On the Identifiability of Negative-Binomial Counts-in-Cells Radio Dipole Fits Under Survey Systematics Templates

Aiden Smith*
Independent Researcher
(Dated: February 9, 2026)

Recent work has advocated a negative-binomial (NB) counts-in-cells dipole estimator for wide-area radio catalogs, in which per-cell overdispersion is absorbed by a single NB parameter inferred from the empirical mean and variance of the observed counts. Because radio dipole inferences are known to be sensitive to large-scale survey selection and calibration structure, a key question is whether a dipole-only NB model is identifiable on real survey footprints.

Using publicly staged NVSS, RACS-low, and LoTSS-DR2 source catalogs with geometrically defined footprints (including zero-count cells at HEALPix $N_{\mathrm{side}} = 32$), we reproduce the paper-style NB dipole-only fits and then perform controlled stress tests by adding nuisance templates. We implement both a linear modulation model (close to the baseline form) and an exponential intensity modulation that preserves positivity by construction and reduces to the linear model at first order. We further include physically motivated template maps available from public products or catalog columns: the Haslam 408 MHz map, NVSS flux uncertainty proxy $\langle e_{S_{1.4}} \rangle$, and LoTSS per-source RMS/mask proxies aggregated into per-cell means, as well as a per-cell fractional MOC coverage proxy for the LoTSS footprint.

Two results are emphasized. First, injection/recovery tests show that modest structured selection templates can cause large axis mis-recovery under the dipole-only model. Second, on the real joint three-survey fit, adding nuisance freedom yields substantially higher likelihood solutions and large direction shifts relative to the dipole-only solution; with the physical-template set, the improvement is large enough to be favored even under BIC. These findings indicate that dipole-only NB fits are not robustly identifiable without explicit, physically grounded survey systematics modeling and validated template choices.

I. INTRODUCTION

The kinematic dipole from our motion with respect to the cosmic rest frame is expected to imprint a dipolar modulation in sufficiently deep, wide-area number-count surveys. In practice, radio dipole measurements can be biased by large-scale survey structure (depth, calibration, scan/striping, masking, and heterogeneous selection), and an analysis must establish that the inferred dipole is stable under plausible modeling choices.

This note audits the identifiability of a negative-binomial counts-in-cells dipole estimator advocated in Ref. [1], by performing (i) an end-to-end reproduction using geometrical footprints (including zero-count cells) and (ii) controlled stress tests with nuisance templates, including several physically motivated proxy maps available from public products.

II. DATA AND FOOTPRINTS

We use three publicly staged catalogs: NVSS at 1.4 GHz [2], RACS-low DR1 at 888 MHz [3], and LoTSS-DR2 at 144 MHz [4]. Selections follow the staged analysis defaults: a Galactic latitude cut $|b| \geq 10°$, survey declination bounds (NVSS $\delta \geq -39°$, RACS-low $-78° \leq \delta \leq 28°$), and for LoTSS a public MOC footprint (best-effort proxy for the "inner masked" region cited in the paper).

Counts are aggregated to HEALPix $N_{\mathrm{side}} = 32$ cells on a nested grid. Crucially, *all* cells inside each survey footprint are included, including zero-count cells, so the map support is footprint-defined rather than data-defined.

III. NB DIPOLE MODEL AND TEMPLATE EXTENSIONS

A. Negative binomial with fixed p

For each survey map, per-cell counts y_i are modeled as NB with success probability p and cell-dependent "shape" r_i,

$$y_i \sim \mathrm{NB}(r_i, p), \tag{1}$$

with p fixed from the empirical mean and variance of the observed counts in the footprint,

$$p \simeq 1 - \frac{\mu}{\mathrm{Var}} \quad (\text{clipped to } (0,1)). \tag{2}$$

This reproduces the core dispersion-handling idea in Ref. [1].

B. Dipole-only modulation

The baseline dipole-only model takes

* aidenblakesmithtravel@gmail.com

$$r_i = r_0 \left[1 + d\,(\hat{\boldsymbol{n}}_i \cdot \hat{\boldsymbol{d}}) \right], \tag{3}$$

where $\hat{\boldsymbol{n}}_i$ is the unit direction of the cell center, $\hat{\boldsymbol{d}}$ is the dipole direction, and d is the dipole amplitude parameter. Joint fits share $(d, \hat{\boldsymbol{d}})$ across surveys while allowing survey-specific $r_{0,s}$.

