Identifiable or not? A Bayesian reanalysis of the Uluburun tin cargo

iddin annakam — "he delivered the tin"

Conor Reid*

August 1, 2026 — working draft

Abstract

The provenance of the Uluburun shipwreck's 105-ingot tin cargo is contested between a Central Asia + Taurus attribution (Powell et al., 2022) and a European counter-reading (Berger et al., 2023b; Williams et al., 2025), both argued from visual comparison of isotope biplots. We ask the prior question — *is deposit-level attribution statistically identifiable from these data at all?* — for the Sn-isotope channel. From a harmonized 11-source cassiterite library we fit per-ingot and batch-level Bayesian mixtures and subject them to posterior predictive checks, fractionation-prior sensitivity, and source-library perturbation. Per-ingot classification fails all three diagnostics: its attributions track whichever source has the smallest variance. A batch-level model survives them and supports exclusions only, with graded strength: Mushiston lies 6–7σ from the 86-ingot cargo bulk and its exclusion as sole batch source withstands fractionation-prior sensitivity, per-batch fractionation, and doubled reference dispersions, whereas Spain's $\approx 3\sigma$ exclusion remains fractionation-conditional, and the five positive-$\delta^{124/116}$Sn sources (Cornwall, Erzgebirge, Taurus, Serbia, Portugal) remain mutually unidentifiable. Negative $\delta^{124/116}$Sn does not imply Central Asia: the tightest negative-$\delta^{124/116}$Sn signature in the library is Iberian — Penouta, a single Galician mine (-0.285 ± 0.075‰, n=6), tighter than Mushiston itself. A concentration sweep over pre-registered trade-route priors shows the ten batch means carry only $N_{\text{eff}} \approx 3$ pseudo-batches of information about the positive apportionment, while every exclusion survives all five prior families: the apportionment is prior-driven; the exclusions are not. One isotopic dimension bounds admixture only weakly. The Pb, trace-element, and comparanda channels remain uncomputable here, which is not unidentifiable in principle. Three structural gaps cap all analyses: figure-only Taurus values, no Karnab ore data, and a ∼857-analysis compilation never deposited as a dataset.

1 Introduction

The Late Bronze Age ship that sank at Uluburun (∼1320 BC) carried about a ton of tin in 105 ingots — the single largest tin assemblage of the Bronze Age Mediterranean (Pulak, 1998). Where that tin was mined has become the sharpest open question in the archaeology of the period's metal trade, and since 2022 it has been the subject of an unusually direct dispute between two research groups. Powell et al. (2022), analyzing 90% of the cargo, attributed one third of the ingots to Central Asian ores (Uzbekistan and Tajikistan, with the Mušiston deposit

*Draft prepared from the *iddin-annakam* analysis repository (Reid, 2026). Contact: conor.reid@hey.com.

central) and two thirds to the Taurus Mountains of Anatolia. Berger et al. (2023b) replied that Mushiston is excluded on chemical and isotopic grounds and that south-west Britain is the likelier source for much of the eastern Mediterranean tin trade, a position elaborated for shipwreck cargoes by Williams et al. (2025). The exchange continues through the two camps' subsequent papers (Powell et al., 2024b,a, 2025, 2026).

Both camps argue from the same kind of evidence — tin and lead isotope ratios and trace-element chemistry, compared visually against ore reference fields — and reach incompatible conclusions. This paper asks the question that sits underneath that impasse: *are the competing attributions statistically identifiable from the data that exist?* That question has a quantitative answer, and obtaining it requires no new measurements: it requires treating the sources as the overlapping statistical distributions they are (Baxter, 1999; Baxter et al., 2000), treating the contested smelting-fractionation correction (Berger et al., 2018) as an uncertainty rather than a constant, and asking a fitted model to survive its own diagnostics before its attributions are believed. Bayesian mixture approaches of this kind are established in adjacent provenance problems (Stock et al., 2018; Mason et al., 2020; Tomczyk, 2025) — indeed Mason et al. (2020) is a probabilistic Sn-isotope apportionment method from within one of the disputing groups. What has not been published, to our knowledge, is an *identifiability* analysis for this cargo: not "what does a model attribute" but "can any model built on these data attribute at all, and under which assumptions does its answer flip."

Our contributions: (i) a harmonized, license-tracked cassiterite $\delta^{124/116}$Sn reference library for eleven real sources, with two documented corrections to how "Erzgebirge" data has been aggregated (§2); (ii) a model-criticism arc — posterior predictive dispersion checks, fractionation-prior sensitivity, and source-library perturbation — showing that per-ingot classification fails three independent diagnostics, with a mechanistic account of *why* its failures produce confident-looking attributions (§4); (iii) a batch-level re-specification that survives those diagnostics and delimits what the isotope data genuinely support (§5); (iv) a pre-registered trade-route-prior sensitivity analysis quantifying the data/prior division of labor: the exclusions are prior-invariant, the positive apportionment tracks the prior nearly one-to-one (§6); and (v) documentation of three structural data gaps — figure-only Taurus values, absent Karnab data, and a ∼857-analysis reference database cited across the dispute but never deposited — that currently cap what any statistical treatment can decide (§7).

