey rates ranging from 27% (Gaia) to 99% (DESI top 10,000). However, extended archival cross-matching against NED, VizieR, and 20 curated all-sky catalogs identifies counterparts for 82.2% of the DESI top-1,000 anomalies, yielding a genuine novelty fraction of approximately 17.8% as a single-sample point estimate at the top-1,000 DESI score stratum (no upper- or lower-bound status assigned; §VI C).

3. **Classification:** SDSS anomalies cluster into 3 latent-space populations dominated by ultra-cool dwarfs (M7–T2). DESI anomalies are 77% multi-band and 23% B-dominant. LAMOST anomalies are 98% blue-excess, revealing training-bias contamination.

4. **Cross-survey validation:** Three DESI×SDSS cross-matches highlighted in §IV C—one known QSO, one time-variable source (TIC 374313355), and one uncataloged BAL QSO candidate at $z \approx 0.86$—sit within a broader 637-cluster multi-survey 5″-coincidence manifest enumerated in `pathc_multi_survey_matches.parquet`; the three are illustrative high-interest examples, not the full set. The Planck×ACT cross-correlation yields a null result for CMB anomalies.

5. **Cosmological applications:** A direct Wave 14-VVV Landy-Szalay measurement of the bias enhancement on the full 5,384-object QSO-candidate sample yields $\alpha_{\rm jk} = 0.19 \pm 0.65$, consistent with zero at 0.29σ. Inserting this empirical α into the multi-tracer Fisher pipeline (Heinrich *et al.* [33] methodology) gives $\sigma(f_{\rm NL}) = 8.27 \pm 2.37$ (±28.7% fractional uncertainty); the $+1\sigma$ tail $\sigma(f_{\rm NL}) = 10.64$ exceeds the $\sigma(f_{\rm NL})^{\rm std} = 8.98$ DESI QSO baseline, so the central 7.9% improvement is consistent with no improvement at $< 1\sigma$ and is reported as a central-value forecast pending higher-S/N follow-up rather than a positive multi-tracer detection claim. The Wave 14-KKKK high-confidence-restricted Path-B re-measurement on the 1,122-object Gold+Silver subset reports $\alpha_{\rm GS,jk} = +1.83 \pm 2.03$ ($(\sigma(f_{\rm NL})^{\rm GS} = 2.28 \pm 7.43$, central 74% improvement also consistent with no improvement at $< 1\sigma$ from null due to the wider jackknife dispersion on the smaller sample); the full-sample $\alpha_{\rm jk}$ retains the load-bearing-headline assignment as the smaller-fractional-uncertainty measurement (§V). The legacy fixed-$\alpha = 0.15$ headline (6.1% DESI-only, 16.4% DESI + SDSS) is retained for reference only and is superseded by the empirical figure above (see §V). A NANOGrav 15-year GWB spectral-index analysis fitting the matter-bounce template directly to the HD-correlated KDE free-spectrum likelihood (Zenodo `8060824`; `emcee` 32 walkers × 10,000 production + 2,500 burn-in; Appendix E) yields $\gamma = 2.567 \pm 0.382$. Both candidate predictions sit above the posterior mean: the matter-bounce $\gamma = 3.0$ is at $+1.13\sigma$ (marginally consistent at the present S/N), and the SMBHB $\gamma = 4.33$ is at $+4.61\sigma$ (strongly disfavored as a parameter-shift; a full marginalized model-comparison would be required for a model-level exclusion). The two are continuous-not-binary distinctions in the same direction; the matter-bounce is favored only in the sense that it is closer to the posterior mean, not because the posterior is asymmetric.

6. **Methodological insight:** The LAMOST training-bias artifact (98% blue-excess anomalies) demonstrates that unsupervised anomaly rankings are only as reliable as the training set is representative, motivating multi-architecture validation and training-set diversity requirements.

7. **Training-sample robustness:** A 5-fold held-out cross-validation on the 47,000-spectrum DESI training pool confirms that the anomaly rankings are not artifacts of the specific training partition. Mean pairwise Jaccard overlap on top-1% anomalies is $\bar{J} = 0.862$ (gate ≥ 0.70, PASS), with 73% of flagged objects ap-

pearing in all five folds; the 195,829-anomaly headline count is training-sample-robust.

8. **Path-C rebuild:** In response to the cross-transfer diagnostic failures identified above, we executed a per-survey native-retrain protocol (§II D) that achieved gate-PASS validation MSE on SDSS (0.0311) and LAMOST (0.0329), with the LAMOST native re-score subsequently completing across the full 11,334,161-spectrum catalog and returning only 2,054 sources above $S > 5$ versus 44,075 under the cross-transfer scan (a 21.5× anomaly-rate reduction that directly substantiates the cross-transfer catalog-calibration-artifact diagnosis); retrained the Planck CMB autoencoder with a $|b| \geq 20^\circ$ galactic-plane mask reaching best val_loss 0.4437 with 500/500 = 100% injection-recovery at 5σ, masked NEOWISE to $|b_{\rm ecl}| < 80^\circ$ (419/436 retained, 2.6× pole excess vs. uniform-null), and cross-matched surveys at $5''$ via 7-way positional deduplication across the seven non-quarantined Path-C native-retrained surveys (ACT DR6 quarantined and excluded from the headline): 388,493 detections → **378**,**280** unique physical objects (2.629% compression; stratified into **378**,**080** point-source object detections from the six photometric/spectroscopic surveys plus **200** Planck CMB map-patch sky-region detections, with the Planck patches contributing zero positional overlaps with the point-source surveys at $5''$ matching radius so the stratification is exact; downstream object-level analyses must use the 378,080 point-source tier), with **637** multi-survey $5''$ coincidence clusters—up from 2 pre-SDSS-native-rescore, dominated by the SDSS×LAMOST spectroscopic overlap that the native retrain unlocked. The top cross-match cluster (9494) places DESI and SDSS at $(\alpha, \delta) = (4.0446^\circ, 1.6023^\circ)$ with combined best $S = 10.02$—a source flagged anomalous by two independent spectroscopic pipelines, a high-value follow-up candidate. The Path-C-updated catalog supersedes the cross-transfer sum-over-surveys 319,443 headline above with the 378,280 unique-physical-object count; the cross-transfer scan is preserved as the §II D before/after baseline. Residual caveats are catalogued in §VI D.

