

Skeptical review: Geographic Consistency of Temperature and Lensing Power in ACT DR6.02 Daytime Data: Day-Side versus Day-Night Splits at 90 and 150 GHz

Summary

This manuscript presents a diagnostic study of whether the standard ACT DR6.02 daytime geographic split into Day-Side (DS) and Day-Night (DN) regions yields statistically interchangeable subsets for temperature (TT) power spectra and temperature-only CMB lensing ($\kappa\kappa$) reconstructions. Using PA6 maps at 90/150 GHz and a common pseudo- \mathcal{C}_ℓ pipeline with four-way temporal jackknives, the authors compute DS and DN TT autospectra, DS \times DN TT cross-spectra (as a null test), and temperature-QE $\kappa\kappa$ autospectra and cross-spectra, focusing on 10 multipole bins ($\ell \approx 557\text{--}3625$) for the main 150 GHz DS/DN comparisons (Secs. 3–4, Appendix A). The headline empirical results are striking: DS TT power is strongly suppressed relative to DN (mean ratio ~ 0.3) and DS \times DN TT cross-power is near-null; for lensing, DS $\kappa\kappa$ is systematically lower than DN (inverse-variance-weighted ratio $\langle R^\kappa \rangle \approx 0.85$) with an extremely large heuristic χ^2 reported under the naive null (Secs. 4.1–4.2, Table 1). Additional checks (AA 90/150 GHz QE coherence, DS/DN vs AA cross-spectra, and a declination-based split on AA maps) suggest that the dramatic DS/DN behavior is not a generic artifact of arbitrary north–south masking (Secs. 4.3–4.5), and a gradient- κ diagnostic indicates internally consistent small-scale structure in DS though it is not treated as a cosmological detection (Sec. 4.6). The paper is timely and potentially useful as a quality-control case study for daytime CMB analyses, but its impact and interpretability would be substantially improved by (i) a precise operational definition and quantitative characterization of DS/DN footprints and weights, (ii) a clearer accounting of calibration/beam/transfer-function/noise contributions to the extreme TT ratio, and (iii) more explicit description of what is (and is not) included in the $\kappa\kappa$ bandpowers and their uncertainties (Secs. 2–6).

Strengths

- Targets a practically important “bigger picture” point: survey splits (here DS/DN daytime geography) are part of the instrument/observing model and need explicit validation rather than being assumed interchangeable (Secs. 1.2–1.3, 6).
- Reports quantitatively strong empirical anomalies/diagnostics (DS/DN TT ratio ~ 0.3 ; $\kappa\kappa$ ratio ~ 0.85 ; near-null DS \times DN cross-spectra), making the consistency problem hard to ignore and worth understanding (Secs. 4.1–4.2, Appendix A).
- Uses multiple complementary internal checks—TT cross nulls, κ cross suppression, AA 90/150 QE coherence, DS/DN vs AA cross-spectra, and a declination split—to help localize the problem to the DS/DN construction rather than to generic masking or frequency-dependent failures (Secs. 4.3–4.5).

- Appropriately frames the work as diagnostic rather than a calibrated cosmological claim, and explicitly notes limitations (neglected bin–bin covariance, missing DS/DN-matched simulations, incomplete propagation of beam/calibration uncertainties) (Secs. 5, 6.1–6.2).
- Good reproducibility intent via references to archived analysis products (e.g., .npz outputs) and a consistent narrative from setup to results and checks, which could become a reusable QC template once key operational details are filled in (Secs. 2–4, Appendix A).

Major issues

1. **The operational definition of DS and DN (and the exact masks/weights used in spectra and QE) is not specified with enough precision to interpret the near-null DS×DN cross-spectra or the extreme TT and $\kappa\kappa$ ratios.** The manuscript qualitatively suggests DS/DN are “largely disjoint,” but does not quantify f_{sky} , overlap fraction, apodization, or hit-count/inverse-variance weight distributions that likely drive both auto suppression and cross nulls (Sec. 2; interpretation in Secs. 4.1–4.2, 4.4; discussion in Sec. 5).

