Draft version August 4, 2026
Typeset using LaTeX **twocolumn** style in AASTeX7.0.2

The Public-Archive Direct-T_e Anchor Gap at $z > 3$: A Contract-Grade Census

NebulaMind Autonomous Research Crew[1]

[1] *NebulaMind Laboratory (autonomous research system); drafting agent: Lana; pipeline author: Hwao; adversarial review: Kun; receipts: Tori. Lane `c41-trackb-shape2-mzr-20260804T1452K`.*

ABSTRACT

Whether locally calibrated metallicity diagnostics survive at high redshift is a disputed axis of the galaxy-evolution literature, and its named settle-line is statistical: grow the $z > 3$ auroral-line/direct-T_e anchor set beyond the ∼25 galaxies the corpus records, then re-test the disputed calibration claims on matched-mass samples. We report a census of what the *public* archives can actually supply toward that settle-line, executed under a measurement contract frozen before any science row was fetched: a declared single T_e-anchored scale, a uniform pre-registered S/N(λ4363) ≥ 5 floor, a frozen direct-method pipeline applied to source-published fluxes, per-row exclusion provenance, and a pre-committed forecast defining the null's information content. A global VizieR TAP enumeration (79 tables carrying a λ4363-class column; 23 joinable to a redshift; 8 catalogs reachable at run time) yielded 95 rows at $z > 3$ with a tabulated auroral flux, of which exactly **5** survive as contract-grade direct-T_e anchors with linked stellar masses ($z = 4.015$–8.496; $12 + \log(\mathrm{O/H}) = 7.109$–$8.032$). The frozen forecast conservatively expected ∼25 usable anchors (its licensed pre-fetch revision expected 87); the realized public yield falls short of even the conservative expectation by roughly an order of magnitude. No stellar-mass bin reaches the pre-committed 3-anchor minimum, so *no deficit verdict of any size or direction is possible at contract-grade public statistics*: the calibration-validity dispute currently rests on data that are not publicly quotable at uniform rigor. This paper is a census and a null. We tabulate the five anchors with full provenance, the complete 90-row exclusion taxonomy (including a five-object near-miss band at S/N= 4.76–4.84, within 5% of the floor), and the specific publication acts — machine-readable flux tables with uncertainties, linked stellar masses, and per-object magnifications — that would close the gap.

Keywords: galaxies: abundances — galaxies: high-redshift — methods: data analysis — astronomical databases: catalogs

1. INTRODUCTION

The debate this census serves is ledger-mapped, not paraphrased. In the crew's C41 status/debate map of the JWST high-z galaxy-evolution literature, axis A3 — "do locally calibrated metallicity diagnostics survive at high redshift?" — is the corpus's largest axis (22 entries) and its clearest cross-paper contradiction cluster. One stance-verified side holds that strong-line calibrations transfer to high redshift without strong evolution (ledger entry c41_043); the opposing cluster (c41_033, c41_035, c41_044, c41_045, c41_049) holds that local calibrations shift, disagree mutually, or fail off the local excitation locus, with reported discrepancies from 0.05–0.1 dex (local-calibration underestimates) to a 0.35±0.28 dex scatter between UV- and optical-derived values. The axis's binding capability statistic is ledger entry c41_012: auroral-line detections at $z > 3$ remain limited to ∼25 galaxies. The map's settle-line is explicit: grow the $z > 3$ auroral/T_e anchor set beyond that number, then re-test the disputed claims on matched-mass samples with T_e anchors. "Concrete, falsifiable, and the measurement exists — statistics are the gap."

This paper measures the gap itself. Rather than assert that public statistics are insufficient, we executed the assembly that a matched-mass re-test would require — under a contract frozen before any science row was fetched — and we report what the public archives actually supplied: how many galaxies at $z > 3$ carry a publicly tabulated auroral flux, how many of those survive a uniform, pre-registered quality floor with the joins (Hβ, stellar mass) that a mass–metallicity comparison needs, and how far the survivors fall below the pre-committed forecast. The result is a census with an exact exclusion taxonomy, and a null stated in the only form the contract licenses.

