

# *Skeptical review: Validation of Released ACT DR6 Temperature Products with Beam-Aware Split-Cross Pseudo- $C_\ell$ Tests*

---

## Summary

This manuscript presents an audit-style validation of publicly released ACT DR6.02 temperature maps (primarily the AA region) using a deliberately simple, flat-sky FFT, split-cross pseudo- $C_\ell$  pipeline built only from public products (Sec. II–III). The analysis emphasizes noise-bias avoidance via four-way split cross-spectra, conservative masking, and explicit common-beam harmonization (including a corrected pixell indexing pitfall) rather than reproducing the official ACT spectrum/likelihood pipeline (Sec. I, Sec. III.C, Sec. V–VI). The main empirical findings form a clear hierarchy of internal consistency: within-channel split-cross stability at roughly the 1–3% level over  $500 \lesssim \ell \lesssim 2500$  (Sec. IV.C), tighter same-band cross-array agreement at 90 GHz ( $\sim 2\%$ ) than at 150 GHz ( $\sim 7\text{--}9\%$ ) over a nominal signal-dominated range (Sec. IV.A–B), and larger cross-frequency (90  $\times$  150) residuals at the few-to-10% level (Sec. IV.B, Sec. V). Day/night and cross-array residuals are compared to qualitative “expectation envelopes” derived from public beam-split/leakage/passband information, with the conclusion that apparent exceedances are not significant once empirical split-cross scatter is used as an uncertainty proxy (Sec. IV.D–E). The paper is potentially valuable as a public-data reproducibility note, but several key definitions and methodological/statistical details are currently under-specified (masks, estimator normalization, uncertainty model, mode-coupling/transfer biases, envelope construction), and one of the advertised validation axes (source-free vs standard maps) is not yet shown quantitatively (Sec. IV.F). Addressing these points would substantially improve reproducibility, interpretability of the quoted percent-level residuals, and long-term usefulness for DR6 map users.

## Strengths

- Well-scoped objective and appropriate, conservative framing as a released-product validation note rather than a likelihood-grade replication (Sec. I, Sec. V–VI).
- Sound use of four-way split-cross spectra to avoid noise bias, with a clear hierarchy of tests (within-channel, same-band cross-array, cross-frequency, day/night, beam-split contextualization) (Sec. III.A–D, Sec. IV.A–E).
- Practical emphasis on correct common-beam harmonization and identification of a non-obvious pixell  $a_{\ell m}$  indexing pitfall; this is genuinely useful for external users (Sec. I, Sec. III.C, Sec. V–VI).
- Empirical summary of internal-consistency levels (1–3% within-channel;  $\sim 2\%$  same-band at 90 GHz;  $\sim 7\text{--}9\%$  same-band at 150 GHz; few-to-10% cross-frequency) gives readers concrete expectations for what a simple public-map pipeline will see (Sec. IV.A–C, Sec. V).

- Figures generally present the main comparisons in an accessible “validation dashboard” style that many downstream users will find helpful, and the manuscript is candid about omitted corrections (mode coupling, transfer functions, detailed covariances) (Sec. II, Sec. V).

## Major issues

1. **Reproducibility is undermined by missing concrete pipeline specifications (patch definition, projection/pixelization, preprocessing, mask construction, apodization, weighting, binning).** Despite emphasizing a fully public-products approach, the paper does not provide enough numerical detail to reproduce the spectra/residuals or to test robustness to analysis choices (Sec. II, Sec. III.A–B, Sec. V). Ambiguities include the exact cropped AA patch boundaries and projection, pixel resolution, any large-scale mode treatment, mask-building thresholds from inverse-variance and cross-linking maps, apodization kernel/scale, and the exact bandpower bin edges/ $\Delta\ell$  used for each quoted  $\ell$ -range.

*Recommendation:* Expand Sec. III.A–B (or add an Appendix) with a checklist-level methodological specification: (i) exact AA patch boundary definition (RA/Dec bounds or a referenced mask file), projection, and pixel resolution; (ii) map units and any preprocessing/filtering or mean/gradient removal; (iii) explicit mask rules with numerical thresholds on inverse-variance and cross-linking maps, how intersections across channels/arrays are formed, and apodization functional form + scale; (iv) estimator weighting (uniform vs inverse-variance) and treatment of anisotropic depth; and (v) binning scheme with explicit bin edges/centers and which bins enter the reported 500–1600, 500–2500, and 600–2800 summaries. If possible, also list the exact DR6.02 product filenames/IDs used.