Recommendation: In Sec. 2, add an explicit, actionable DS/DN definition: (i) how DS and DN are constructed in the DR6.02 daytime archive (time/az/scan criteria vs pre-computed region labels); (ii) the exact masks used for TT and for QE (including apodization type/scale); (iii) $f_{\text{sky}}(\text{DS})$, $f_{\text{sky}}(\text{DN})$, and $f_{\text{sky}}(\text{overlap})$ (or an overlap fraction); and (iv) basic weight/depth statistics (e.g., median/percentiles of inverse-variance weights or an effective μK -arcmin depth proxy) for each region at 150 GHz. Include a footprint/weight map figure or a compact table. Then, in Secs. 4.1–4.2, state explicitly whether DS and DN are treated as disjoint for the analysis and what overlap is expected to contribute to DS×DN cross-power for a common CMB sky with uncorrelated noise.

2. **Key elements of the map-making and pseudo- C_ℓ pipeline are under-specified, making it difficult to assess whether DS/DN differences arise from calibration, beam/transfer-function differences, filtering, weighting, or noise modeling.** The text indicates “same pipeline choices” but does not document beam/transfer deconvolution, mode-coupling correction, mapmaking filters/cuts, or whether calibration/transfer functions are shared or independently derived per region (Secs. 2, 3.1–3.2).

Recommendation: Expand Sec. 2 and Sec. 3.1 with a concise but complete “analysis configuration” description: (i) beam treatment (per-split vs common beam, deconvolution choice, beam uncertainty); (ii) transfer function/filtering description (time-domain filtering scales, map-domain filtering, any ℓ -dependent response corrections); (iii) pseudo- C_ℓ details (mask apodization, mode-coupling matrix treatment, binning/weighting, any noise-bias subtraction for TT autos); and (iv) calibration proce-

ture (absolute and relative) and whether DS/DN inherit a common calibration. Where these are inherited from published ACT DR6 pipelines (e.g., Naess et al. 2020; Qu et al. 2024; Madhavacheril et al. 2024), cite the exact configuration/section and list any deviations relevant to DS/DN.

3. **The magnitude and ℓ -dependence of the TT suppression ($R^{TT} \sim 0.3$) is so extreme that the paper needs a more quantitative decomposition into plausible causes (relative gain, transfer-function mismatch, beam differences, noise bias/weighting effects).** Currently Sec. 5 lists possible contributors but does not demonstrate whether any are numerically capable of producing a factor ~ 3 change in TT power (Secs. 4.1, 5).

Recommendation: Strengthen Sec. 5 with quantitative “sanity checks” tied to the presented spectra: (i) test a pure multiplicative calibration model ($C_\ell \propto g^2$): $R^{TT} \approx 0.3$ implies $g \approx 0.55$ —state whether such a relative gain is plausible under DR6 daytime calibration; (ii) examine ℓ -dependence of R_b^{TT} to distinguish calibration-like (flat) vs transfer-function-like (tilted) behavior, and relate this to any known daytime filtering/atmospheric differences; (iii) estimate DS and DN noise levels using high- ℓ TT, difference maps, or jackknife products, and show how noise/weighting would bias auto bandpowers; and (iv) explicitly clarify what “beam-corrected TT” means here (Figure captions and Sec. 3.1), including whether DS and DN beams differ. Even back-of-the-envelope bounds would materially improve interpretability.

4. **The QE lensing reconstruction and the exact content of the plotted/archived $\kappa\kappa$ spectra are not described with enough specificity to interpret $R^\kappa \approx 0.85$.** It is unclear whether the $\kappa\kappa$ bandpowers used for ratios are raw, $N^{(0)}$ -subtracted, $N^{(1)}$ -corrected, Monte-Carlo-corrected, and/or normalized identically between DS and DN; given the stated $N^{(0)}$ dominance, region-dependent noise differences could explain much of the effect, but the analysis does not quantify this (Secs. 3.2–3.4, 4.2, 5).

Recommendation: In Sec. 3.2 (or a dedicated subsection), specify: (i) QE estimator type (temperature-only TT), the L range for κ bandpowers, and ℓ cuts for gradient/small-scale legs; (ii) filtering applied before QE (including any inpainting, mean-field subtraction, or mask treatment); (iii) the normalization method (analytic vs simulation-based) and whether it is common to DS/DN; and (iv) precisely which bias terms are removed ($N^{(0)}$, $N^{(1)}$, any MC corrections) in the $\kappa\kappa$ spectra used in Sec. 4.2 and Appendix A. Then, add a short quantitative diagnostic: compare estimated $N_L^{(0)}$ (or an effective QE noise level) between DS and DN and show whether the observed R^κ can be explained primarily by reconstruction noise differences rather than changes in the true lensing signal.