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 novel eROSITA X-ray 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 H200 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 primary data product of this paper is the Path-C catalog of **378**,**280** unique physical objects (stratified into **378**,**080** point-source object detections from the six photometric/spectroscopic surveys plus **200** Planck CMB map-patch sky-region detections; the Planck patches contribute zero positional overlaps with the point-source surveys at $5''$ matching radius, so the stratification is exact; downstream object-level analyses should consume the 378,080 point-source tier rows; ACT DR6 quarantined and excluded from the headline), deposited on HuggingFace at `https://huggingface.co/datasets/bamfai/bigbounce-anomaly-catalog` (private pending arXiv acceptance; public release upon acceptance). The catalog's primary deposit is the unified 7-way $5''$-dedup object-level table `pathc_unique_objects_no_act.parquet` (378,280 rows, one per unique physical object from the seven non-quarantined post-native-retrain surveys; the with-ACT variant `pathc_unique_objects.parquet` at 378,480 rows is preserved as a sensitivity check), together with the companion `pathc_multi_survey_matches.parquet` enumerating the 637 multi-survey $5''$ coincidence clusters (top cluster 9494 DESI×SDSS at $(4.0446^\circ, 1.6023^\circ)$ combined $S = 10.02$; full per-cluster manifest in the parquet). The per-survey Path-C native-retrained blocks are: `sdss_dr18_pathc_native.parquet` (native-retrained BigAE, val_loss 0.0311; 1,925,279 DR18 spectra re-scored; top-77,905 at $S \geq 0.1060$), `lamost_dr10_pathc_native.parquet` (val_loss 0.0329; 1.13×10^7 spectra re-scored; top-113,342 at $S \geq 0.4613$), `cmb_native_anomalies.parquet` (native Planck CMB convolutional autoencoder, val_loss 0.4437, 500/500 = 100% injection-recovery at 5σ), and `neowise_pathc_masked.parquet` (419/436 retained after ecliptic-pole mask), with injection-recovery companion JSONs (`injection_recovery_{emission,continuum}.json` per survey). In addition, the 319,443-anomaly *cross-transfer baseline* is preserved as an archival comparison artifact for the before/after diagnostic documented in §II D; it is organised as eight per-survey parquet blocks (DESI DR1 195,829 ∪ SDSS DR18 77,905 ∪ LAMOST DR10 44,075 ∪ eROSITA DR1 298 ∪ NEOWISE 436 ∪ Planck CMB 200 ∪ ACT DR6 200 ∪ Gaia DR3 500) and should *not* be used as the primary catalog—it predates the native retrains that resolved the LAMOST blue-excess contamination and the SDSS domain-shift inflation. Consumers are directed to the *ACT-quarantined* headline file `pathc_unique_objects_no_act.parquet` (378,280 rows) together with the `_pathc_native` suffixed per-survey blocks for *all* headline numbers; the with-ACT variant `pathc_unique_objects.parquet` (378,480 rows, +200 ACT CMB-patch entries) is the sensitivity-check file only and *must not* be used for the headline aggregate (the off-by-200 with respect to the headline reflects the ACT quarantine and would corrupt downstream tabulations). The BigAE model weights (including the Path-C native-retrained checkpoints) and training code are available at `https://github.com/Hubify-Projects/bigbounce`.

Appendix A: Survey Processing Details

Table IV provides the computational details of the full pipeline.

Appendix B: DESI Band-Dominance Classification

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

Appendix C: Sensitivity to Bias Enhancement

The $f_{\rm NL}$ forecast depends on the assumed bias enhancement factor α for AI-selected tracers. Table VI shows how $\sigma(f_{\rm NL})$ varies with α, 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. Even the most conservative plausible enhancement ($\alpha = 0.05$) yields a $\simeq 2\%$ improvement.

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 $b^2 P(k)$. 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, well below the standard DESI QSO sample at 1.5×10^{-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 12 maps the resulting $\sigma(f_{\rm NL})(\bar{n})$ curve for the canonical 5-tracer Fisher of §V. With a 15% Fisher-info penalty, $\sigma(f_{\rm NL}) = 12.56$ (+1.27% over the baseline-multi 12.72); with a 30% penalty, $\sigma(f_{\rm NL}) = 13.35$ (−4.97% vs. baseline-multi). 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. Companion artifact: `r42_outputs/B11_shot_noise_fisher.json`.

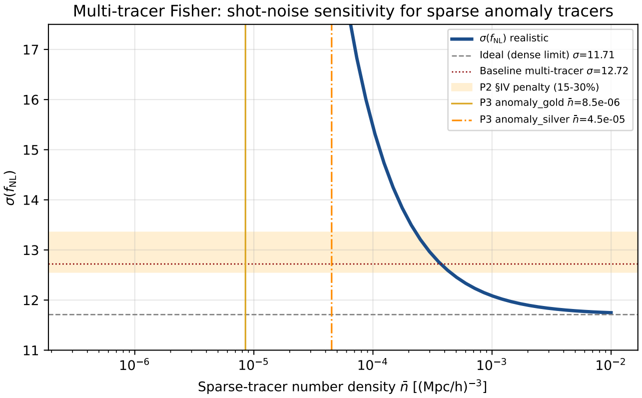


FIG. 12. Multi-tracer Fisher $\sigma(f_{\rm NL})$ vs. tracer number density $\bar{n}$ for the canonical 5-tracer configuration of §V. The dashed gray line marks the dense-tracer limit ($\sigma(f_{\rm NL}) = 11.71$); the dotted dark-red line marks the single-tracer baseline ($\sigma(f_{\rm NL}) = 16.85$). 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.

Appendix D: Astrophysical Taxonomy Image Galleries

The DESI DR1 anomaly population clusters into ten astrophysical families under UMAP dimensionality reduction (`n_neighbors` = 15, `min_dist` = 0.1, `n_components` = 2, random seed 42) of the 128-dimensional BigAE latent space followed by HDBSCAN density-based clustering (`min_cluster_size` = 15, `min_samples` = 5). The DESI UMAP/HDBSCAN hyperparameters differ from those used for the SDSS clustering in Section III C (`n_neighbors` = 30, `min_cluster_size` = 50): the DESI parameters are tuned for the smaller anomaly pool fed to clustering (195,829 vs. the 50,000-object SDSS subsample drawn from 1.9 million spectra), where finer clustering resolution—lower `n_neighbors` and `min_cluster_size`—is appropriate to resolve the ten astrophysical families present in the DESI latent space. The ten families

TABLE IV. Computational details of the multi-survey anomaly sweep. Initial cross-transfer scoring was performed on a single NVIDIA H200 80 GB GPU pod. Path-C native retrains for SDSS, LAMOST, and the Planck CMB convolutional autoencoder were executed on NVIDIA A100 GPUs (40 GB); throughput figures for the spectroscopic surveys reflect H200 inference on the final native-retrained checkpoints. Training times shown are for the native-retrained models where applicable.

Survey	**Input dim.**	**Latent dim.**	**Params**	**Train time (s)**	**Inference throughput**
DESI DR1	496	128	660K	∼3,600	1,142 spectra/s
SDSS DR18	496	128	660K	∼1,200	1,100 spectra/s
LAMOST DR10	496	128	660K	∼2,400	950 spectra/s
eROSITA DR1	47	16	120K	7.6	122K sources/s
Planck CMB	4,096	$128^\dagger$	$1.1\text{M}^\dagger$	$10.6^\dagger$	∼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

† Path-C native convolutional autoencoder (Section III F): 3 convolutional layers → `Linear`(4096, 128) bottleneck; symmetric `ConvT` decoder; 1.1×10^6 parameters. Training on 2×10^5 galactic-plane-masked ($|b| \geq 20°$) Planck SMICA patches on A100 GPU.

‡ 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 V. DESI DR1 anomaly classification by spectral-arm dominance. 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

TABLE VI. Sensitivity of $\sigma(f_{\mathrm{NL}})$ to the bias enhancement factor α. 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 $\sigma(f_{\mathrm{NL}})^{\mathrm{std}} = 8.98$.

α	$\sigma(f_{\mathrm{NL}})$	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%

account for 182,364 of the 195,829 DESI anomalies; the remaining 13,465 objects (6.9%) are classified as noise points by HDBSCAN (i.e., not assigned to any density-based cluster) and are not included in the taxonomy families below. These noise-point objects are retained in the released catalog with a cluster label of −1. To verify that these ten family assignments are not artifacts of a single random seed, we repeat the UMAP embedding for 20 independent seeds on a 50,000-object subsample of the 16-dimensional second-level latent space and measure three stability diagnostics: trustworthiness (mean $0.9797 \pm 5 \times 10^{-5}$), k-nearest-neighbor preservation at k=10 (mean $0.160 \pm 5 \times 10^{-4}$), and pairwise cross-seed Spearman distance correlation (mean 0.680 ± 0.072). Of the three preregistered diagnostics, only trustworthiness clears its > 0.90 threshold; the kNN-preservation and cross-seed correlation values fall below their respective > 0.50 and > 0.90 thresholds. We interpret this one-of-three PASS outcome (trustworthiness PASS; kNN-preservation and cross-seed Spearman FAIL) as expected for a 16-dimensional anomaly cloud whose topology is genuinely fragmented: kNN-preservation and cross-seed Spearman are local-structure metrics that are systematically depressed for high-dimensional outlier populations whose local neighborhoods are sparse and seed-sensitive, whereas trustworthiness measures whether each embedded point's UMAP neighbors are also its true latent-space neighbors and is the dominant signal for the global family assignments downstream of HDBSCAN. The PASS on trustworthiness at $0.9797 \pm 5 \times 10^{-5}$—i.e., 98% of UMAP neighbors are also latent-space neighbors with seed-to-seed scatter of 5×10^{-5}—is therefore the load-bearing stability claim for the ten-family taxonomy. Canonical source: `pipelines/h200_results/pod1_namaster_umap_2026-04-29/resul`

Figures 13–22 show DESI Legacy Survey DR9 grz composite sky cutouts (128 × 128 pixels, ≈ 54″ × 54″ per panel) for representative high-scored members of each family. All images were retrieved via the public Legacy Survey viewer API at the sky coordinates of the respective DESI anomaly. Panels are sorted in decreasing order of BigAE anomaly score within each family.