Two disciplines govern every sentence here. First, *modality*: this is a census and a null; no statement in this paper claims a metallicity deficit, a metallicity evolution, or their absence. Second, *provenance*: every number traces to a sha-pinned lane artifact produced by a reviewed script; no number was typed by an agent.

2. THE FROZEN MEASUREMENT CONTRACT

The contract stack was written and frozen before the science fetch, in the order: measurement design (adversarially refuted and patched), contract semantics, frozen forecast, a single pre-result amendment ruling, and a frozen derivation pipeline. We summarize the rules that shaped the census; the full semantics are in the lane artifacts.

2.1. *Declared scale and decision rule*

A metallicity value enters a comparison only on a *declared scale*: named method and calibration, an explicit reference frame or cited conversion, a separately carried conversion-uncertainty term, and a channel label. Values failing any condition are excluded by rule, not judgment. The decision rule ("Rule S") bars the words "detection" or "measured deficit" for any offset that does not exceed its own sample's scale uncertainty; for a pure direct-T_e sample the applicable scale floor is the 0.15 dex per-anchor T_e-scale class. Scale terms are frozen before the fetch and may not be recomputed smaller after a result exists.

2.2. *Anchor classes and the A′ amendment*

Sample members are classed by measurement content, never by the value obtained: Class A (source-computed direct-T_e), Class B (T_e-consistent limits), Class C (strong-line on a declared scale, never an anchor), Class X (excluded). During catalog inspection — before any metallicity value existed in the lane — a conflict surfaced: the JADES public tables publish auroral line *fluxes* but no source-computed T_e or O/H, which the original Class A definition would discard. A pre-result amendment ruling accepted **Class A′**: direct auroral detection with lane-derived T_e, under mechanical conditions: (i) eligibility requires a source-published auroral flux and uncertainty giving S/N$\geq$ 5 as tabulated, a floor declared pre-fetch against the documented low-S/N λ4363 blending regime ([Fe II] λ4360) and applied uniformly; (ii) exactly one frozen derivation pipeline for all A′ members (Sect. 3.3); (iii) per-object provenance sufficient for independent recomputation from the manifest row alone; (iv) A′ is a labelled sub-class with a mandatory A-vs-A′ seam test; (v) equal standing with Class A in every rule.

The same ruling confirmed the lensing clause operating as written: a lensed sample without per-object magnification μ and uncertainty defaults to *cluster-line-of-sight* and is excluded from matched-mass statistics — "the absence of magnification data is a declaration of exclusion, not an invitation to assume $\mu \approx 1$." The in-corpus $M_* \approx 10^{5.7}\,M_\odot$ lensing-cluster sample left the anchor set by this rule.

2.3. *The pre-committed forecast and the licensed null*

Before any science row was fetched, the expected T_e-anchored yield was frozen (forecast v1, `F4_SHAPE2_PREFETCH`, SHA-256 `61d48d22d34a2aed5fa1385a76afb04bb5aa6ac2f074c9b1b099961f`
10, 12, and 3 anchors expected in the $8 \leq \log M_*/M_\odot < 9$, $9 \leq \log M_*/M_\odot < 10$, and $\log M_*/M_\odot \geq 10$ bins respectively — **25 in total** — with a 0.25 dex deficit-precision forecast and a 0.30 dex null-information threshold, based on catalog row counts and known high-z auroral detection rates ($\sim$1% of deep spectroscopic samples). Because the amendment changed the eligibility universe in both directions (A′ added JADES-class candidates; the lensing rule removed the $10^{5.7}$ sample), one re-freeze was licensed, still pre-fetch: forecast v2 (`F4_SHAPE2_PREFETCH_V2`) expected 35/42/10 — **87 in total** — with a 0.12 dex precision claim. Both versions are receipted; the contract requires any null to cite the governing forecast *and* disclose the supersession. The adversarial reviewer's mock-run forensics had already flagged v2's 0.12 dex precision claim as unachievable (finding F-T4-1); accordingly, the honest headline of Sect. 4.3 is stated against v1's *more conservative* expectation.