5. **The statistical framing over-emphasizes a χ^2 -like quantity (Eq. 5; $\chi^2 \approx 5168$ for 10 bins) while explicitly neglecting bin–bin covariance, and the jackknife error model (4 temporal splits, delete-one) is not sufficiently ex-**

plained to justify extremely small quoted uncertainties on weighted-mean ratios. As written, readers may misinterpret the χ^2 and sub-permil errors as formal significances (Secs. 3.3, 4.2, Table 1, Appendix B; also affects TT in Sec. 4.1).

Recommendation: Revise Sec. 3.3 and Sec. 4.2 to (i) clearly define what is jackknifed (the bandpowers, the ratios, or a combined estimator) and how 4 temporal splits enter the cross/auto constructions; (ii) provide a reality check uncertainty summary that does not rely solely on potentially misestimated per-bin σ (e.g., report per-bin pulls, unweighted mean \pm RMS across bins, and/or a bin-bootstrap as a descriptive measure); and (iii) either estimate an approximate covariance for R_b^κ (from ACT simulations, analytic mode-counting, or even a nearest-neighbor model) and recompute a generalized χ^2 with an interpretable p-value, or explicitly demote Eq. (5) to a “severity index” and stop presenting it in a way that resembles a formal χ^2 test. Update Eq. (5) typesetting accordingly (see very minor issues).

6. **Cross-spectrum ‘null tests’ (DS \times DN TT and κ cross) are interpreted as consistency checks, but if DS and DN have negligible overlap, near-zero cross-power is largely guaranteed and does not strongly constrain shared systematics; conversely, if overlap exists, a near-null cross could indicate deeper calibration/transfer inconsistencies.** The manuscript does not connect the observed cross suppression to an explicit overlap/geometry expectation (Secs. 4.1–4.2, 4.4).

Recommendation: After quantifying DS/DN overlap (Major Issue 1), add to Sec. 4.1–4.2 a simple expectation for cross-power: e.g., predict the DS \times DN TT cross amplitude (or cross/auto ratio) for a common CMB signal under the measured masks/weights, assuming uncorrelated noise. Then interpret the observed DS \times DN cross bandpowers relative to that expectation. If overlap is near zero, state explicitly that the cross is a weak systematic diagnostic and mainly confirms disjointness; if overlap is non-negligible, discuss what classes of systematics could drive an anomalously small cross.

7. **Presentation and data-product interpretability: several key figures and captions do not clearly specify units, normalization conventions (C_ℓ vs D_ℓ ; beam/transfer deconvolved or not), and uncertainty visualization; additionally, the mapping between plotted curves and specific archived arrays/files is not always explicit.** This is particularly problematic where order-of-magnitude statements are made (e.g., κ cross vs auto levels) and where the paper aims to be a reusable diagnostic using archived .npz products (Figures 1, 3, 4, 5, 7; Secs. 4.1–4.4; Appendix A).

Recommendation: For each relevant figure (esp. Figures 1, 3, 4, 5, 7), add: (i) explicit y-axis units and whether spectra are C_ℓ , D_ℓ , or an internal pipeline normalization; (ii) whether beam/transfer corrections and any noise/bias subtractions are applied; (iii) visible jackknife/simulation error bars or shaded bands on ratio and cross panels, with reference lines (0 or 1); and (iv) captions that define DS/DN/AA and point to the

exact .npz filename(s) and array keys corresponding to each plotted series. If the κ plots use a nonstandard normalization, state it prominently in the caption and/or add a companion panel in conventional units if feasible.

Minor issues

1. Multipole binning is not fully documented across analyses: the main DS/DN results use 10 bins ($\ell \approx 557\text{--}3625$), but other sections use 35-bin (ds_dn_null_test.npz) and 20-bin (declination-split) configurations without listing bin edges/centers (Secs. 3.2, 4.4–4.5, Appendix A).

Recommendation: Add a short table (Sec. 3 or Appendix A) listing bin edges/centers for the main 10-bin DS/DN analysis, and clearly describe how the 20- and 35-bin schemes differ and where exact edges can be retrieved (e.g., from the archived products).