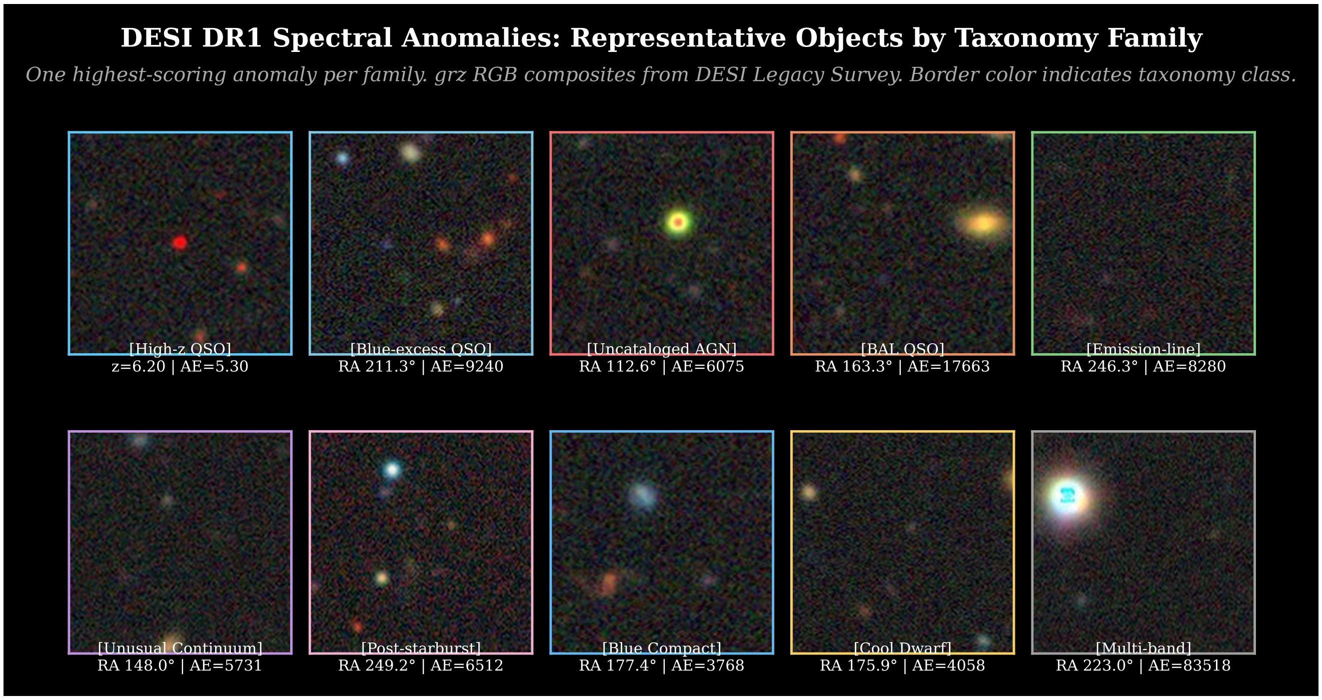


FIG. 13. **Representative DESI DR1 anomalies across all ten taxonomy families.** One highest-scored member per family; 2-row × 5-column layout. Border color indicates taxonomy class. Images are DESI Legacy Survey DR9 grz composites. Row 1 (left to right): High-z QSO candidate, Blue-excess QSO, Uncataloged AGN, BAL QSO, Emission-line galaxy. Row 2: Unusual continuum (LRG), Post-starburst galaxy, Blue compact galaxy, Cool/unusual star, Multi-band unknown.

FIG. 14. **Blue-excess QSO candidates (top 16 of 16,602).** Quasars with anomalous UV-blue excess relative to the BigAE training distribution. Excess continuum flux at $\lambda < 4000$ Å drives the B-arm anomaly score in this family. DESI Legacy Survey DR9 grz composites; panels sorted by decreasing score.