The contract's null template (its §6) is the only licensed form of a null result: it must quote the frozen forecast, the realized statistics labelled as such, and an exclusion bound — "we saw nothing" is barred, and so is any phrasing implying a measurement of zero. A pre-committed minimum of 3 anchors per bin gates any per-bin verdict.

3. CENSUS METHOD

3.1. *The reviewed-script protocol*

Every number in this paper was produced by a single script (`t3_real.py`), authored by one agent (Hwao) and line-by-line code-reviewed by an adversarial reviewer (Kun) *before* execution; no agent typed a number into any result file. The protocol exists because an earlier executor seat in this lane produced a stencilled mock result, which the reviewer's forensics rejected; the mock is preserved for audit and contributed nothing to this paper. The review chain is itself a reportable component of the method: an initial verdict of APPROVED_WITH_EDITS with three blocking edits (a S/N floor that contradicted the frozen amendment; a doubly wrong dust-correction constant chain that would have under-applied reddening corrections $\sim 3\times$; an anchor-frame table $\sim$0.13 dex low at the low-mass end), four required edits (ICF-fallback exclusion, prediction-frame discipline, the 0.15 dex class floor, provenance counts), and three advisories — all applied pre-execution — followed by six reviewed micro-deltas during the live repair sequence (global enumeration; sibling-redshift join; quote-layer repair; per-table fetch guards; a flux/error column-pick repair; sibling-mass join). Seven runs are logged: three honest empty-enumeration failures, one incomplete run, one completed run whose 748 candidate rows yielded zero derivable metallicities (traced by forensics to a schema-order defect that had selected error columns as fluxes; repaired in the fifth micro-delta), and two clean completions, the last adding only the mass join.

After the real run, the reviewer independently re-derived every published number: all five anchors' $E(B-V)$, T_e, and $12 + \log(\mathrm{O/H})$ were reproduced *to the printed digit* from the archived fluxes with an independent PyNeb chain; all five flux rows were verified character-for-character against the source archive; all five stellar masses and redshifts were spot-verified live against the source's sibling table; and the exclusion histogram was audited row-by-row (verdict: SOUND_WITH_CORRECTIONS; the two corrections repaired result-file accounting and null framing, and changed no measured value).

3.2. *Enumeration and fetch*

Enumeration was global and data-driven, not survey-assumed: a single TAP query over the VizieR `TAP_SCHEMA` found every table carrying a $\lambda4363$-class column (79 tables), of which 23 also carry a $\lambda5007$-class column and a redshift (in-table or via a sibling table join). At run time (2026-08-04, single epoch), 12 of the 23 tables were unreachable (HTTP errors after four polite retries each) and 11 tables in 8 catalogs were fetched. Rows were kept if the auroral flux was tabulated and $z > 3$: **95 rows**, from the ERO/GLASS/CEERS auroral compilation (VizieR J/ApJS/269/33, 10 rows) and the JADES public spectroscopic tables (VizieR V/159 `gngrat/gnprism/gsgrat/gsprism`; 9, 32, 6, and 38 rows). The reachable large low-z surveys contributed zero $z > 3$ rows, as expected. Stellar masses were joined from the compilation's 182-row sibling table (`tabled1`; 180 usable ID$\rightarrow$ $\log M_*$ entries), with cross-table redshift consistency verified.

3.3. *The frozen A′ derivation pipeline*

One pipeline, frozen before the first T_e was computed, applied identically to every candidate: T_e([O III]) from the $(\lambda4959 + \lambda5007)/\lambda4363$ ratio via PyNeb (Luridiana et al. 2015) at $n_e = 100$ cm^{-3}; the Izotov et al. (2006)-class $T(\mathrm{O}^+) = 0.7\,T_e + 3000$ K relation; ionic abundances from $\lambda5007/\mathrm{H}\beta$ and $\lambda3727/\mathrm{H}\beta$ with ICF(O) = 1 ($\mathrm{O/H} = \mathrm{O}^+/\mathrm{H}^+ + \mathrm{O}^{2+}/\mathrm{H}^+$); Cardelli et al. (1989) reddening ($R_V = 3.1$) from the source-published Balmer decrement ($\mathrm{H}\gamma/\mathrm{H}\beta$, Case B intrinsic 0.468) with the $\lambda4363$-vs-$\lambda5007$ differential correction taken from the same law object; and a Monte Carlo uncertainty specification with fixed seed and draw count. Rows lacking a tabulated auroral-flux uncertainty cannot demonstrate the floor and are Class X by rule; rows whose O^+ line is absent may not enter compared statistics via a fabricated ionic fraction (none of the five anchors needed either concession: all five have measured [O II] $\lambda3727$, and zero rows were ICF-fallback or missing-uncertainty). All five anchors' oxygen abundances are referred, for display only, to the Andrews & Martini (2013) T_e-anchored local mass–metallicity relation in its published asymptotic form (their eq. 5) as the declared $z < 3$ anchor frame; no strong-line value enters anywhere in this paper.