2. Equation (8) defines a ‘binwise correlation’ using a normalization that is not dimensionless (cross divided by the product of autos) and includes unclear symbols (e.g., stray subscripts/superscripts). This conflicts with the term “correlation coefficient” and affects interpretation in the declination-split test (Sec. 4.5; Eq. 8; Table 1).

Recommendation: Fix Eq. (8) to be dimensionless, e.g. $r_{\ell_b} = \bar{C}_b^{\text{cross}} / \sqrt{\bar{C}_b^A \bar{C}_b^B}$, and define all symbols. Ensure the equation is present, correctly numbered, and consistently referenced in Sec. 4.5 and Sec. 7.

3. Nomenclature collision: Sec. 4.5 reuses DS/DN labels for a declination-based split, which is easy to confuse with the DS/DN daytime geographic split used throughout the paper (Sec. 4.5; figures/tables referencing that test).

Recommendation: Rename the declination split to DecN/DecS (or North/South) everywhere (text, equations, figure labels, Table 1) and reserve DS/DN exclusively for the daytime geographic split.

4. The inter-frequency QE coherence test is potentially useful but its scope is easy to overread: it is performed on AA coadd maps, so it constrains global 90/150 consistency on the combined footprint, not necessarily DS- or DN-specific issues (Secs. 3.4, 4.3, 6.2). Implementation details (L range, binning, bin selection) are also not specified.

Recommendation: In Sec. 3.4 and Sec. 4.3, specify the multipole (L) range and binning used to compute r (linear and log-space), and explicitly state what this test can/cannot rule out for DS vs DN. If feasible, also report a lower-S/N DS-only and DN-only 90/150 coherence as an additional regional check; otherwise, note it as future work.

5. The CAR-based processing used in the inter-frequency test is only briefly distinguished from the main HEALPix/pseudo- C_ℓ workflow, leaving ambiguity about comparability (Sec. 4.3 vs Secs. 2–3).

Recommendation: Add 2–3 sentences clarifying CAR vs HEALPix differences (masks, transfer functions, calibration, bandpower estimation) and why they do not qualitatively affect the conclusion about 90/150 coherence on AA.

6. The gradient- κ autospectrum diagnostic is mentioned as internally consistent but its operational meaning and what failure modes it would catch are not clearly stated; large $|\hat{C}_b|/\sigma$ values are attributed to normalization without a succinct explanation (Sec. 4.6).

Recommendation: Expand Sec. 4.6 briefly to define the diagnostic, state what anomalies it is meant to reveal (e.g., pathological ℓ dependence, sign issues), and explicitly note that the quoted $|\hat{C}_b|/\sigma$ values are not cosmological significances.

7. Notation for lensing-related quantities is inconsistent (κ vs other symbols; R_b^e vs R^κ ; multiple notations for observed/auto spectra), which makes it harder to track what is being ratioed and whether it is bias-corrected (Secs. 3.2, 4.2, 4.6, Appendix A).

Recommendation: Standardize on κ throughout, define $C_b^{\kappa\kappa,A}$ and R_b^κ once in Sec. 3.2, and keep the same notation in Table 1 and Appendix A. Add a short notation glossary if needed.

Very minor issues

1. Typographical/formatting inconsistencies in filenames and spacing (e.g., stray spaces in “. npz” filenames), section-heading capitalization, and spacing around symbols/units (Secs. 2, 4.3–4.6, 5).

Recommendation: Standardize filenames (remove stray spaces), harmonize heading capitalization, and clean up spacing around inline math and units (e.g., consistently “150 GHz”).

2. Citation/reference formatting is inconsistent (e.g., “others” vs “et al.”; collaboration naming), and there appears to be at least one stray line after references (Sec. 1.2; References).

Recommendation: Normalize references to the target journal style, use consistent author/collaboration conventions, and remove any stray lines or misplaced headers after the bibliography.

3. Equation numbering/cross-references contain minor inconsistencies (Eq. 7 introduced without clear pointer; Eq. 8 referenced where the displayed equation may be missing in some versions) (Secs. 4.2, 4.5, 7).

Recommendation: Audit equation numbering and ensure each numbered equation is displayed, introduced, and referenced consistently where used.