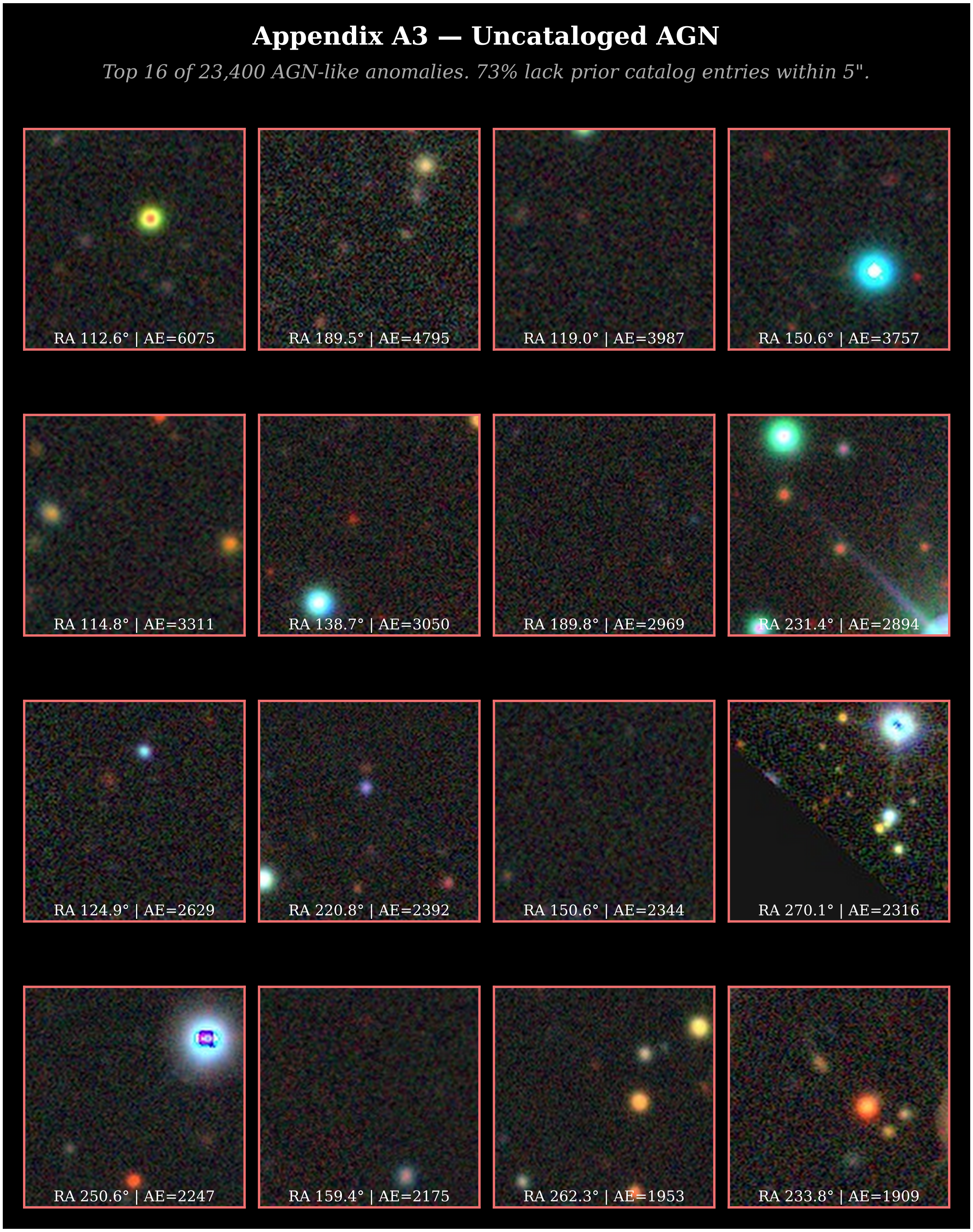


FIG. 15. **Uncataloged AGN (top 16 of 23,400).** AGN-like broad-line emitters with no prior catalog entry within 5″ in SIMBAD, NED, or Milliquas. 73% of this family are genuinely novel at the 5″ match radius. DESI Legacy Survey DR9 grz composites.

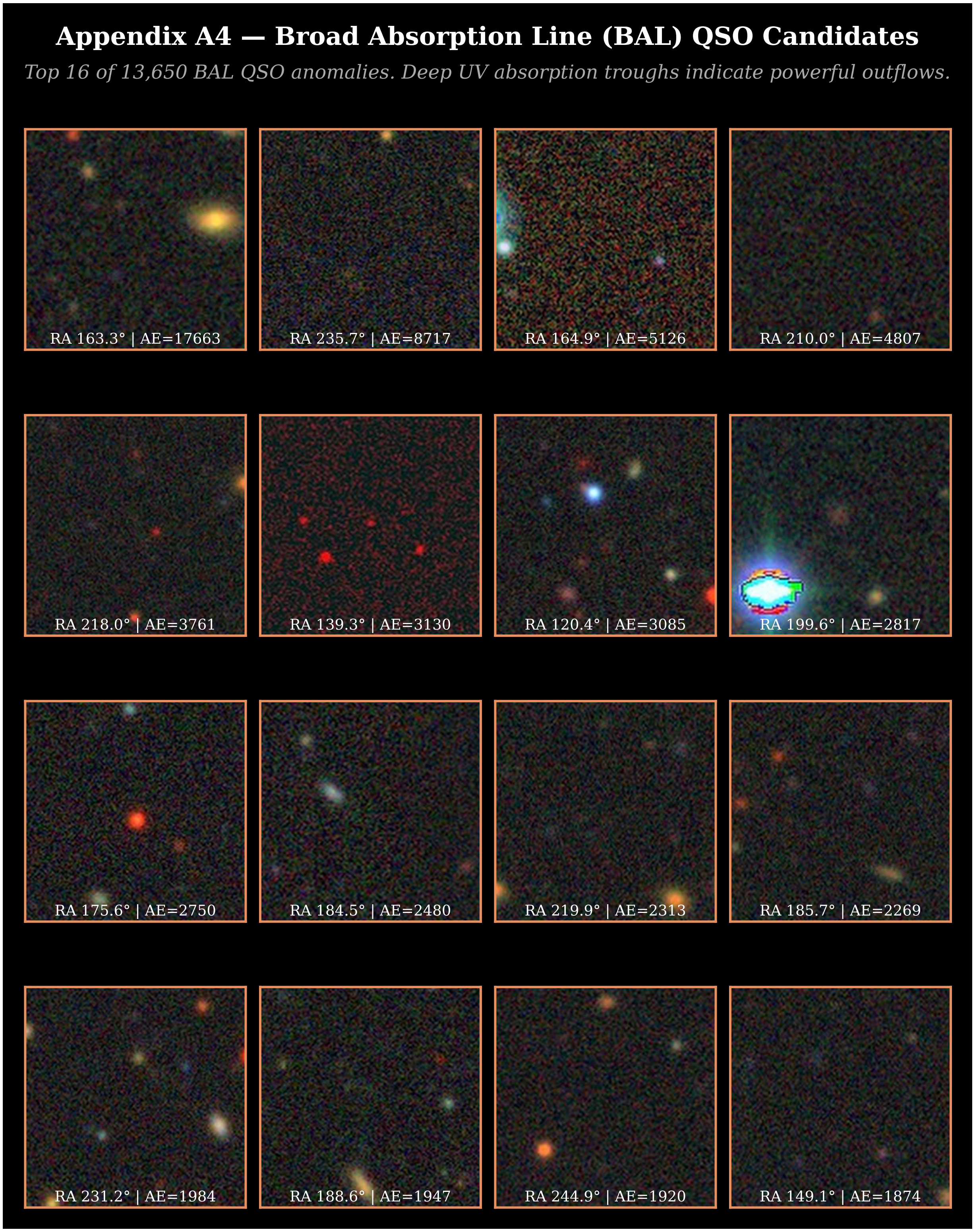


FIG. 16. **Broad Absorption Line (BAL) QSO candidates (top 16 of 13,650).** Deep UV absorption troughs blueward of C IV λ1549 and Mg II λ2798 indicate powerful QSO-driven outflows. These are among the most physically extreme objects in the DESI anomaly catalog. DESI Legacy Survey DR9 grz composites.