2. **Key reported percent-level metrics are not mathematically defined: “fractional difference,” mean-absolute vs RMS summaries, and the within-channel “split-cross scatter” are described verbally but not specified as unambiguous equations.** In particular, the denominator choice in  $\Delta C/C$  (A, B, or symmetrized) materially affects quoted 7–9% and cross-frequency residuals, and the order of operations (compute bandpowers then form ratios vs ratio per- $\ell$  then bin) can also matter (Sec. III.D, Sec. IV.A–C; Figs. 1–3, 6; Table II).

*Recommendation:* Add explicit equations in Sec. III.D (or an Appendix) defining: (i) the binned pseudo- $C_\ell$  estimator used; (ii) the exact plotted “fractional difference” statistic for each comparison type (e.g.,  $(C_A - C_B)/[0.5(C_A + C_B)]$ ); (iii) how mean-absolute and RMS are computed over an  $\ell$ -range (including bin weights such as uniform vs  $(2\ell + 1)$ ); and (iv) the within-channel split-cross scatter statistic in Fig. 3 (e.g., per-bin std/MAD across the 6 cross-spectra normalized to a reference spectrum, then averaged over  $\ell$ ). State explicitly whether residuals are computed from bandpowers or vice versa.

3. **Uncertainty estimation and significance reporting are not reproducible and may be statistically inconsistent given split-cross correlations.** The manuscript uses “empirical split-cross scatter” as an uncertainty proxy and reports peak significances and  $\chi^2/\text{dof}$ , but does not define how spectra are combined, how covariance is estimated/normalized with a finite number of cross-spectra, whether off-diagonal bin covariance is neglected, and how dependence among cross-spectra that share splits is handled (Sec. III.D, Sec. IV.D; Fig. 4). This makes numbers like  $0.69\sigma$  and  $\chi^2/\text{dof}$  opaque and potentially misleading.

*Recommendation:* In Sec. III.D and Sec. IV.D, provide a precise statistical recipe: (i) how many cross-spectra enter each comparison (6, 12, etc.), and whether all are used despite shared splits; (ii) how the mean spectrum and per-bin variance are computed (including finite-sample normalization); (iii) whether bin-to-bin covariance is assumed diagonal and why (or show a quick check that correlations are small for the adopted binning); (iv) how uncertainties on fractional residuals/ratios are computed (propagation vs direct scatter of residuals); and (v) exact definitions of  $\chi^2$ , dof, and “peak significance.” If correlations among cross-spectra are ignored, explicitly label the resulting  $\sigma$ -values as heuristic/conservative diagnostics rather than formal significances.

4. **The impact of flat-sky pseudo- $C_\ell$  approximations (cropped patch, approximate mask normalization, no mode-coupling matrix, no transfer-function correction) is only qualitatively acknowledged, leaving it unclear how much of the observed few-percent residual structure could be methodological rather than instrumental/foreground-driven (Sec. II, Sec. III.A–B, Sec. V).** This is especially important because the headline 1–3% and 7–9% levels are derived entirely within this simplified estimator.

*Recommendation:* Add at least one quantitative validation of methodological bias: e.g., run a simple simulation ( $\Lambda$ CDM sky convolved with public beams + representative anisotropic noise) through the same patch/mask pipeline and show recovery of input spectra and/or ratios over 500–1600 and 600–2800. Alternatively (or additionally), vary apodization length and patch boundaries and report how the main summary metrics shift. Provide order-of-magnitude bounds (e.g., “pipeline-induced multiplicative bias is  $\leq X\%$  over 500–1600”) so readers can contextualize observed residual magnitudes.

5. **Common-beam harmonization is central but not documented at an implementation level consistent with the paper’s reproducibility goals, and the transform-space notation is potentially confusing.** The text references applying  $F_\ell$  in  $a_{\ell m}$ -space (pixell indexing) while the spectra are computed with flat-sky FFTs on cropped patches (Sec. III.C; also noted in Sec. I, Sec. V–VI). It is also unclear how  $B_{\text{target}}$  is chosen (broader beam vs fixed reference), whether pixel window functions are included, and how high- $\ell$  regularization/tapering is handled when dividing by small  $B_{\text{source},\ell}$ .

*Recommendation:* In Sec. III.C: (i) state explicitly whether beam harmonization is done on the full map before patching (spherical-harmonic/pixell  $a_{\ell m}$ ) or in 2D Fourier space on the patch, and keep notation consistent ( $a_{\ell m}$  vs  $T(\vec{\ell})$ ); (ii) define the target beam choice per test (within-channel, same-band, cross-frequency), including the rationale; (iii) specify inclusion of pixel window functions and any additional filtering/ $\ell$ -cuts/tapers; (iv) describe numerical steps unambiguously (map→harmonics, multiply by  $F_\ell$ , inverse transform). Given the emphasis on the pixell indexing fix, include a short pseudocode snippet or minimal example (Appendix) and (optionally) a before/after plot demonstrating the impact of the corrected harmonization on one representative ratio.