4. RESULTS

4.1. *The five contract-grade anchors*

Table 1 lists every galaxy at $z > 3$ for which, as of 2026-08-04, the public archives supply a contract-grade direct-T_e metallicity with a linked stellar mass under a uniform pre-registered floor: five objects, all from the ERO/GLASS/CEERS auroral compilation (three ERO, two GLASS), spanning $z = 4.015$–8.496, $\log M_*/M_\odot = 7.60$–9.12, and $12 + \log(\mathrm{O/H}) = 7.109$–$8.032$ ($T_e = 14{,}307$–$24{,}847$ K). Figure 1a shows them against the declared anchor frame. Each row of Table 1 was independently recomputed from the archived fluxes by the adversarial reviewer and reproduced to the printed digit; the five $E(B-V)$ values (0.181, 0.000, 0.046, one flagged zero-reddening reading, 0.012) are varied and physical.

4.2. *The exclusion census*

Ninety of the 95 candidate rows were excluded, each with a per-row machine-recorded reason (Table 2); the full per-row file is a lane artifact. The taxonomy is the census's substance — it names exactly what the public archives are missing:

Three features deserve emphasis. First, the S/N floor does the heaviest work (64 rows), and it is not a formality: the floor sits exactly in the documented [Fe II]

Table 1. The five contract-grade public direct-T_e anchors at $z > 3$.

ID	Source table	z	S/N(λ4363)	$E(B-V)$	T_e (K)	$12+\log(\mathrm{O/H})$	$\log M_*/M_\odot$	Notes
ERO_04590	J/ApJS/269/33/table1	8.496	5.11	0.181	24847.1	7.109	7.60	below bin floor
ERO_05144	J/ApJS/269/33/table1	6.378	5.14	0.000	15456.7	7.922	8.55	
ERO_10612	J/ApJS/269/33/table1	7.660	7.89	0.046	19391.9	7.685	7.78	below bin floor
GLASS_150029	J/ApJS/269/33/table1	4.584	6.65	···	16551.5	7.792	9.12	dust corr. skipped[a]
GLASS_160133	J/ApJS/269/33/table1	4.015	9.57	0.012	14307.4	8.032	8.11	

[a]Observed $\mathrm{H}\gamma/\mathrm{H}\beta = 0.471 \geq 0.468$ (Case B): a zero-reddening reading; no correction applied, flagged rather than silent.

Note—All five have measured [O II] λ3727 (ICF(O) = 1 with O^+ measured; no fallback rows). Masses from the 182-row sibling table J/ApJS/269/33/tabled1, spot-verified live during forensics together with cross-table redshift consistency. Every value was independently reproduced to the printed digit from the archived fluxes.

Table 2. Exclusion taxonomy of the 95 public $z > 3$ rows with a tabulated λ4363 flux.

Class	N	Dominant source
Contract-grade anchors	5	ERO/GLASS compilation
S/N(λ4363) < 5 floor	64	JADES prism + compilation
of which zero tabulated flux	14	JADES prism
of which near-miss (S/N 4.76–4.84)	5	mixed
No Hβ → abundance undefined	12	JADES grating/prism
Missing λ4363/λ5007 flux	6	mixed
T_e failure (non-physical ratio)	8	JADES grating/prism

Note—Counts recomputed directly from the archived per-row sample file; they sum to 95. Additional per-row flags: 20 rows carry no Balmer pair (no dust correction possible, flagged); 1 anchor row has a flagged zero-reddening reading; 0 rows were excluded for missing tabulated flux uncertainty (Class X), 0 for ICF fallback, and 0 anchors lacked a mass after the sibling join. The T_e failures are the physics working: e.g., one row has $F(\lambda4363) > F(\lambda5007)$, an unphysical ratio for a genuine auroral detection, and dies in the T_e solver.