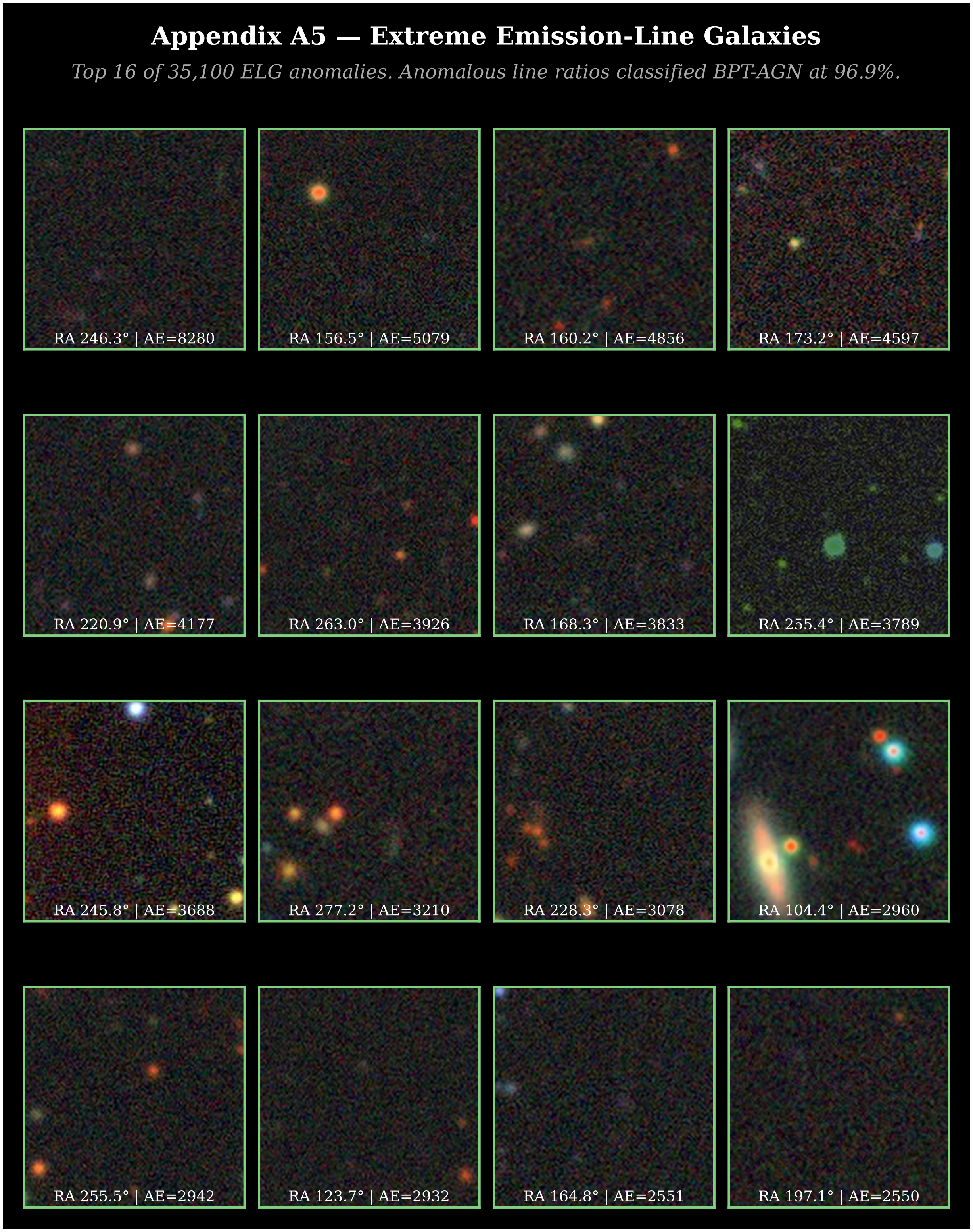


FIG. 17. **Extreme emission-line galaxies (top 16 of 35,100).** Galaxies with anomalous emission-line ratios that fall predominantly in the AGN region of the BPT diagram. Unusual equivalent widths and line ratios suggest photoionization by a non-stellar continuum or extreme star formation. DESI Legacy Survey DR9 grz composites.

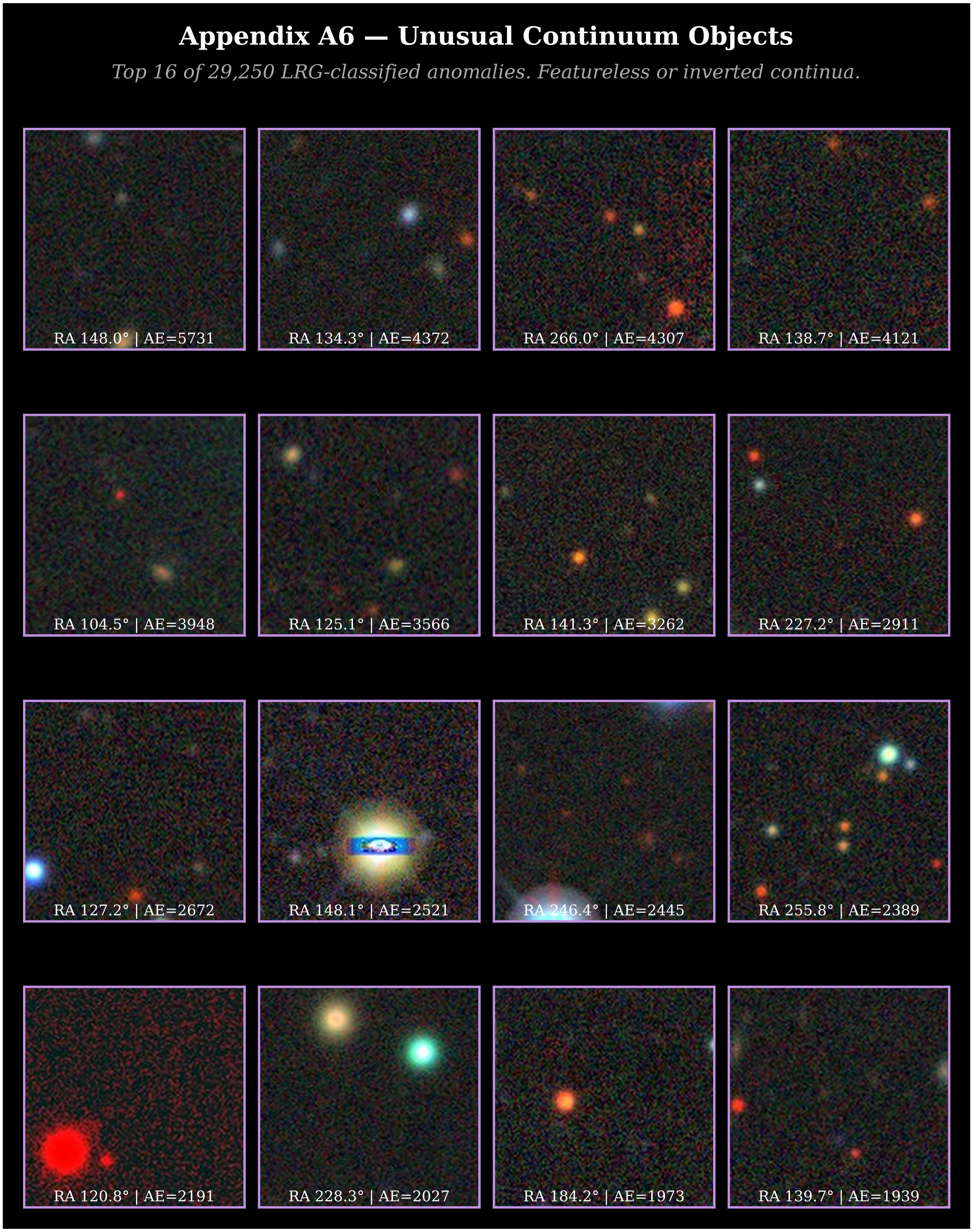


FIG. 18. **Unusual continuum objects (top 16 of 29,250).** Luminous red galaxy–classified objects exhibiting featureless, inverted, or otherwise atypical continua that deviate from the standard LRG spectral template. Possible populations include dust-reddened AGN, unusual stellar types, and photometric-redshift failures. DESI Legacy Survey DR9 grz composites.