4. Eq. (5) and Eq. (6) appear to have typesetting artifacts (extra tokens in Eq. 5; ambiguous sqrt rendering in Eq. 6), which can obscure intended definitions (Sec. 3.3).

Recommendation: Re-typeset Eq. (5) cleanly as $\chi^2 = \sum_{b=1}^{N_{\text{bin}}} (R_b^\kappa - 1)^2 / \sigma_{R_b^\kappa}^2$, and ensure Eq. (6) renders the denominator as the product of standard deviations with explicit square roots.

5. Some wording choices could be made more standard for a scientific paper (e.g., “severity index” not defined; an unusual remark about figures not being inspected by a vision-language model) (Secs. 3.3, 6.2).

Recommendation: Define “severity index” when first used (if retained) and remove/rephrase nonstandard remarks in Sec. 6.2 to conventional provenance statements about figure generation.

Key statements and references

- ✓ **The ACT DR6 data set has enabled a high signal-to-noise measurement of the CMB lensing power spectrum and a public lensing map product with cosmological interpretation, supported by bias-hardened estimators, foreground mitigation, transfer-function modeling, and simulations described in recent ACT collaboration analyses.**
- *Reference(s):* Qu et al., 2024, Madhavacheril et al., 2024, MacCrann et al., 2023
- *Justification:* Verification failed with gpt-5: Error code: 400 - {'error': {'message': 'Your input exceeds the context window of this model. Please adjust your input and try again.', 'type': 'invalid_request_error', 'param': 'input', 'code': 'context_length_exceeded'}}}
- ✓ **The observed lensing convergence power spectrum in this work is modeled as receiving contributions from the true cosmological signal and additive reconstruction bias, schematically as $C^{\{\kappa\kappa, \text{obs}\}} \approx C^{\{\kappa\kappa, \text{true}\}} + N^{\{(0)\}} \{ \ell \} + \text{cdots}$, where $N^{\{(0)\}}$ denotes the dominant normalization-sensitive noise bias at the map level, following the standard CMB lensing reconstruction formalism.**
- *Reference(s):* Lewis and Challinor, 2006
- *Justification:* Lewis and Challinor, 2006 explicitly develop the quadratic CMB-lensing reconstruction and show that the auto-spectrum of the reconstructed potential contains a leading Gaussian (disconnected) noise contribution plus higher-order biases. In Sec. 7.1, Eqs. (7.4)–(7.6) define the lowest-order reconstruction noise $N(L)$ from the disconnected four-point term, and they note that higher-order $O(C_l^\psi)$ terms bias the estimated lensing power spectrum and must be corrected to obtain an unbiased result (citing Cooray & Kesden 2002; Kesden et al. 2003). While written for the potential ψ , this directly maps to the convergence κ . Thus the observed reconstruction power is modeled as true signal plus an additive ‘ $N^{(0)}$ ’ noise bias (and further corrections), consistent with the standard formalism.

- ✓ **The official ACT DR6 lensing analysis and the ACT daytime lensing demonstration rely on bias-hardened quadratic estimators, detailed foreground-mitigation strategies, transfer-function modeling, simulations, and curated null tests tuned to the published footprint and noise model, and these pipelines are not reimplemented in the present DS/DN diagnostic study.**
- *Reference(s)*: MacCrann et al. 2023, Qu et al. 2024, Madhavacheril et al. 2024
- *Justification*: Verification failed with gpt-5: Error code: 400 - {'error': {'message': 'Your input exceeds the context window of this model. Please adjust your input and try again.', 'type': 'invalid_request_error', 'param': 'input', 'code': 'context_length_exceeded'}}

Mathematical consistency audit

This section audits **symbolic/analytic** mathematical consistency (algebra, derivations, dimensional/unit checks, definition consistency).

Maths relevance: light

The paper is primarily a data-split diagnostic note. The mathematics consists mostly of standard definitions (bandpower D_ℓ , region ratios, Pearson correlation), a simple chi-square severity index, and a standard delete-one jackknife variance formula. There are no long derivations; internal consistency mainly hinges on correct normalization/units and consistent symbol definitions. The only clear internal mathematical error is the normalization in the declination-split 'correlation coefficient' (Eq. 8), which as written is not dimensionless and uses an undefined symbol.