2 Data

2.1 Cargo and reference data

Ingot data are Powell et al. (2022)'s published table for the 105 Uluburun ingots ($\delta^{124/116}$Sn vs. NIST 3161a, Pb isotope ratios, trace elements, and the paper's own group labels MC1A–C, MC2A/B, P1, P2A/B, plus two ungrouped ingots). Reference (ore) data are drawn from four open datasets (Table 1), all reported on or harmonized to the same $\delta^{124/116}$Sn NIST 3161a scale. Harmonization is not a formality in this literature: two of this project's own early errors — a mislabeled $\delta^{124/116}$Sn column in one supplement, and a 4-amu ($\delta^{124/120}$Sn) value briefly compared against 8-amu values — were each caught only by cross-checking a source's parameters against an independent dataset for the same region, and both are documented in the repository (Reid, 2026). We recommend that practice as standard.

Table 1: Data sources. All were re-parsed from the published files; parser fidelity checks and per-source licenses are documented in the repository.

Dataset	Used here for	License
Powell et al. (2022) supplement	105 Uluburun ingots: $\delta^{124/116}$Sn, Pb, trace elements, group labels	CC BY-NC 4.0
Mathur et al. (2025) Tables 1–3	Cassiterite reference: Cornwall region cross-check, Erzgebirge, Slavkov, Serbia, Spain, Portugal	CC BY
Berger et al. (2023a) Suppl. S2	Central Asian cassiterite: Mushiston (28), Dara-i-Pech (22), Paprok (10), Kunar (9); dual NIST/Puratronic panels	CC BY
Williams et al. (2025) OSM	Cornwall/Devon paired Sn+Pb ore data (17); Erzgebirge ore Pb	CC BY
Powell et al. (2024b) Fig. 6	Taurus cassiterite (8 points, digitized; see §7)	—
Westner et al. (2025)	Mushiston ore Pb (primary table: Berger et al., 2022)	open (cite DOI)

2.2 The source library, with two corrections

Table 2 and Figure 2 give the eleven-source library; Figure 1 places it on the map. Two aggregation errors in our own earlier build are worth reporting because they are easy to make and materially move parameters. First, filtering Mathur et al. (2025) by a Czech country prefix *excluded* that compilation's seven German rows — Altenberg, Ehrenfriedersdorf, Sauberg, Sadisdorf, i.e. Saxon Erzgebirge proper — while *including* fourteen rows from Horní Slavkov and Krásno, which belong to the Slavkov Forest, a related but distinct Variscan district ∼40 km southwest whose mean (+0.15‰) sits half a permil below the Czech-side Erzgebirge sensu stricto (+0.63‰, n=25). The blended "Erzgebirge" (+0.46‰) was an average of two different distributions — precisely the single-Gaussian-per-source failure the statistical literature warns about (Baxter, 1999). We split them. Second, the corrected Erzgebirge library entry — Czech-side and German rows pooled (n=32, mean +0.584‰; the German side alone averages +0.41‰) — is nearly identical to Cornwall's (+0.632‰), which sharpens, rather than softens, the indistinguishability results below.

A third correction, found by auditing the library's aggregation levels, is of the same class: assigning Iberian rows by study-ID country prefix had placed a sample whose own locality reads "Portugal, Pyrite Belt, Neves Corvo" (−1.11‰) inside the *Spain* component, dragging its mean down and inflating its dispersion. Assigning by locality and excluding Neves Corvo entirely — its two samples sit 1.88‰ apart, split across both country entries, and represent massive-sulfide mineralization foreign to the vein/greisen deposits dominating both countries — gives Spain n=9 (-0.191 ± 0.228‰) and Portugal n=8 ($+0.251 \pm 0.439$‰). The audit behind the correction exposes a structural fact the library cannot escape: "Spain" is two-thirds one mine (Penouta, n=6) and "Portugal" is seven-eighths one mine (Panasqueira), so the library unavoidably mixes deposit-level entries (Mushiston, Paprok), district-level entries (Slavkov, West Serbia), and country-level entries (Spain, Portugal) whose dispersions are not commensurable — and it cannot be made granularity-consistent at deposit level with present sampling, since outside Penouta the per-district n is 1–2. We state the rule this forces: verdict strengths that run through component dispersion must be read against the entry's aggregation level (§8).