C. Template stress tests

To probe identifiability, we add nuisance templates $t_k(\hat{\boldsymbol{n}})$ and coefficients α_k in two ways.

a. Linear template extension. We extend Eq. (3) as

$$r_i = r_0 \left[1 + d\,(\hat{\boldsymbol{n}}_i \cdot \hat{\boldsymbol{d}}) + \sum_k \alpha_k\, t_{k,i} \right], \tag{4}$$

noting that this model requires $r_i > 0$ for all cells and therefore can exhibit boundary behavior.

b. Exponential (positivity-preserving) stress model. We also consider

$$r_i = r_0\, \exp\left[d\,(\hat{\boldsymbol{n}}_i \cdot \hat{\boldsymbol{d}}) + \sum_k \alpha_k\, t_{k,i} \right], \tag{5}$$

which guarantees $r_i > 0$ and reduces to the linear model at first order in the bracketed argument. This is used as a stress test for coherent large-scale selection structure not captured by the dipole-only model.

c. Template basis. The minimal geometric basis includes z-scored templates of $(\delta, \delta^2, \sin\alpha, \cos\alpha)$ (declination and right ascension). We additionally include a set of physically motivated templates available from public products or catalog columns: the Haslam 408 MHz map [5] sampled at cell centers, a NVSS flux-uncertainty proxy from the catalog column $\langle e_{S_{1.4}} \rangle$ aggregated per cell, and LoTSS per-source proxies $\langle$Isl_rms$\rangle$ and $\langle$Masked_Fraction$\rangle$ aggregated per cell, as well as a per-cell fractional MOC-coverage proxy for the LoTSS footprint. All templates are z-scored on the survey footprint.

D. Optimization and complexity bookkeeping

We maximize the NB log likelihood with bounded L-BFGS-B, using multistart initialization to reduce local-minimum pathologies. To provide a coarse model-comparison diagnostic, we report AIC and BIC,

$$\mathrm{AIC} = 2k - 2\log\mathcal{L}, \qquad \mathrm{BIC} = k\log n - 2\log\mathcal{L}, \tag{6}$$

where k is the number of fitted parameters and n is the total number of included cells for the given fit (single survey or joint).

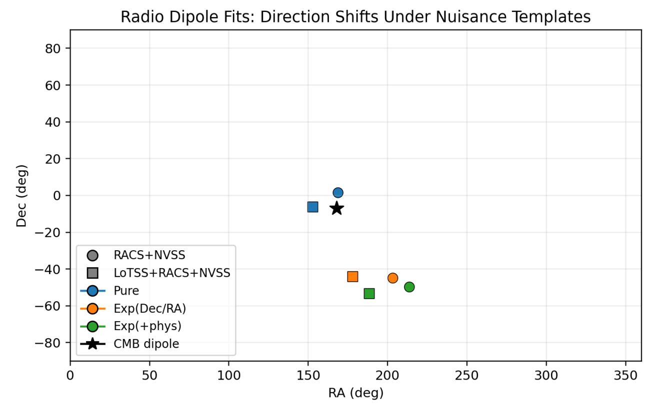

FIG. 1. Best-fit dipole directions for joint two-survey and three-survey fits under three model families: pure dipole-only, exponential modulation with geometric Dec/RA templates, and exponential modulation with an expanded physical-template set. The CMB dipole direction is shown for reference.

IV. RESULTS

A. Direction sensitivity under nuisance templates

Figure 1 compares best-fit directions for joint two-survey (RACS+NVSS) and three-survey (LoTSS+RACS+NVSS) fits under the dipole-only model and two stress-test models. The baseline three-survey pure fit yields $d \simeq 0.0205$ with $(\alpha, \delta) \simeq (152.9^\circ, -6.2^\circ)$, about 15° from the CMB dipole direction. In contrast, nuisance-template models produce substantially different directions and (under the chosen conservative bound) push the recovered amplitude to the upper bound $d_{\max} = 0.05$.

B. Information criteria: dipole-only vs stress models

Figure 2 summarizes information-criteria differences relative to the pure model. For the joint three-survey fit, the physical-template exponential model achieves a large likelihood gain and is favored even under BIC in this configuration (despite a larger parameter count), indicating that the map contains coherent structure not captured by the dipole-only model. This directly challenges the identifiability of a dipole-only interpretation.

C. Which templates drive the improved fit?

Figure 3 shows the per-survey nuisance-template coefficients for the joint three-survey exponential physical-template fit. Large coefficients on LoTSS RMS/mask proxies and the NVSS flux-uncertainty proxy are consis-

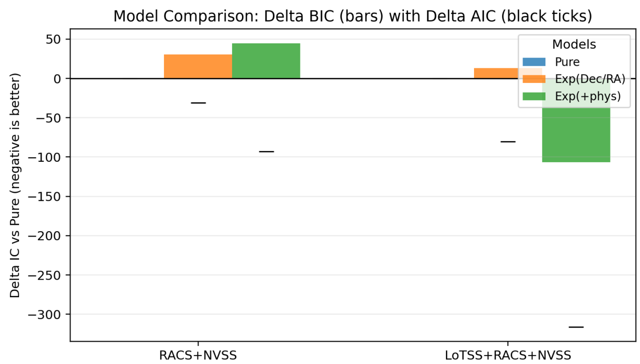

FIG. 2. Delta information criteria relative to the pure dipole-only model for joint fits. Bars show ΔBIC and black ticks show ΔAIC. Negative values indicate preference for the extended model.