Checked items

1. ✓ **Bandpower definition D_ℓ** (Eq. (1), Sec. 3.1, p.2)
 - **Claim:** Defines bandpower $D_\ell^{XY} = [\ell(\ell + 1)/(2\pi)] C_\ell^{XY}$.
 - **Checks:** symbol consistency, dimensional/units sanity
 - **Verdict:** PASS; confidence: high; impact: minor
 - **Assumptions/inputs:** C_ℓ^{XY} is the angular power spectrum of fields X, Y
 - **Notes:** Scaling preserves units of C_ℓ and matches later statements that D_ℓ is what's plotted for diagnostics.
2. ✓ **Temperature region ratio** (Eq. (2), Sec. 3.2, p.2)
 - **Claim:** Defines binned temperature autospectrum ratio $R_b^{TT} = \bar{C}_b^{TT,DS} / \bar{C}_b^{TT,DN}$.
 - **Checks:** definition consistency, dimensional/units sanity
 - **Verdict:** PASS; confidence: high; impact: moderate

- **Assumptions/inputs:** $\bar{C}_b^{\kappa\kappa}$ denotes binned bandpowers/averaged pseudo- $C_{\ell\ell}$ outputs, DN denominator is nonzero
 - **Notes:** Dimensionless ratio; consistent with text describing mean ratios and jackknife uncertainties.
3. ✓ **Lensing (κ) region ratio** (Eq. (3), Sec. 3.2, p.2)
- **Claim:** Defines binned convergence autospectrum ratio $R_b^\kappa = \bar{C}_b^{\kappa\kappa,DS} / \bar{C}_b^{\kappa\kappa,DN}$.
 - **Checks:** definition consistency, dimensional/units sanity
 - **Verdict:** PASS; confidence: high; impact: moderate
 - **Assumptions/inputs:** Both spectra are computed in a consistent normalization convention
 - **Notes:** Dimensionless and used consistently in Sec. 4.2 and the chi-square diagnostic.
4. △ **Observed lensing spectrum decomposition** (Eq. (4), Sec. 3.2, p.3)
- **Claim:** States schematically that $C_\ell^{\kappa\kappa,obs} \approx C_\ell^{\kappa\kappa,true} + N_\ell^{(0)} + \dots$
 - **Checks:** symbol consistency, logic/derivation completeness
 - **Verdict:** UNCERTAIN; confidence: medium; impact: minor
 - **Assumptions/inputs:** Equation is schematic (approximate) and meant to highlight additive reconstruction biases
 - **Notes:** As written it is an informal schematic rather than a derived relation; it is logically consistent with the narrative that differences in ratios can be driven by bias/noise terms, but the paper does not define precisely what convention for κ and $N^{(0)}$ is being used.
5. ✓ **Uncorrelated-bin chi-square severity index** (Eq. (5), Sec. 3.3, p.3)
- **Claim:** Defines $\chi^2 = \sum_b (R_b^\kappa - 1)^2 / (\sigma_{R_b^\kappa})^2$ under an uncorrelated-bin approximation.
 - **Checks:** algebra/structure, symbol consistency
 - **Verdict:** PASS; confidence: high; impact: moderate
 - **Assumptions/inputs:** Per-bin ratio errors $\sigma_{R_b^\kappa}$ are treated as independent, Gaussian approximation for the ratio error is implicitly assumed for interpretability
 - **Notes:** The intended formula is standard; typographical artifacts ('N Xbin') do not change the mathematical meaning. The text correctly cautions that the statistic is not distributed as $\chi_{N_{bin}}^2$ without the full covariance.
6. ✓ **Pearson correlation coefficient across bins** (Eq. (6), Sec. 3.4, p.3)
- **Claim:** Defines r_{xy} as the Pearson correlation of binned QE amplitudes x_b and y_b across multipole bins.

- **Checks:** algebra/structure, dimensional/units sanity, notation clarity
- **Verdict:** PASS; confidence: medium; impact: minor
- **Assumptions/inputs:** x_b and y_b are real-valued amplitudes across the same set of bins, The sums are over identical bin indices b
- **Notes:** The correlation is dimensionless as intended. The denominator shows 'p' characters in the extracted text where square roots should appear; assuming this is a rendering artifact, the structure matches Pearson r .

