Decision-Grade Calibration of Negative-Slope Discriminability in a Dark-Siren-Motivated Growth-Mapping Pipeline

Aiden B. Smith[1]

[1] *Independent Researcher*

This Letter reports a decision-grade calibration test of whether a dark-siren-motivated growth-mapping pipeline can distinguish relative versus absolute coupling constructions in synthetic truth-labeled injections. A fixed threshold ($\tau = 0.0646$) from prior calibration is held fixed and evaluated on a large power-map suite with 192 tasks (4 truth points, 48 holdouts each, 256 replicates per holdout). For the baseline decision run, the mean true-positive rate across relative-truth points is 0.4043, the mean false-positive rate on the absolute-truth point is 0.2983, and the separation is $\Delta_{\rm sep} = 0.1059$. Three additional full-core replicate seeds yield $\Delta_{\rm sep} = \{0.1000, 0.1067, 0.1089\}$, giving a four-run mean 0.1054 with run-to-run standard deviation 0.0038. These results show stable, nonzero out-of-sample discriminability under fixed scoring and threshold rules, but only at moderate strength. The test therefore supports a calibrated and reproducible directional signal within this model family, while remaining insufficient for model-independent physical confirmation on its own.

I. TEST DEFINITION

For each replicate, let

$$\Delta s \equiv s_{\rm rel} - s_{\rm abs}, \tag{1}$$

where $s_{\rm rel}$ and $s_{\rm abs}$ are paired scores under relative and absolute growth mappings. A replicate is classified as "relative" if $\Delta s \geq \tau$ and "absolute" otherwise, using fixed threshold

$$\tau = 0.06459924567842923. \tag{2}$$

II. DECISION-GRADE RUN

The decision configuration used:

- truth grid: relative $\nu = \{0.2, 0.5, 0.8\}$ and absolute $\nu = 0.0$;
- holdouts per truth point: 48;
- replicates per holdout: 256;
- workers: 256 CPU processes;
- run output: `outputs/growth_mapping_power_map_decision_20260210_044918UTC/`.

Point-level rates from the baseline decision run:

$$\mathrm{TPR}_{\nu=0.2} = 0.3432, \tag{3}$$
$$\mathrm{TPR}_{\nu=0.5} = 0.3805, \tag{4}$$
$$\mathrm{TPR}_{\nu=0.8} = 0.4892, \tag{5}$$
$$\mathrm{FPR}_{\nu=0.0} = 0.2983. \tag{6}$$

Global means are

$$\langle \mathrm{TPR}_{\rm rel} \rangle = 0.4043, \tag{7}$$
$$\langle \mathrm{FPR}_{\rm abs} \rangle = 0.2983, \tag{8}$$
$$\Delta_{\rm sep} = 0.1059. \tag{9}$$

Metric	Visual scale (0 to 0.60)	Value
TPR ($\nu = 0.2$)		0.3432
TPR ($\nu = 0.5$)		0.3805
TPR ($\nu = 0.8$)		0.4892
FPR ($\nu = 0.0$)		0.2983

Figure 1. Point-level classification rates from the baseline decision-grade run, showing monotonic increase in TPR across relative-truth points and lower absolute-truth FPR.

III. REPLICATE STABILITY

Three additional full-core seeds were run sequentially with all settings fixed:

- `outputs/growth_mapping_power_map_decision_replicates_s`
- `outputs/growth_mapping_power_map_decision_replicates_s`
- `outputs/growth_mapping_power_map_decision_replicates_s`

The separation values were

$$\Delta_{\rm sep} = \{0.1000,\ 0.1067,\ 0.1089\}, \tag{10}$$

and across all four decision-grade runs (baseline + 3 replicates),

$$\langle \Delta_{\rm sep} \rangle = 0.1054, \qquad \sigma_{\rm run} = 0.0038. \tag{11}$$

Run	Visual scale (0 to 0.12)	$\Delta_{\rm sep}$
Baseline		0.1059
Rep-1		0.1000
Rep-2		0.1067
Rep-3		0.1089
Mean		0.1054

Figure 2. Run-to-run stability of separation across baseline plus three replicate seeds. The spread is small relative to the mean positive separation.

IV. INTERPRETATION

This test establishes that the calibrated classifier produces a stable positive separation under independent seeds and fixed rules. The effect is reproducible but moderate, so this result is best interpreted as calibration-grade evidence of nonzero discriminability in the tested model family, not as a standalone physical proof.

DATA AND SOFTWARE AVAILABILITY

Software archive for this pipeline: doi:10.5281/zenodo.18582609.

External data/product DOIs used by the underlying stack include:

- GWTC-3 products: doi:10.1103/PhysRevX.13.041039.
- GLADE+: doi:10.1093/mnras/stac1443.
- Pantheon+ papers and dataset: doi:10.3847/1538-4357/ac8b7a, doi:10.3847/1538-4357/ac8e04, doi:10.5281/zenodo.16365279.
- Cosmic-chronometer compilations: doi:10.1088/1475-7516/2012/08/006, doi:10.1088/1475-7516/2016/05/014.
- BOSS DR12 BAO+FS: doi:10.1093/mnras/stx721.
- eBOSS DR16 LRG BAO+RSD: doi:10.1093/mnras/staa2455.
- DESI 2024 BAO analyses: doi:10.1088/1475-7516/2025/04/012, doi:10.1088/1475-7516/2025/02/021.
- Planck 2018 lensing: doi:10.1051/0004-6361/201833886.

AI-USE STATEMENT

AI systems were used extensively for software prototyping, experiment orchestration, diagnostics, and drafting/editing support in preparing this analysis package.

[1] R. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration), Phys. Rev. X **13**, 041039 (2023), doi:10.1103/PhysRevX.13.041039.

[2] G. Dálya *et al.*, Mon. Not. R. Astron. Soc. **514**, 1403 (2022), doi:10.1093/mnras/stac1443.