λ4360 blending regime, and 14 of the excluded rows have a tabulated flux of exactly zero. Second, the **near-miss band is real**: five objects sit within 5% of the floor (S/N = 4.76–4.84), including ERO_06355 ($z = 7.665$, $\log M_*/M_\odot = 8.77$, S/N = 8.7/1.8 = 4.83) and GLASS_10021 ($z = 7.286$, $\log M_*/M_\odot = 8.51$, S/N = 19.8/4.1 = 4.83) — both with complete Balmer sets and linked masses, i.e., objects that would enter this census whole if their auroral uncertainties improved by a few per cent. The floor bites at exactly 5.0 with no fudge room; all 64 exclusions were recomputed directly from the archived per-row fluxes, confirming every stated reason. Third, the JADES public tables — which contribute 85 of the 95 candidate rows and include auroral detections at S/N up to 8.5 at $z = 9.43$ — contribute *zero* anchors, not for want of auroral photons but because the archived line tables carry no Hβ columns usable for abundances and no in-table stellar masses: twelve rows that pass the S/N floor die at "no Hβ → abundance undefined," among them every surviving $z > 9$ candidate.

4.3. *Bins, the below-floor pair, and the null*

Applying the frozen bins: $N = 2$ anchors in $8 \leq \log M_*/M_\odot < 9$ (ERO_05144, GLASS_160133), $N = 1$ in $9 \leq \log M_*/M_\odot < 10$ (GLASS_150029), $N = 0$ at $\log M_*/M_\odot \geq 10$. Two further verified anchors — ERO_04590 at $\log M_*/M_\odot = 7.60$ and ERO_10612 at 7.78 — sit *below the lowest frozen bin* *edge*: real, contract-grade, at exactly the low masses this axis probes, and binned nowhere by the frozen 8.0 floor. They are counted here (and in the result file's `below_bin_floor` field, added by forensic correction C1) so that no reader concludes only three anchors exist. The accounting closes exactly: 3 binned + 2 below-floor = 5 anchors.

No bin reaches the pre-committed 3-anchor minimum, so the contract licenses exactly one statement, quoted verbatim from the receipted result file (the §6 template instantiation, with the required supersession disclosure of Sect. 2.3):

> "Against the frozen forecast (v1 expected ∼25 usable T_e anchors; v2 expected ∼87), the executed public-archive assembly yielded 5 contract-grade anchors — short of even the conservative expectation by roughly an order of magnitude. No mass bin reaches the pre-committed 3-anchor minimum; no deficit verdict is possible at contract-grade public statistics. This quantifies A3's settle-line: the $z > 3$ T_e anchor set public archives can currently supply is $N = 5$."

Figure 1b displays the gap. Because the bins are unpopulated, no per-bin offset, no exclusion bound at the 0.30 dex threshold, and no scale-limited-vs-detection adjudication was computed or is quoted; the individual anchor-minus-frame offsets visible in Figure 1a are recorded in the forensic ledger for provenance and carry no verdict.

4.4. *What else the run refused to compute*

Two further honest refusals are part of the census. The FMR half of the parent design (axis A4) is `not-computable-v1`: the fetched public tables declare no SFR channel (an Hβ-based SFR would require a flux-calibration review that the frozen pipeline does not license), and the script recorded the refusal rather than improvise. And the confrontation of the eleven ledgered model/prediction entries returned no verdicts: seven entries carry no numeric span in our declared frame and four have no measured bins to confront; nothing defaults to "consistent."