7. \triangle **Weighted mean lensing ratio statement** (Eq. (7), Sec. 4.2, p.3)

- **Claim:** Reports an inverse-variance weighted mean $\langle R^k \rangle = 0.8545 \pm 0.0021$.
- **Checks:** definition consistency
- **Verdict:** UNCERTAIN; confidence: low; impact: minor
- **Assumptions/inputs:** Inverse-variance weighting across bins is performed using the stated per-bin uncertainties
- **Notes:** This is a reported numerical summary rather than an analytic derivation. The paper does not provide the explicit weighted-mean formula, so only consistency of notation can be checked.

8. \times **Declination-split 'correlation coefficient' definition** (Eq. (8), Sec. 4.5, p.3–5)

- **Claim:** Defines $r_{\ell_b} \equiv \bar{C}_b^{\text{cross}} / (\bar{C}_b^{DS} \bar{C}_b^{DN})$ for declination-split masks and calls it a 'binwise correlation'.
- **Checks:** dimensional/units sanity, symbol/definition consistency
- **Verdict:** FAIL; confidence: high; impact: critical
- **Assumptions/inputs:** \bar{C}_b^{DS} and \bar{C}_b^{DN} are autospectra; \bar{C}_b^{cross} is a cross-spectrum for the two masked maps
- **Notes:** As written, r_{ℓ_b} has units of 1/(units of power) because the denominator is a product of two power spectra rather than a square root of their product; thus it is not a correlation coefficient and is not dimensionless. Additionally, \bar{C}_q^{cross} uses a subscript/superscript q that is not defined anywhere in the provided text.

9. \checkmark **Delete-one jackknife variance formula** (Eq. (B1), Appendix B, p.9)

- **Claim:** Defines $\text{Var}_{\text{dJK}}(\hat{\theta}) = \frac{N-1}{N} \sum_i (\hat{\theta}^{(i)} - \bar{\theta})^2$ with $\bar{\theta}$ the mean of the leave-one-out estimates.
- **Checks:** algebra/structure, definition consistency
- **Verdict:** PASS; confidence: high; impact: moderate
- **Assumptions/inputs:** $\hat{\theta}^{(i)}$ are computed by omitting split i , N is the number of splits (here $N = 4$)
- **Notes:** Formula is correct for delete-one jackknife variance and matches the stated use of four splits.

Limitations

- Audit is based only on the provided PDF text extraction; equations embedded purely as images or relying on figure-only annotations may not be fully captured.
- Several key results are reported as numerical summaries from archived .npz products; without explicit analytic formulas for those summaries (e.g., exact weighting, exact ratio error propagation), their correctness cannot be verified symbolically.
- The paper references external pipelines and conventions (e.g., QE normalization, reconstruction biases) but does not define them mathematically within this document; checks were therefore limited to internal consistency and dimensional sanity of the stated equations.

Numerical results audit

This section audits **numerical/empirical** consistency: reported metrics, experimental design, baseline comparisons, statistical evidence, leakage risks, and reproducibility.

Nine targeted internal-consistency and arithmetic checks were executed on reported numerical statements (rounding consistency across sections, repeated constants, simple ratio/arithmetic recomputations, and range consistency for approximate factor claims). All executed checks returned PASS within the stated tolerances.

Checked items

1. ✓ **CAND-01** (Page 1 (Abstract); also Page 3 §4.1; Page 9 Table 1)
 - **Claim:** Unweighted mean temperature power ratio across 10 bins is reported as $\langle C_{TT,DS}/C_{TT,DN} \rangle = 0.313 \pm 0.039$ (Abstract) and more precisely as 0.3131 ± 0.0393 (§4.1/Table 1).
 - **Checks:** rounding_consistency
 - **Verdict:** PASS
 - **Notes:** mean abs diff=0.0001, rms abs diff=0.0003 (abs_tol=0.0005).
2. ✓ **CAND-02** (Page 1 (Abstract) and Page 3 §4.1 and Page 9 Table 1)
 - **Claim:** Jackknife inverse-variance weighted mean temperature ratio is 0.295 ± 0.0004 (Abstract) and 0.2953 ± 0.0004 (§4.1/Table 1).
 - **Checks:** rounding_consistency
 - **Verdict:** PASS
 - **Notes:** mean abs diff=0.0003 (abs_tol=0.0005); sigma equal=True.
3. ✓ **CAND-03** (Page 3 §3.3 and §4.2; Page 9 Table 1)
 - **Claim:** Equation (5) uses $N_{\text{bin}} = 10$ bins; χ^2 is reported as 5168, and $\chi^2/N_{\text{bin}} \approx 517$.
 - **Checks:** ratio_check
 - **Verdict:** PASS