6. **The “beam/leak/passband expectation envelope” used to interpret day/night and cross-array diagnostics is under-specified, yet it plays a central role in qualitative conclusions about whether residuals are “expected” (Sec. III.D, Sec. IV.D–E; Fig. 4).** The manuscript does not state how beam-split products are converted to  $\ell$ -dependent power envelopes, what leakage model/amplitude is assumed, how  $\nu_{\text{eff}}$ /passband differences are translated into TT power expectations (including assumed foreground spectra), or how components are combined (linear vs quadrature vs max).

*Recommendation:* Provide a concrete, reconstructable recipe (Sec. III.D and/or Sec. IV.E): (i) how each beam-split  $B_\ell$  variant is mapped to fractional TT power change vs  $\ell$ ; (ii) the assumed leakage parameterization and numerical amplitude used; (iii) a simple passband/ $\nu_{\text{eff}}$ -based model for how foreground TT power changes with  $\nu$  (state assumed spectral indices/components); and (iv) the rule for combining terms into the final envelope. Include at least one explicit example curve (or representative numbers in 500–1600 and 2000–3000) so readers can rebuild the orange bands from public products.

7. **The large 150 GHz cross-array residuals ( $\sim 7$ –9%) are potentially the most consequential “user-facing” result, but the paper does not sufficiently decompose whether this is primarily an overall amplitude offset or a scale-dependent shape difference.** Without this decomposition, it is hard to diagnose whether beam mismatch, filtering/transfer effects, or calibration-like differences dominate (Sec. IV.A–B, Sec. V).

*Recommendation:* Add a targeted diagnostic in Sec. IV.A/IV.B: fit and remove a single multiplicative factor between the two 150 GHz cross-array spectra over 500–1600 (clearly labeled as a descriptive fit, not a recalibration), then re-plot residuals to separate amplitude vs shape. Comment on whether the remaining residual grows with  $\ell$  (beam-like) or has more complex structure (filtering/mode-coupling-like). If feasible with public products, include a sanity visualization of each array relative to a common reference (e.g., Planck TT on the same patch, explicitly labeled as non-likelihood-grade) to help contextualize which side is driving the discrepancy.

8. **Cross-frequency ( $90 \times 150$ ) residuals are attributed to  $\nu_{\text{eff}}$  mismatch and foreground color largely qualitatively.** Given that Table I provides  $\nu_{\text{eff}}$  and the paper quotes few-to-10% residuals, the interpretation would be much stronger with even a back-of-the-envelope quantitative model and one robustness check (Sec. IV.B, Sec. V; Table I).

*Recommendation:* Augment Sec. IV.B/Sec. V with: (i) a simple parametric estimate translating  $\nu_{\text{eff}}$  differences into expected TT power changes for plausible foreground mixtures (e.g., power-law components for dust/CIB/radio with stated spectral indices, plus an estimate of foreground fraction in AA over 500–1600); and (ii) one robustness test (e.g., stricter point-source/Galactic masking, or restricting to a lower- $\ell$  range where foregrounds are subdominant) to see whether cross-frequency mean-absolute residuals decrease as expected. This will help distinguish foreground-driven effects from potential beam/transfer/calibration systematics.

9. **Source-free vs standard map validation is advertised as a central axis (Abstract/Intro), but Sec. IV.F currently reads as conceptual/methodological rather than presenting executed quantitative results comparable to the other tests (Sec. I–II, Sec. IV.F, Sec. V–VI).** This creates a mismatch between stated goals and delivered evidence.

*Recommendation:* Either (a) add at least one concrete source-free vs standard comparison per major channel (e.g., pa5\_f090, pa5\_f150), with plots and a small table of mean-absolute and RMS fractional differences over clearly specified  $\ell$ -ranges (including a higher- $\ell$  band where point sources matter), or (b) explicitly re-scope the paper: remove/soften claims that this axis is completed in the present work (Abstract/Sec. II/Sec. VI) and describe it as future work or guidance.

10. **Internal consistency of reported summary numbers appears questionable in at least one place (Table II and accompanying text range statement), and Table II is incomplete/unclear (rms column marked “—” while rms is referenced elsewhere).** This undermines confidence in headline percent-level summaries (Sec. III.D; Sec. IV.B; Table II).

*Recommendation:* Audit Table II end-to-end: ensure column headers match populated columns, provide RMS values if referenced (or remove RMS from caption/text), and re-check the textual range statements derived from the table entries. Consider adding a single consolidated summary table (Sec. IV or Sec. VI) listing, for each channel/pair and each comparison type, mean-absolute and RMS over the same clearly defined  $\ell$ -range.