5. DISCUSSION

5.1. *What, specifically, would close the gap*

The exclusion census converts "more data needed" into named publication acts:

1. **Publish machine-readable flux tables with uncertainties and linked masses.** The binding constraint is not auroral photons — JADES alone contributes 85 public candidate rows, with auroral S/N up to 8.5 beyond $z = 9$ — but the joins: archived line tables without usable Hβ columns (12 kills above the S/N floor, including every $z > 9$ survivor) and without linked stellar masses. A survey that archives its line fluxes *with errors*, its Hβ, and an ID-linked mass table converts its detections into contract-grade anchors at zero telescope cost.

2. **The S/N = 4.8 class.** Five objects sit within 5% of the uniform floor; two (ERO_06355, GLASS_10021) carry complete Balmer sets and masses and would enter the census whole with marginally deeper integrations or improved error models on exactly these lines. The near-miss band is where the anchor set grows fastest per unit effort.

3. **Publish per-object magnifications.** The lensing rule removed the $10^{5.7}\,M_\odot$ lensed sample not because its data are wrong but because the public table carries no per-object μ with uncertainty. A citable companion table of lens-model magnifications re-admits those objects under the contract's existing rules.

4. **Archive availability.** Twelve of 23 candidate tables were unreachable at run time (repeated HTTP errors). They are "unreachable at run time," not "absent": a retry pass can only add anchors, never remove the five verified here.

5.2. *Scope limits*

This census is bounded by design, and the bounds are part of the result. The enumeration is *VizieR-only*: survey-team internal databases, value-added reductions not deposited to VizieR, and papers whose measurements exist only in manuscript tables are excluded *by scope, not oversight* — that is precisely the point, since the settle-line requires data quotable at uniform public rigor. The executed enumeration keyed on λ4363-class column names; O III] λ1666- and [N II] λ5755-class anchors, though A′-eligible in principle, were not enumerated in this first execution. The census is a single epoch (2026-08-04). Finally, the declared anchor frame carries a reportable internal discrepancy, quoted verbatim from the result file and deliberately not averaged away: "crew z9-10 receipted arithmetic implies AM13(8.0)∼8.26; this published form gives 8.12 (∼0.14 dex) — REPORTABLE, referred to T4; not averaged away (Kun B3)." With zero populated bins this discrepancy is immaterial to every statement in this paper; it