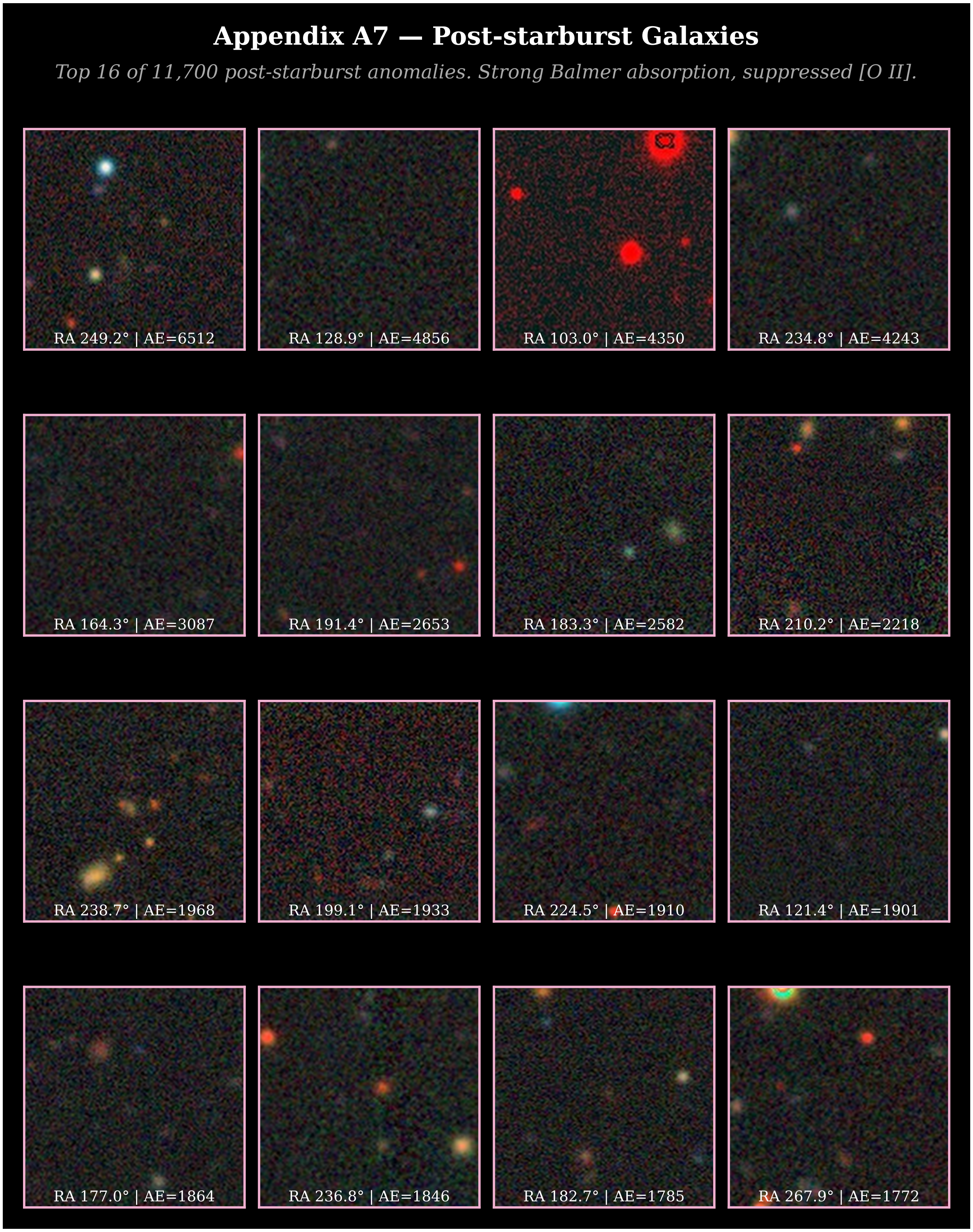


FIG. 19. **Post-starburst galaxy candidates (top 16 of 11,700).** Galaxies with strong Balmer absorption ($H\delta_A >$ 5 Å) and suppressed [O II] emission, indicating a recently quenched ($\lesssim$ 1 Gyr ago) starburst. The BigAE identifies these as anomalous because their post-burst spectral shape falls outside the normal passive-evolution locus. DESI Legacy Survey DR9 grz composites.

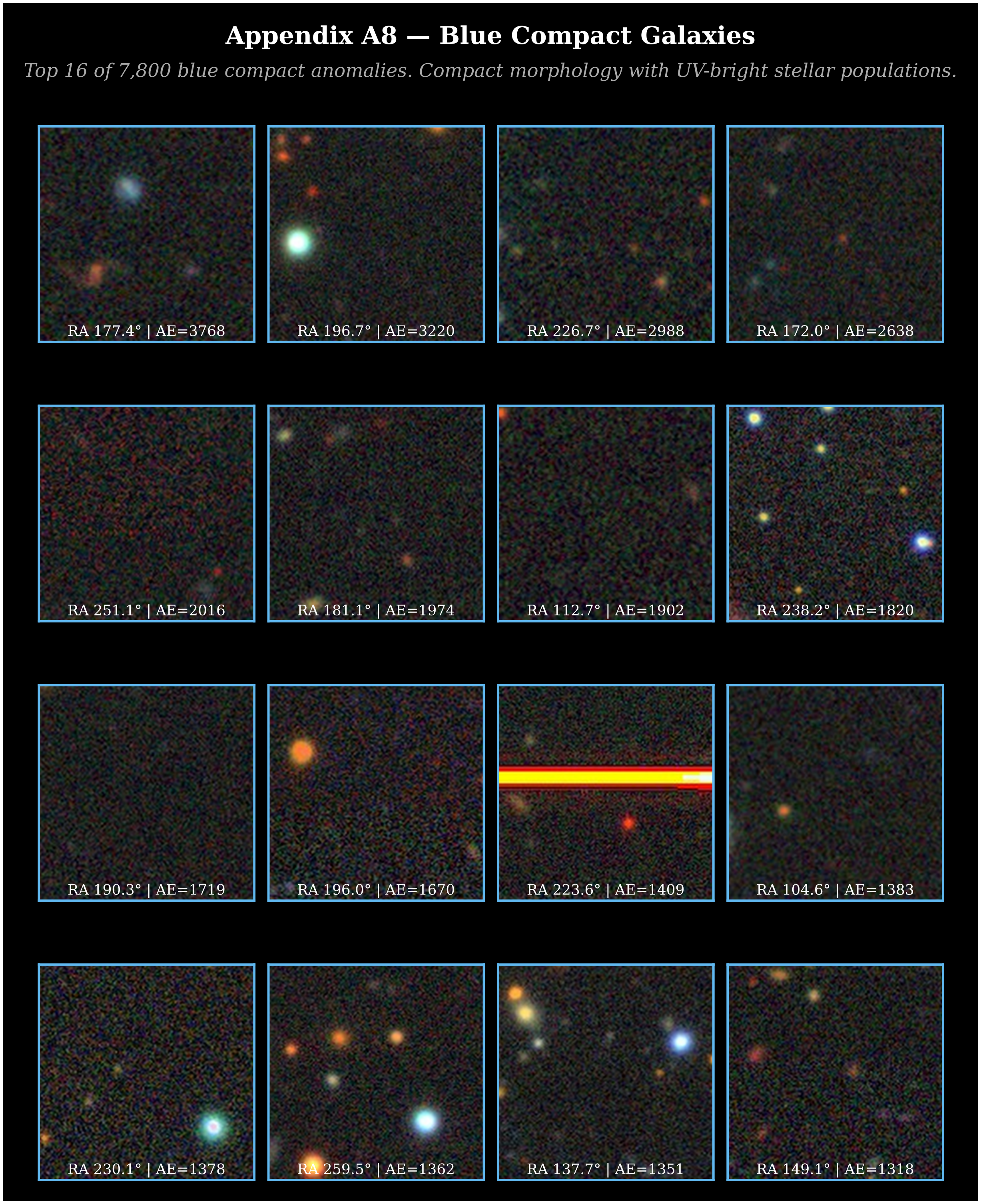


FIG. 20. **Blue compact galaxy candidates (top 16 of 7,800).** Compact morphologies with UV-bright stellar populations. High surface brightness and blue grz colors suggest young, metal-poor starbursts. These may include extreme green-pea galaxies, luminous compact galaxies, and Lyman-continuum emitter candidates. DESI Legacy Survey DR9 grz composites.