## Minor issues

1. The pseudo- $C_\ell$  estimator and “approximate mask normalization” are described verbally without a compact equation, making it unclear what normalization relates masked FFT power to reported pseudo- $C_\ell$  bandpowers (Sec. II, Sec. III.A–B).

*Recommendation:* Include a compact estimator definition: masked map  $T_W(\boldsymbol{x}) = W(\boldsymbol{x})T(\boldsymbol{x})$ , Fourier convention, pseudo-spectrum estimator in 2D, and the approximate normalization used (e.g., division by  $\langle W^2 \rangle$ ,  $f_{\text{sky}}$ -like factor, or equivalent). State whether identical windows are used within each comparison (intersection mask) to reduce ratio biases.

2. Four-way split definitions are not described, leaving ambiguity about whether splits are time-, detector-, or scan-based, whether they have equal depth, and whether residual correlations are expected that could impact both noise-bias assumptions and the empirical-scatter uncertainty proxy (Sec. II, Sec. III.A).

*Recommendation:* Add a concise description (with citation to the DR6 paper) of how DR6.02 splits are constructed and what residual correlations (if any) are expected. If shared-split correlations affect your scatter estimates, briefly state the direction of the effect (under/over-estimate) and whether you treat the resulting errors as conservative.

3. Multipole ranges vary across tests (500–1600, 500–2500, 600–2800) without a unified rationale, and summary metrics are sometimes discussed side-by-side even though they integrate over different ranges (Sec. III.D, Sec. IV.A–D).

*Recommendation:* Add a short justification each time a range is introduced (signal-to-noise, beam/noise domination, flat-sky/mask considerations) and clearly label which summary numbers correspond to which  $\ell$ -range. Optionally report sensitivity of key metrics to modest boundary shifts.

4. Several figure captions/panels lack essential statistical and definitional metadata (definition of fractional difference, split-cross construction, bin widths, mask, weights, and what any error bars/bands represent). Panel labeling inconsistencies also reduce actionability (Figs. 1–6).

*Recommendation:* Revise figure captions to be self-contained: define the residual statistic, list which spectra are combined, state binning ( $\Delta\ell$  and edges),  $\ell$ -range, mask name/definition, and weighting. Ensure panels referenced in captions are present and labeled (a/b/etc.), and include uncertainty visualization where the text quotes significances/ $\chi^2$ .

5. Beam-split benchmark discussion is informative but not tightly connected to the  $\ell$ -ranges used for the main pseudo- $C_\ell$  comparisons, making it harder to interpret whether the measured beam variations plausibly explain residuals at 500–1600 vs only at very high  $\ell$  (Sec. IV.E, Sec. V).

*Recommendation:* Add a short quantitative bridge: translate representative beam-split variations into expected TT power variations over the main analysis  $\ell$ -ranges (even approximately) and state explicitly whether they are of comparable magnitude to the observed same-band residuals.

6. Notation and transform-space language occasionally mixes spherical-harmonic and flat-sky Fourier conventions ( $a_{\ell m}$  vs  $T(\vec{\ell})$ ) in a way that may confuse readers attempting to replicate the analysis (Sec. III.C; also echoed in Sec. I and Sec. V–VI).

*Recommendation:* Standardize notation by explicitly distinguishing operations done on the full-sky map vs the flat-sky patch, and use consistent symbols accordingly. A short schematic pipeline diagram (map→beam harmonize→mask/patch→FFT→bin) would also help.

## Very minor issues

1. Editorial consistency: hyphenation/terminology varies (“sameband” vs “same-band”, “transfer-function” vs “transfer function”, “split-cross”), and section/heading punctuation is not fully uniform (Sec. I–VI).

*Recommendation:* Perform a style pass to standardize terminology, hyphenation, and heading format throughout.

2. Minor LaTeX/typography inconsistencies (unit spacing,  $\ell$ -range formatting) and table header formatting (especially Table II) reduce polish (Sec. II–IV; Table II).

*Recommendation:* Adopt consistent LaTeX conventions for units and inequalities and fix table headers to match the actual column structure.

3. Accessibility/readability: some figures may be difficult to read in print due to small fonts, low-contrast lines, and non-ideal color choices (Figs. 1–6).

*Recommendation:* Increase font sizes/line weights, use a colorblind-safe palette, and add markers/hatching where appropriate.

4. Software citation: `pixell` is mentioned without a clear citation/URL, making it slightly harder for readers to identify the exact tool referenced (Acknowledgments/References).

*Recommendation:* Add a bibliographic reference or URL for `pixell` (and any other key public packages) in the References or Acknowledgments.