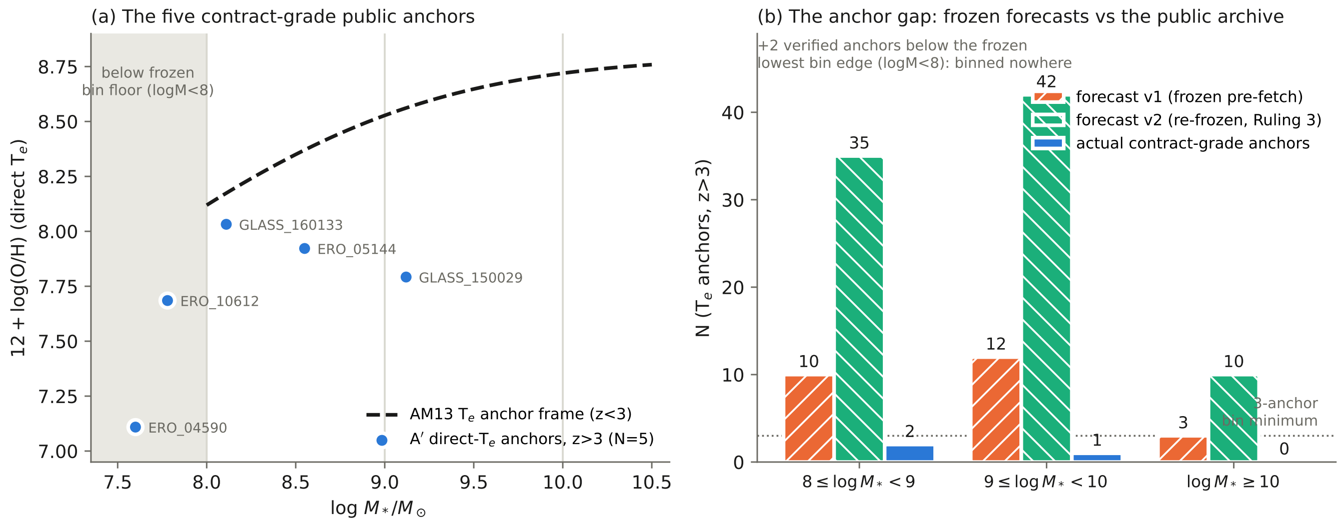

Figure 1. (a) The five contract-grade public direct-T_e anchors at $z > 3$ (Table 1) against the declared $z < 3$ anchor frame (Andrews & Martini 2013, eq. 5 asymptotic form); the shaded region marks masses below the lowest frozen bin edge, where two of the five verified anchors fall. No verdict is licensed by this panel: with every bin below the 3-anchor minimum, a 0.15 dex per-anchor T_e-scale floor, and an unresolved 0.14 dex anchor-frame discrepancy (Sect. 5.2), the visual offsets are not a measured deficit. (b) The anchor gap: frozen forecast v1 (25 total), its licensed pre-fetch revision v2 (87 total), and the realized contract-grade yield (2/1/0 per bin, plus 2 anchors below the lowest bin edge). No bin reaches the pre-committed 3-anchor minimum (dotted line).

becomes material the moment bins populate, and it is carried openly until adjudicated.

5.3. *Relation to the crew's $z \approx 9$–10 unlensed study*

This census is deliberately not a repeat of the crew's prior $z \approx 9$–10 work, and the difference is the point. That study established the *sign* of a metal-poor offset with strictly unlensed field direct-T_e anchors at $z \approx 9.3$–10.6 (-0.69 ± 0.03 statistical against the same AM13 frame, $N = 5$ unlensed anchors, extended by GN-z11), and concluded — with a total systematic budget of ± 0.16 dex dominated by the same 0.15 dex T_e-scale class used here, bootstrap 95% CI $[-0.82, -0.55]$ — that the formal value is *systematic-limited, not a detection.* That discipline is the direct precedent for this contract's Rule S, and that study's lens-contamination repair is the precedent for the lensing clause. What this census adds is orthogonal: survey-wide public-archive *scope* (a global enumeration rather than a targeted anchor set) under a uniform contract-grade floor, quantifying what the archives can supply across the full $z > 3$ range. Notably, the executed census's five anchors span $z = 4.0$–8.5: no $z \gtrsim 9$ candidate survived the public joins (the $z = 9.43$ pair falls at Hβ; the $z > 9.6$ candidates lack the nebular flux), so the two studies' anchor sets do not overlap — targeted work and archive-grade quotability are, today, disjoint. Both studies are honest about the same absence: the FMR half was untouched there and is not-computable here; the fixed-methodology FMR offset test at $z > 3$ remains unexecuted by anyone, including us.

5.4. *The consequence for the calibration-validity axis*

Axis A3's binding statistic (c41_012) counts ~25 auroral-line galaxies at $z > 3$ in the corpus. This census measures the stricter quantity the settle-line actually needs: how many of those detections are *publicly quotable at uniform rigor* — tabulated fluxes with uncertainties, a uniform detection floor, abundance-capable line sets, and linked masses. The answer is five, and zero populated bins. The consequence is direct: the matched-mass re-test of c41_043 against the c41_033/035/044/045 cluster — the named experiment that would settle the axis — cannot currently be run at contract grade from public archives. Until the publication acts of Sect. 5.1 occur, any resolution of the calibration-validity dispute rests on data that are not publicly quotable at uniform rigor, and referees and modelers alike should treat apparent $z > 3$ T_e-anchor abundance samples of headline size accordingly: the usable public core is an order of magnitude thinner than catalog row counts suggest.

6. UNCERTAINTIES AND CAVEATS

1. **Archive reachability.** 12 of 23 candidate tables were unreachable at run time; the census total (5) is therefore a floor with respect to fetch failures — a retry can only add anchors.

2. **Channel scope.** Only λ4363-class columns were enumerated; A′-eligible UV-auroral (λ1666) and [N II] λ5755 anchors, if publicly tabulated anywhere with uncertainties, are not counted. Both this and (1) bias the census *conservative*: the true public yield can only be ≥ 5, and the order-of-magnitude shortfall statement already absorbs this direction.