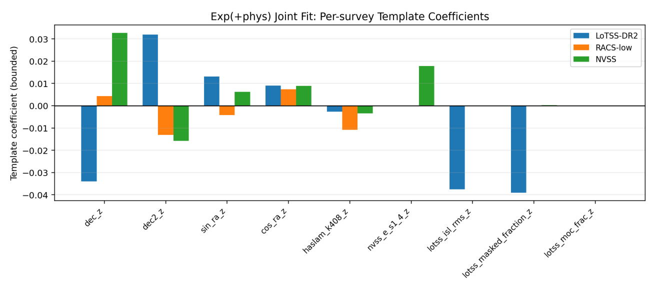

FIG. 3. Per-survey nuisance-template coefficients for the joint three-survey exponential physical-template model.

tent with survey-selection structure contributing to the apparent large-scale anisotropy.

D. Injection/recovery: axis mis-recovery under a dipole-only fit

Finally, Fig. 4 summarizes an injection/recovery diagnostic in which modest structured selection templates are injected (together with a CMB-direction dipole), and the resulting maps are fit with the pure model versus a template-augmented model. The pure dipole-only fit can exhibit large axis error and biased recovered amplitude under this controlled setup, reinforcing the point that dipole-only fits are not robust in the presence of structured survey selection.

V. CONCLUSIONS

The NB counts-in-cells dipole estimator can accommodate overdispersion, but this does not by itself ensure that a dipole-only model is identifiable on real radio survey footprints. Across joint two- and three-survey fits, adding simple nuisance templates (including physically motivated proxies tied to depth/masking and foreground structure) yields large likelihood gains and significant direction shifts, and injection/recovery tests show that dipole-only fits can mis-recover the dipole axis under modest structured systematics.

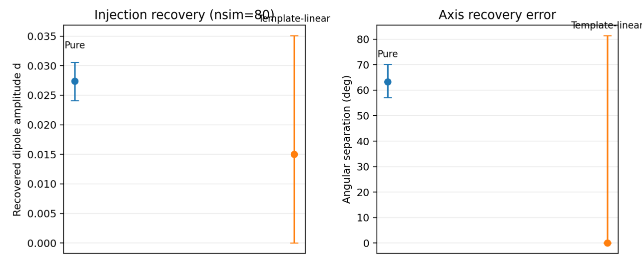

FIG. 4. Injection/recovery summary (one configuration). Points show p50 with p16–p84 error bars for the recovered dipole amplitude d and the angular separation between recovered and injected dipole directions.

These results motivate treating dipole-only NB inferences as incomplete unless the analysis includes explicit, physically grounded systematics templates, robustness to template choices, and end-to-end injection/recovery validation on the relevant footprints.

DATA AND CODE AVAILABILITY

The software and reproducibility materials for this radio NB dipole audit are archived on Zenodo at doi:10.5281/zenodo.18530376. This study is extracted from the broader Quasars-Systematics archive at doi:10.5281/zenodo.18476711. Related external products used in this analysis include NVSS doi:10.1086/300337, RACS DR1 doi:10.1017/pasa.2020.41, LoTSS-DR2 doi:10.1051/0004-6361/202142484 (with released LoTSS-DR2 data products at doi:10.25606/SURF.LoTSS-DR2), and the Remazeilles 408 MHz map doi:10.1093/mnras/stv1274.

ACKNOWLEDGMENTS

The author used AI assistance during this project for brainstorming, drafting/editing text, and software development.

[1] L. Böhme, D. J. Schwarz, P. Tiwari, M. Pashapour-Ahmadabadi, B. Bahr-Kalus, M. Bilicki, C. L. Hale, C. S. Heneka, and T. M. Siewert, "Overdispersed radio source counts and excess radio dipole detection," *arXiv:2509.16732* (2025).

[2] J. J. Condon *et al.*, "The NRAO VLA Sky Survey," *Astron. J.* **115**, 1693 (1998), doi:10.1086/300337.

[3] D. McConnell *et al.*, "The Rapid ASKAP Continuum Survey I: Design and first results," *Publ. Astron. Soc. Aust.* **37**, e048 (2020), doi:10.1017/pasa.2020.41.

[4] T. W. Shimwell *et al.*, "The LOFAR Two-metre Sky Survey. V. Second data release," *Astron. Astrophys.* **659**, A1 (2022), doi:10.1051/0004-6361/202142484.

[5] M. Remazeilles, C. Dickinson, A. J. Banday, M. A. Bigot-Sazy, and T. Ghosh, "An improved source-subtracted and destriped 408 MHz all-sky map," *Mon. Not. R. Astron. Soc.* **451**, 4311 (2015), doi:10.1093/mnras/stv1274.