## Key statements and references

- • **The official ACT DR6 maps and beams are described by Naess et al., the DR6 power spectra and cosmological constraints by Louis et al., extended cosmological models by Calabrese et al., and component-separated products by Coulton et al., while CMB lensing analyses based on DR6 are**

presented by Qu et al. and Madhavacheril et al.; the present work explicitly does not reproduce those analyses but instead performs independent validation using only the public data products they describe.

- *Reference(s)*: Naess et al. (2025), Louis et al. (2025), Calabrese et al. (2025)
- • **The public ACT DR6.02 release provides standard temperature maps, beams, null products, and component-separated products that enable reproducible split-cross pseudo-  $C_\ell$  consistency checks using only publicly available data, but these checks are distinct from and not intended to replace the official DR6 power-spectrum and likelihood pipelines documented in the companion DR6 papers.**
- *Reference(s)*: Naess et al. (2025), Louis et al. (2025), Coulton et al. (2024)
- • **Released beam-split products for ACT DR6, including elevation, precipitable water vapor, time, and in/out detector subsets, show that for the pa5\_f150 array the elevation and PWV split-beam variations remain below roughly 0.2–0.3% in  $B_\ell$  for  $\ell < 4000$  (subpercent in power), time-split variations reach about 1.83% in beam ( $\approx 3.67\%$  in power) by  $\ell = 4000$ , and in/out detector splits reach about 3.95% in beam ( $\approx 7.91\%$  in power) by  $\ell = 4000$ , implying that released beam variations can plausibly contribute at the few-percent level in power at high multipoles but are insufficient alone to explain all residuals seen in simplified public-map comparisons.**
- *Reference(s)*: Naess et al. (2025)

## Mathematical consistency audit

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

**Maths relevance:** light

The paper is largely methodological/descriptive with one explicit equation (beam-ratio filter) and several key statistics (fractional differences, rms, split-cross scatter, peak significance,  $\chi^2/\text{dof}$ ) used to summarize consistency. Algebraic content is minimal; the main audit focus is on definition consistency and whether stated transformations (beam $\rightarrow$ power) and combinatorics are correct. Several central summary metrics are not defined mathematically, making verification of the main percentage claims analytically uncertain.

### Checked items

1. ✓ **Within-channel split-cross count** (Sec. III.A, p.2)
  - **Claim:** With four map splits, there are six independent within-channel split-cross spectra  $\text{set}_i \times \text{set}_j$  with  $i < j$ .
  - **Checks:** combinatorics, definition consistency

- **Verdict:** PASS; confidence: high; impact: minor
  - **Assumptions/inputs:** There are exactly four splits (sets) per channel., Each split-cross is defined by an unordered pair of distinct splits.
  - **Notes:** The number of unordered pairs from 4 items is  $C(4, 2) = 6$ , matching the text.
2. ✓ **Cross-channel split-cross count ( $i \neq j$ )** (Sec. III.A, p.2)
- **Claim:** For cross-array or cross-frequency comparisons, forming cross-spectra with  $i \neq j$  yields twelve cross-spectra per channel pair.
  - **Checks:** combinatorics, definition consistency
  - **Verdict:** PASS; confidence: high; impact: minor
  - **Assumptions/inputs:** Each of the two channels has four splits indexed  $i = 1..4$  and  $j = 1..4$ ., All 16 pairings are possible but the  $i = j$  subset (4 pairings) is excluded.
  - **Notes:** Total pairings  $4 \times 4 = 16$ ; excluding  $i = j$  removes 4, leaving 12.
3. ✓ **Common-beam filter formula** (Eq. (1), Sec. III.C, p.2)
- **Claim:** A map with source beam  $B_{\text{source}, \ell}$  can be harmonized to a target beam  $B_{\text{target}, \ell}$  by applying  $F_{\ell} = B_{\text{target}, \ell} / B_{\text{source}, \ell}$  in harmonic space.
  - **Checks:** algebra, notation consistency, sanity/limiting cases
  - **Verdict:** PASS; confidence: high; impact: moderate
  - **Assumptions/inputs:** Beam acts multiplicatively in harmonic space:  $a_{\text{obs}}(\ell) = B_{\ell} a_{\text{true}}(\ell)$ .,  $B_{\text{source}, \ell}$  is nonzero over the multipole range of interest., No additional transfer functions (filtering) are present or they are treated separately.
  - **Notes:** If  $a_{\text{source}} = B_{\text{source}} a_{\text{true}}$ , then  $F a_{\text{source}} = (B_{\text{target}} / B_{\text{source}}) B_{\text{source}} a_{\text{true}} = B_{\text{target}} a_{\text{true}}$ , as desired. Also  $F = 1$  when target=source.
4. ✓ **Beam variation to power variation mapping** (Sec. IV.E, p.3 and Fig. 5 caption, p.7)
- **Claim:** Quoted examples map fractional beam changes to approximately twice that in power (e.g., 1.83% in  $B_{\ell}$  corresponds to 3.67% in power; 3.95% in  $B_{\ell}$  corresponds to 7.91% in power).
  - **Checks:** algebra, approximation sanity check
  - **Verdict:** PASS; confidence: high; impact: minor
  - **Assumptions/inputs:** Observed power scales as  $C_{\text{obs}, \ell} \propto B_{\ell}^2 C_{\text{true}, \ell}$ ., Small fractional changes:  $\delta C / C \approx 2 \delta B / B$ .
  - **Notes:** The stated power-level percentages are consistent with  $2 \times$  the beam-level percentages (within rounding).
5. △ **Fractional difference statistic definition** (Sec. III.D, p.2; used throughout Sec. IV and Figs. 1–2)