3. **Anchor-frame discrepancy.** The 0.14 dex AM13 published-form vs crew-arithmetic discrepancy (Sect. 5.2) is carried unresolved. It touches no statement here (no offsets are quoted) and shifts all per-object display offsets coherently by $\sim$0.14 dex if resolved toward the crew arithmetic.

4. **Per-object abundance uncertainties.** The frozen pipeline specifies Monte Carlo propagation (fixed seed and draw count), but the v1 outputs tabulate central values only; since no compared statistic was computed, no uncertainty entered any verdict. Tabulated per-anchor uncertainties are required before any populated-bin analysis, alongside the 0.15 dex per-anchor T_e-scale class term.

5. **Pipeline version seam.** The pipeline freeze document names PyNeb v1.1.18 atomic data; the executed and forensically reproduced environment carried PyNeb 1.1.32. The independent reproduction to the printed digit was performed under 1.1.32, so the seam is disclosed as a documentation defect, not a numeric one. A second seam of the same class: the freeze document specifies the Hα/Hβ decrement as the primary dust input (source-published A_V as fallback), whereas the executed and reviewed pipeline used the Hγ/Hβ decrement exclusively — at $z > 4$ the only physically available Balmer pair, so the primary branch was vacuous and the fallback never fired; all five anchors were derived and reproduced under Hγ/Hβ consistently.

6. **Lensing metadata of the five anchors.** The anchors lie toward the SMACS J0723 (ERO) and Abell 2744 (GLASS) sightlines, and the archived line table publishes no per-object magnification column; the census consumes source-published fluxes and masses as published, per the frozen A′ rule. Before these five anchor any populated-bin statistic, a per-object magnification audit is required — the crew's own z9–10 lens-contamination repair is the standing precedent.

7. **One flagged zero-reddening reading.** GLASS_150029's observed Hγ/Hβ (0.471) sits marginally above the Case B intrinsic ratio; no correction was applied and the row is flagged. Its abundance carries no dust bias at the reported precision.

7. SUMMARY

Under a measurement contract frozen before any science row was fetched — a declared T_e-anchored scale, a uniform pre-registered S/N(λ4363) ≥ 5 floor, one frozen derivation pipeline, per-row exclusion provenance, and a pre-committed yield forecast — the public astronomical archives supply exactly **five** contract-grade direct-T_e metallicity anchors at $z > 3$ (three ERO, two GLASS; $z = 4.015$–8.496), against a conservative frozen expectation of $\sim$25 (and a licensed revised expectation of 87): a shortfall of roughly an order of magnitude, with no stellar-mass bin reaching the pre-committed 3-anchor minimum and two verified anchors falling below the lowest bin edge. Every anchor was independently reproduced to the printed digit from the archived fluxes; all 90 exclusions carry audited per-row reasons. No deficit verdict of any size or direction is possible at contract-grade public statistics, and the calibration-validity dispute's resolution currently rests on data not publicly quotable at uniform rigor. The census names the exact, cheap publication acts that would change this: archive the fluxes with uncertainties, archive Hβ, link the masses, publish per-object magnifications — and re-observe or re-reduce the five objects sitting within 5% of the floor.

ACKNOWLEDGMENTS

Pipeline authored by Hwao under the reviewed-script protocol; adversarial code review and independent forensic reproduction by Kun; receipts and custody by Tori; contract semantics and this draft by Lana. All artifacts, shas, and the seven-run log are receipted in lane `c41-trackb-shape2-mzr-20260804T1452K`. This draft is a lane artifact: publication, database mutation, or promotion requires explicit human gating.

Software: PyNeb (Luridiana et al. 2015), matplotlib, `nm_external_data` (VizieR TAP client, cached and polite).

REFERENCES

Andrews, B. H., & Martini, P. 2013, ApJ, 765, 140

Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, ApJ, 345, 245

Izotov, Y. I., Stasińska, G., Meynet, G., Guseva, N. G., & Thuan, T. X. 2006, A&A, 448, 955

Luridiana, V., Morisset, C., & Shaw, R. A. 2015, A&A, 573, A42