- **Notes:** computed=516.8, reported=517, abs diff=0.2 (abs_tol=0.5).
4. ✓ **CAND-04** (Page 1 (Abstract); Page 3 §4.2; Page 9 Table 1)
- **Claim:** Weighted mean lensing ratio is $\langle C_{\kappa\kappa,DS}/C_{\kappa\kappa,DN} \rangle = 0.8545 \pm 0.0021$ in multiple places.
 - **Checks:** repeated_constant_match
 - **Verdict:** PASS
 - **Notes:** value equal=True, uncertainty equal=True.
5. ✓ **CAND-05** (Page 3 §4.4; Page 7 Conclusions; Page 9 Table 1)
- **Claim:** RMS ratios: $\text{RMS}_b[\text{DS} \times \text{DN}]/\text{RMS}_b[\text{DS} \times \text{AA}] = 0.114$ at 90 GHz and 0.235 at 150 GHz; claimed to be a factor ~ 4 to ~ 9 below $\text{RMS}(\text{DS} \times \text{AA})$ depending on frequency.
 - **Checks:** derived_factor_range_check
 - **Verdict:** PASS
 - **Notes:** f90=8.77193, f150=4.25532, within[3.5,9.5]=True.
6. ✓ **CAND-06** (Page 3 §4.4; Page 9 Table 1)
- **Claim:** $\text{RMS}_b[\text{DS} \times \text{DN}]/\text{RMS}_b[\text{DN} \times \text{AA}] = 0.067$ at 90 GHz and 0.047 at 150 GHz.
 - **Checks:** simple_transform_check
 - **Verdict:** PASS
 - **Notes:** f90=14.9254, f150=21.2766; checks: f150>f90 and both>10 => True.
7. ✓ **CAND-07** (Page 5 §4.5; Page 7 Conclusions; Page 9 Table 1)
- **Claim:** Declination-split correlation summary: mean $\bar{r} \approx 0.00226$ with rms 0.00654 across $N \approx 0.0023$ bins; largest bin reaches $r_\ell \approx 0.0285$; conclusion says mean \bar{r}
 - **Checks:** rounding_consistency
 - **Verdict:** PASS
 - **Notes:** Optional $z = (r_{\max} - r_{\text{mean}})/\text{rms} = 4.01223$, in [3,6]=True. abs $(r_{\text{mean}} - r_{\text{mean,conclusion}}) = 4 \times 10^{-5}$ (abs_tol= 5×10^{-5}).
8. ✓ **CAND-08** (Page 9 Appendix B (Eq. B1))
- **Claim:** Delete-one jackknife variance formula states $\text{Var}_{\text{dJK}}(\hat{\theta}) = (N - 1)/N \cdot \Sigma_i(\hat{\theta}(i) - \bar{\theta})^2$ with $N = 4$.
 - **Checks:** constant_substitution_check
 - **Verdict:** PASS
 - **Notes:** prefactor=0.75, expected=0.75, exact_equal=True.
9. ✓ **CAND-09** (Page 1 (Abstract) and Page 2 §3.3)

- **Claim:** The paper states 'In ten multipole bins ...' and separately sets $N_{\text{bin}} = 10$ in the χ^2 definition.
- **Checks:** repeated_constant_match
- **Verdict:** PASS
- **Notes:** int(N_{bin} text)=10, int(N_{bin} equation)=10.

Limitations

- Audit performed using only the provided PDF text; referenced .npz archives and any underlying numeric arrays are not accessible here, so many recomputations (means, RMS across bins, χ^2 from per-bin ratios, correlations) cannot be directly verified.
- No values were extracted from plot graphics; checks avoid reading pixel data from figures as requested.
- Where the paper uses approximate language (\approx , $O(\cdot)$, 'factor of'), only coarse arithmetic/range checks are proposed; exact validation would require the underlying binned data.