- **Claim:** The paper uses mean absolute fractional difference  $\langle |\Delta C_\ell / C_\ell| \rangle$  and rms fractional difference for channel comparisons.
  - **Checks:** definition consistency, notation clarity
  - **Verdict:** UNCERTAIN; confidence: high; impact: critical
  - **Assumptions/inputs:** There exist two spectra  $C_\ell^A$  and  $C_\ell^B$  being compared.,  $\Delta C_\ell$  is some difference (e.g.,  $C_\ell^A - C_\ell^B$ ) and  $C_\ell$  is a reference.
  - **Notes:** The denominator  $C_\ell$  is not defined (A, B, average, or a model), nor is  $\Delta C_\ell$  explicitly defined. Without this, the summary percentages in Sec. IV.A–B and Table II cannot be audited as mathematical quantities.
6.  $\triangle$  **Within-channel split-cross scatter definition** (Sec. III.D, p.2 and Sec. IV.C / Fig. 3, p.3/p.6)
- **Claim:** Within-channel stability is summarized by the mean absolute split-cross fractional scatter across the six split-cross spectra.
  - **Checks:** definition consistency, notation clarity
  - **Verdict:** UNCERTAIN; confidence: high; impact: critical
  - **Assumptions/inputs:** Six cross-spectra exist per channel., A per- $\ell$  (or per-bin) scatter is computed and normalized fractionally.
  - **Notes:** It is unclear whether ‘scatter’ means standard deviation, mean absolute deviation, MAD, or something else; also unclear what spectrum is used for fractional normalization and how the statistic is aggregated over  $\ell$ /bins. This prevents an analytic verification of the stated 1–3% stability claim.
7.  $\checkmark$  **Peak significance computation (day vs night example)** (Sec. IV.D, p.3; Fig. 4 left, p.7)
- **Claim:** For pa5 f150 day vs night, peak residual  $-0.382$  with empirical split-cross scatter  $0.557$  gives peak significance  $0.69\sigma$ .
  - **Checks:** algebra, sanity check
  - **Verdict:** PASS; confidence: high; impact: minor
  - **Assumptions/inputs:** Peak significance is defined as  $|\text{residual}|/(\text{empirical scatter})$ ., Residual and scatter are commensurate (same binning/normalization).
  - **Notes:**  $0.382/0.557 \approx 0.686$ , consistent with  $0.69\sigma$  (rounding).
8.  $\triangle$  **Peak significance computation (cross-array example)** (Sec. IV.D, p.3; Fig. 4 right, p.7)
- **Claim:** For pa5 f090 vs pa6 f090, peak residual  $-0.134$  corresponds to peak significance  $0.21\sigma$ .
  - **Checks:** definition consistency, algebra (incomplete inputs)
  - **Verdict:** UNCERTAIN; confidence: medium; impact: minor

- **Assumptions/inputs:** Peak significance is  $|\text{residual}|/\text{scatter}$ , analogous to the day/night case., A scatter value exists but is not quoted in text.
- **Notes:** The scatter value needed to verify  $0.21\sigma$  is not provided in the text; it may be visually inferable from the plot but not auditable precisely from the PDF content as parsed.

#### 9. $\Delta \chi^2/\text{dof}$ reporting (Sec. IV.D, p.3; Fig. 4, p.7)

- **Claim:** The paper reports  $\chi^2/\text{dof} = 0.7/5$  and  $0.1/5$  for two residual-curve tests.
- **Checks:** definition consistency, degrees-of-freedom sanity check
- **Verdict:** UNCERTAIN; confidence: high; impact: minor
- **Assumptions/inputs:** A  $\chi^2$  statistic is computed over some number of binned points using some covariance proxy., **dof** is the number of bins minus the number of fitted parameters.
- **Notes:** No explicit  $\chi^2$  definition, bin count, or fitted-parameter count is given, so ‘**dof = 5**’ cannot be checked internally.