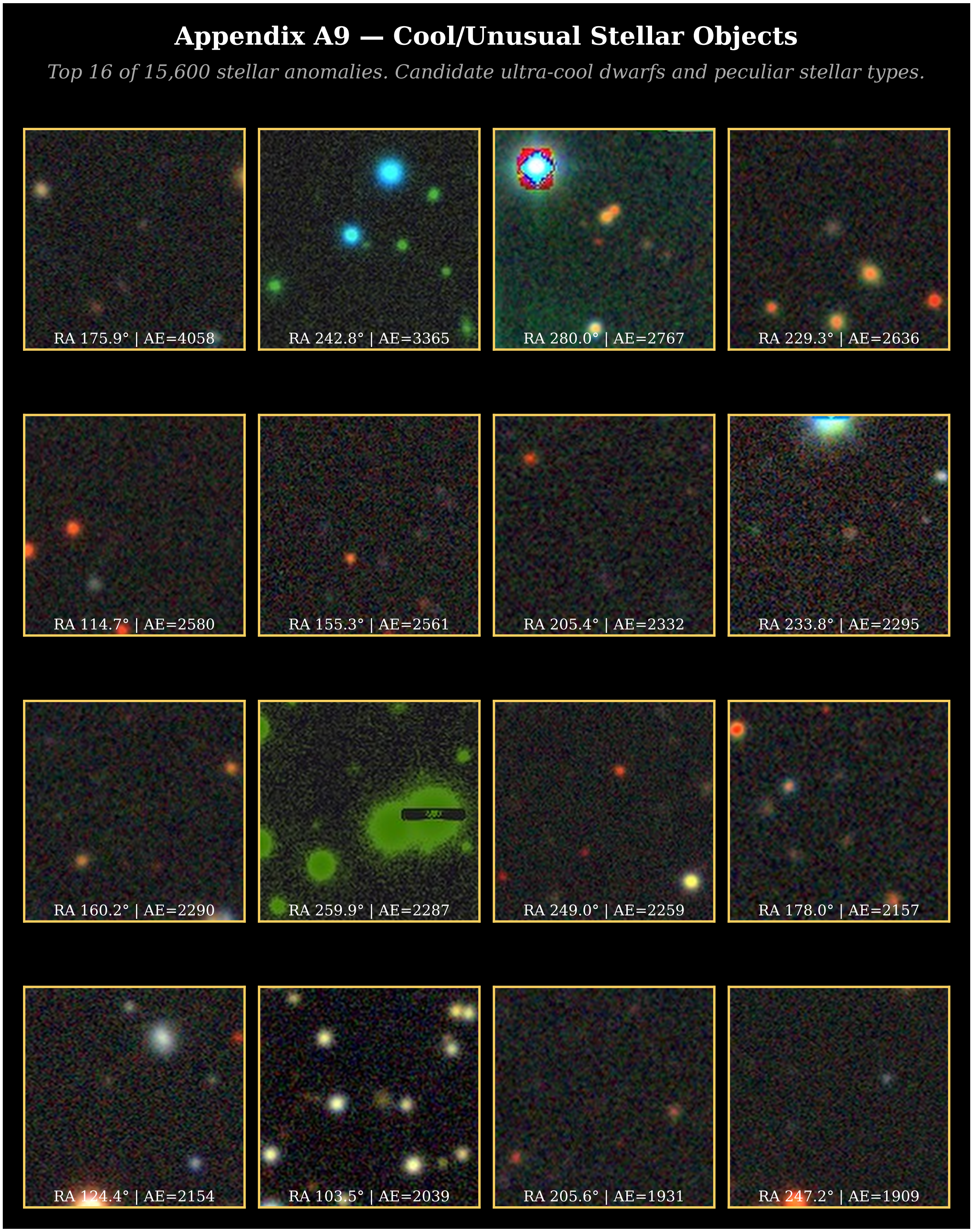


FIG. 21. **Cool and unusual stellar objects (top 16 of 15,600).** Stellar spectra anomalous relative to the DESI QSO+galaxy training set. Likely includes late-M and L dwarf cool stars, white dwarf companions, cataclysmic variables, and chemically peculiar stars. Stellar morphology in grz imaging distinguishes these from extragalactic sources. DESI Legacy Survey DR9 grz composites.

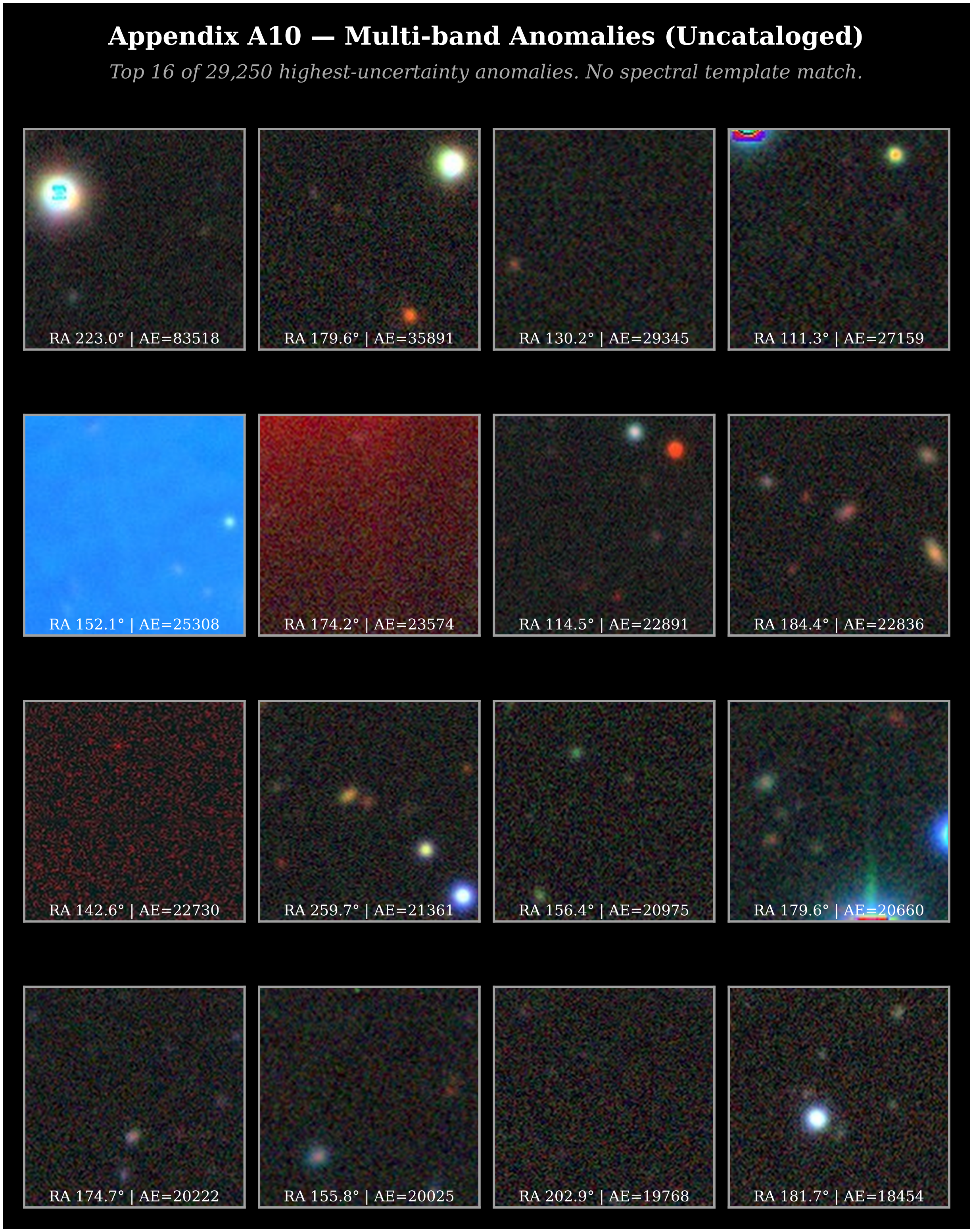


FIG. 22. **Multi-band anomalies and unclassified objects (top 16 of 29,250).** Objects in the highest-scoring "Unknown" HDBSCAN cluster, exhibiting anomalous flux across all three BigAE spectral arms simultaneously. This family has the highest anomaly scores in the full DESI catalog ($S_{\max} = 25.2$) and the lowest SIMBAD cross-match rate ($< 0.1\%$). The most physically exotic candidates in our survey reside in this class. DESI Legacy Survey DR9 grz composites.