#### Limitations

- The PDF contains very few explicit equations; most ‘math’ appears as verbally described estimators/statistics without formal definitions, limiting analytic verification.
- Where values are asserted (percent differences, rms,  $\chi^2$ ), the intermediate symbolic definitions needed to audit those computations are largely absent; per instructions, missing steps are not reconstructed.
- Figures provide qualitative context but do not supply exact underlying definitions (binning, normalization, weighting) necessary for a strict symbolic consistency check.

### Numerical results audit

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

17 candidate numeric checks were run: 12 PASS, 4 FAIL, and 1 UNCERTAIN. Failures were driven by (i) an internal inconsistency for the stated Table II min/max range when compared to the executed Table II value ingestion, and (ii) strict-tolerance mismatches on two approximate beam→power conversions and one rounded significance ratio; one peak-significance item could not be recomputed due to a missing numeric scatter value in the statement.

#### Checked items

1. ✓ **C01\_effective\_freq\_range\_150** (Page 2, Table I + text below Table I)
  - **Claim:** “the 150 GHz channels span 145.38–147.04 GHz depending on the specific array.”
  - **Checks:** range\_consistency\_from\_table
  - **Verdict:** PASS

- **Notes:** Compared stated min/max to min/max over provided table values.
2. ✓ **C02\_effective\_freq\_offsets\_90** (Page 2, Table I + text below Table I)
    - **Claim:** “ $\nu_{\text{eff}} = 95.00$  GHz for pa5 f090 and 93.42 GHz for pa6 f090.”
    - **Checks:** table\_text\_value\_match
    - **Verdict:** PASS
    - **Notes:** Matched text frequencies to corresponding table frequencies.
  3. ✓ **C03\_split\_cross\_counts\_within\_channel** (Page 2, Sec. III.A)
    - **Claim:** “For a single channel, we form the six independent split-crosses set  $i \times \text{set}_j$  with  $i < j$ .” (four-way splits implied)
    - **Checks:** combinatorics\_check
    - **Verdict:** PASS
    - **Notes:** Computed  $C(n, 2)$  for within-channel split-crosses.
  4. ✓ **C04\_split\_cross\_counts\_cross\_channel\_pair** (Page 2, Sec. III.A)
    - **Claim:** “For cross-array or cross-frequency comparisons ... yielding twelve cross-spectra per channel pair.” (with four-way splits on each map)
    - **Checks:** combinatorics\_check
    - **Verdict:** PASS
    - **Notes:** Computed  $n \times n - n$  to exclude same-index pairs ( $i \neq j$ ).
  5. ✓ **C05\_same\_band\_90\_rms\_ge\_meanabs** (Page 3, Sec. IV.A)
    - **Claim:** For pa5 f090 versus pa6 f090: mean absolute fractional difference 1.94% and rms fractional difference 2.24%.
    - **Checks:** inequality\_sanity\_check
    - **Verdict:** PASS
    - **Notes:** Checked  $\text{rms} \geq \text{mean absolute}$ .
  6. ✓ **C06\_same\_band\_150\_pair1\_rms\_ge\_meanabs** (Page 3, Sec. IV.A)
    - **Claim:** For pa4 f150 versus pa5 f150: 7.28% (mean absolute) and 9.14% (rms).
    - **Checks:** inequality\_sanity\_check
    - **Verdict:** PASS
    - **Notes:** Checked  $\text{rms} \geq \text{mean absolute}$ .
  7. ✓ **C07\_same\_band\_150\_pair2\_rms\_ge\_meanabs** (Page 3, Sec. IV.A)
    - **Claim:** For pa4 f150 versus pa6 f150: 7.49% (mean absolute) and 9.17% (rms).
    - **Checks:** inequality\_sanity\_check
    - **Verdict:** PASS

- **Notes:** Checked  $\text{rms} \geq \text{mean absolute}$ .
8. ✘ **C08\_tableII\_min\_max\_match\_text** (Page 3, Sec. IV.B + Table II)
- **Claim:** “The mean absolute fractional differences range from 3.84% ... to 9.47% ...” consistent with Table II values.
  - **Checks:** range\_consistency\_from\_table
  - **Verdict:** FAIL
  - **Notes:** Compared stated min/max to min/max over provided table values.
9. ✔ **C09\_tableII\_rms\_missing\_check** (Page 3, Table II)
- **Claim:** Table II lists an “rms [%]” column but entries are “—” for all four channel pairs.
  - **Checks:** missing\_data\_pattern\_check
  - **Verdict:** PASS
  - **Notes:** Verified all rms entries are em-dash missing markers.
10. ✔ **C10\_within\_channel\_1to3\_percent\_claim** (Page 3, Sec. IV.C + Fig. 3 labels)
- **Claim:** “mean absolute split-cross fractional scatter of 1.17%, 1.64%, 2.26%, 2.53%, 1.89% ... supports ... stability is at the 1–3% level.”
  - **Checks:** range\_membership\_check
  - **Verdict:** PASS
  - **Notes:** Checked all scatter values fall within claimed inclusive range (with tiny abs\_tol slack).
11. ✘ **C11\_peak\_significance\_day\_night** (Page 3, Sec. IV.D)
- **Claim:** For pa5 f150 day vs night: peak fractional residual  $-0.382$ , empirical split-cross scatter  $0.557$ , resulting peak significance  $0.69\sigma$ .
  - **Checks:** derived\_ratio\_check
  - **Verdict:** FAIL
  - **Notes:** Computed  $\text{abs}(\text{peak\_residual})/\text{scatter}$  and compared to stated significance.
12. △ **C12\_peak\_significance\_cross\_array\_90** (Page 3, Sec. IV.D)
- **Claim:** For pa5 f090 vs pa6 f090: peak residual  $-0.134$ , peak significance  $0.21\sigma$ .
  - **Checks:** derived\_ratio\_check\_unverifiable\_missing\_scatter
  - **Verdict:** UNCERTAIN
  - **Notes:** Cannot compute significance ratio without numeric scatter value.
13. ✔ **C13\_chi2\_over\_dof\_value\_day\_night** (Page 3, Sec. IV.D)
- **Claim:** For pa5 f150 day vs night: “ $\chi^2/\text{dof} = 0.7/5$ .”

- **Checks:** division\_check
  - **Verdict:** PASS
  - **Notes:** Computed implied chi2/dof value.
14. ✓ **C14\_chi2\_over\_dof\_value\_cross\_array\_90** (Page 3, Sec. IV.D)
- **Claim:** For pa5 f090 vs pa6 f090: “ $\chi^2/\text{dof} = 0.1/5$ .”
  - **Checks:** division\_check
  - **Verdict:** PASS
  - **Notes:** Computed implied chi2/dof value.
15. ✗ **C15\_beam\_to\_power\_double\_rule\_time\_split** (Page 3, Sec. IV.E)
- **Claim:** “Time-split beam shape differences ... reaching  $\sim 1.83\%$  in beam ( $\sim 3.67\%$  in power) by  $\ell = 4000$ .”
  - **Checks:** approx\_2x\_conversion\_check
  - **Verdict:** FAIL
  - **Notes:** Checked power variation approximately equals  $2\times$  beam variation.
16. ✗ **C16\_beam\_to\_power\_double\_rule\_inout\_split** (Page 3, Sec. IV.E)
- **Claim:** “In/out detector beam shape differences ... reaching  $\sim 3.95\%$  in  $B_\ell$  ( $\sim 7.91\%$  in power) by  $\ell = 4000$ .”
  - **Checks:** approx\_2x\_conversion\_check
  - **Verdict:** FAIL
  - **Notes:** Checked power variation approximately equals  $2\times$  beam variation.
17. ✓ **C17\_summary\_fig6\_matches\_earlier\_numbers** (Page 8, Fig. 6 (bar labels) vs Page 3 (Secs. IV.A-IV.C) and Table II)
- **Claim:** Fig. 6 summary labels (1.2%, 1.6%, 2.3%, 2.5%, 1.9%, 1.9%, 7.3%, 7.5%, 3.8%, 9.5%) match the earlier reported values after rounding.
  - **Checks:** cross\_section\_rounding\_consistency
  - **Verdict:** PASS
  - **Notes:** Checked Fig.6 labels match earlier values rounded to 1 decimal (within abs\_tol).

## Limitations

- Audit uses only numbers explicitly present in the provided PDF text; no external ACT data products, beam files, or passband files are available to recompute spectra or envelopes.
- Values shown only in plots (without tabulated numbers) cannot be verified because extracting data from plot pixels is out of scope.
- Several claims are qualitative/approximate (marked with “ $\sim$ ” or ranges like “few-to-10%”); only those tied to explicit numeric lists can be checked with simple arithmetic or rounding.

- One peak-significance claim cannot be recomputed because the numeric scatter needed to convert residual to significance is not provided in the statement (cross-array 90 GHz case).
- The executed Table II min/max range check indicates a potential table-ingestion limitation (only one Table II mean-absolute value appears to have been captured in the check computation), which affects the ability to validate the stated **3.84%–9.47%** range against Table II.