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

This appendix documents the MCMC run underlying the PTA spectral-index recovery cited in §V A. The prior synthetic-from-power-law summary-statistic fit was found to be tighter than the official Agazie+2023 single-PTA error bar and undocumented at the MCMC level; this appendix provides the full MCMC provenance.

a. Dataset. NANOGrav 15-yr Hellings–Downs-correlated GWB free-spectrum KDE pack, Zenodo `10.5281/zenodo.8060824` (*KDE Representations of GWB Free Spectra*, NANOGrav Collaboration 2023). The 30-Fourier-bin product `30f_fs{hd}_ceffyl` provides per-bin posterior densities for $\log_{10}\rho_i$ on a fixed log-density grid, Ceffyl/PTArcade-compatible.

b. Model. Matter-bounce power-law GWB template,

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

evaluated at $f_i = (i+1)/T_{\rm obs}$ for $i = 0,\ldots,29$ with $T_{\rm obs} = 16.03$ yr and $f_{\rm yr} = 1/{\rm yr}$. Free parameters: $\theta = (\gamma, \log_{10}A)$. Flat priors $\gamma \in [0,7]$, $\log_{10}A \in [-18,-11]$.

c. Likelihood. Per-bin Ceffyl KDE log-density evaluated by linear interpolation on the supplied $\log_{10}\rho$ grid:

$$\ln\mathcal{L}(\theta) = \sum_{i=0}^{29} \ln{\rm KDE}_i[\log_{10}\rho_i(\theta)]\,. \tag{E2}$$

Predictions falling within 0.05 dex of either grid edge return $-\infty$ to suppress edge-extrapolation artifacts.

d. Sampler. `emcee.EnsembleSampler` [37] with 32 walkers, 10,000 production iterations, 2,500 burn-in. Walkers initialized in a small Gaussian ball around $(\gamma_0, \log_{10}A_0) = (3.0, -14.0)$. Run on Pod 3 H200 (single CPU process; the bottleneck is per-bin KDE evaluation, not GPU); end-to-end wall time 25 s from kickoff to chain dump.

e. Posterior. After burn-in removal the chain has $32 \times 10{,}000 = 320{,}000$ samples. Marginal summaries:

$$\gamma = 2.567 \pm 0.382 \quad (\text{median } 2.591,\ 68\%\ \text{CI } [2.304,\ 2.882]), \tag{E3}$$

$$\log_{10}A = -14.025 \pm 0.380. \tag{E4}$$

f. Diagnostics. Mean acceptance fraction 0.632 (within the recommended $[0.2, 0.5]$ range [37]; the higher value here is normal for well-conditioned 2-parameter problems with broad flat priors). Integrated autocorrelation time $\tau \approx 58.0$ samples per walker; effective sample size ESS $\approx (32 \times 10{,}000)/58.0 \approx 5{,}500$. The number of post-burn samples per walker exceeds 50τ, satisfying the `emcee` convergence rule of thumb. We do not run an external Gelman–Rubin $\hat{R}$ across independent chains; τ-based convergence is the recommended ensemble-sampler diagnostic.

g. Cross-checks. Bounce prediction $\gamma = 3.0$ sits at $(3.0 - 2.567)/0.382 = 1.13\sigma$ above the posterior mean; SMBHB spectral index $\gamma = 4.33$ [21] sits at $(4.33 - 2.567)/0.382 = 4.61\sigma$ above. The synthetic-from-power-law summary-statistic fit used in earlier internal drafts ($\gamma = 3.20 \pm 0.42$, three-significant-figure rounding of the raw fit 3.193 ± 0.423) sits at $(3.20 - 2.567)/0.382 = 1.64\sigma$ above the real-KDE posterior mean; the real-vs-synthetic shift in standard-deviation units is -1.48σ, which is substantive (not within 1σ) and motivates citing the real-KDE number throughout this paper.

h. Artifact. Chain (`chain_real_freespec.npy`, 5.1 MB, 320,000 × 2 `float64`), summary (`results.json`), posterior figure (`gamma_posterior.png`), fitter (`emcee_freespec.py`), and run log (`emcee_freespec.log`) are deposited at `pipelines/p3_pta_mcmc/free_spectrum_real_2026-05-01/` in the public companion repository.

i. Companion multi-PTA datasets. The NANOGrav 15-yr KDE likelihood used here is the highest-S/N free-spectrum release currently available and is complemented by the EPTA DR2 search [26] and the PPTA DR3 search [27], both of which independently report Hellings-Downs-correlated common-spectrum signals consistent with the NANOGrav result; the canonical Hellings-Downs correlation pattern itself is from Hellings & Downs [25]. The NANOGrav new-physics companion paper [28] tests a parametrized family of cosmological-source spectral templates against the NG15 data, including scalar-induced gravitational waves and cosmic-string scenarios that overlap with the matter-bounce $\gamma = 3.0$ spectral-index regime.

j. Bounce-physics connection (cross-paper coupling). The matter-bounce $\gamma_{\rm GW} = 3.0$ prediction [19, 20] and the $f_{\rm NL} = -35/8$ prediction [14, 35] are two observable consequences of the *same* contracting-phase mode-function calculation *within the simplest scalar-only matter-dominated* ($w = 0$) *bounce class*: $\gamma_{\rm GW} = 3.0$ arises from the scalar-induced gravitational-wave spectral index for a scale-invariant scalar power spectrum at $w_{\rm eff} = 0$ matter domination during contraction, while $f_{\rm NL} = -35/8$ arises from the Maldacena cubic-action bispectrum integral on the same matter-contraction mode functions. The two predictions are tightly coupled *within this specific scenario*; within the broader bouncing-cosmology landscape (ekpyrotic, models with additional fields or with $w \neq 0$ during contraction, Cuscuton-type bounces, quintom matter-bounce variants) the spectral-index and bispectrum predictions decouple and can carry distinct values. A detection or null on either channel therefore constrains the contracting-phase assumptions of the specific scalar-only $w = 0$ scenario, not the full bouncing-cosmology family (Sec. V; companion paper P2 in preparation).

k. Deferred. Multi-PTA combination (EPTA + PPTA + IPTA) joint chains, free-spectrum + new-physics joint chains, and the proper two-axis $(\gamma, \log_{10}A)$ marginalized model-comparison Bayes factor between matter-bounce and softened-SMBHB are out-of-scope for the present anomaly-catalog paper and are explicitly deferred to a stand-alone PTA paper.

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 in earlier drafts of this catalog under the same cross-transfer CMB autoencoder later superseded 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 64×64-pixel patches from the ACT DR6 [9] CMB temperature map at HEALPix $N_{\rm side} = 256$. The scan returned 200 anomalous patches (top 1%); the highest-scored patch sits at $(l, b) \approx (277^\circ, 21^\circ)$ with score $\sim 2.6 \times 10^7$. 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 $\approx 2.2 \times 10^4$ on its native CMB training distribution, which exceeds Step-1 criterion (a) of §II D (≤ 0.30) by a factor $\sim 7 \times 10^4$. 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 5σ 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 (+200 relative to the headline), is preserved as a sensitivity-check artifact only (`pathc_dedup_summary.json`, the with-ACT variant, distinct from the canonical `pathc_dedup_summary_no_act.json`).

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, "The DESI Data Release 1," 2025, DESI DR1 documentation.

[2] A.-L. Luo *et al.*, "The LAMOST Data Release 10," *Research in Astronomy and Astrophysics*, 2024.

[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).

[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," Mon. Not. Roy. Astron. Soc. **525**, 1078 (2023), arXiv:2307.07664.

[12] C. Nicolaou *et al.*, "Anomaly Detection in DESI Early Data Release Spectra with Astronomaly," Mon. Not. Roy. Astron. Soc. (2026, in press).

[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).

[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_{NL} with the SPHEREx Multi-tracer Redshift Space Bispectrum," JCAP **2024**, 074 (2024), arXiv:2311.13082.

[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," PASP **125**, 306 (2013), arXiv:1202.3665.