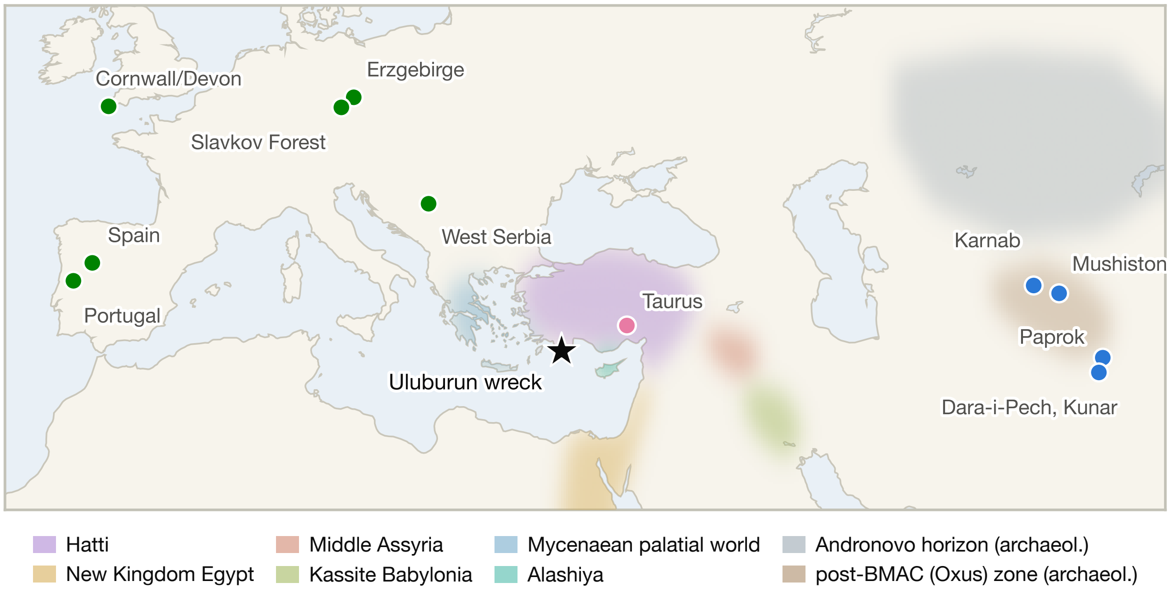

Figure 1: The candidate sources and the political landscape at the time of the wreck (c. 1320 BC). State territories (tints) are schematic reconstructions for the reign of Muršili II and the late 18th Dynasty; the Central Asian tin belts lie beyond state geography in the Andronovo steppe horizon and the post-BMAC (Oxus) zone — archaeological horizons, drawn hatched because they are a different kind of entity. Marker color is region family (as in Fig. 2). All extents and site locations are approximate; the Taurus marker stands jointly for Kestel, Bolkardağ, and Hisarcık; Dara-i-Pech and Kunar share a marker at this scale.

The library also records two facts that shape everything downstream. Penouta alone (-0.285 ± 0.075‰, n=6) is an isotopic doppelgänger for Mushiston (-0.270 ± 0.129‰, n=28) — tighter, even: *a negative* $\delta^{124/116}$Sn *does not imply Central Asia, and the tightest negative-*$\delta^{124/116}$Sn *signature in the library is Iberian.* And the three Afghan deposits from Berger et al. (2023a) span tight-and-negative to extremely wide (Dara-i-Pech sd 1.23‰), so "Central Asia" is not one signature but several.

3 Models and diagnostics

Per-ingot classifier (model A). Neither camp fit a mixture model; both, however, publish per-ingot attribution counts ("all 26 P1 ingots classify to…"-style statements and ingot-by-ingot source tables), and model A is the generative model under which such counts carry the evidential weight with which they are quoted. In that specific sense it is the charitable formalization of both camps' implicit reasoning: each ingot i is a draw from exactly one source,

$$y_i \sim \textstyle\sum_k \pi_k \mathcal{N}\big(\mu_k + \delta,\ \sigma_k^2\big), \qquad \pi \sim \text{Dirichlet}(\mathbf{1}), \qquad \mu_k \sim \mathcal{N}\big(\hat{\mu}_k,\ \hat{\sigma}_k^2/n_k\big),$$

with σ_k fixed at the specimen-level estimates of Table 2 and δ the ore→metal fractionation shift. The baseline prior $\delta \sim \mathcal{N}(0.14, 0.05)$ encodes Berger et al. (2018)'s complete-reduction range; the sensitivity prior is the mixture that reduction chemistry actually implies, $0.7\,\mathcal{N}(0.135, 0.03) + 0.3\,\mathcal{N}(0.50, 0.22)$, whose heavy component reaches the experimentally observed incomplete-reduction maximum (+0.88‰).

Table 2: Source parameters ($\delta^{124/116}$Sn, ‰, NIST 3161a scale; specimen-level statistics). Serbia's samples are Mason et al. (2020)'s, folded into Mathur et al. (2025) — the two must never be pooled. Taurus is digitized from a published figure (§7).

Source	Region	n	mean	sd
Cornwall/Devon (GB)	Europe	17	+0.632	0.420
Erzgebirge s.s. (DE/CZ)	Europe	32	+0.584	0.489
Taurus (TR; digitized)	Anatolia	8	+0.479	0.441
West Serbia (= Mason)	Europe	17	+0.415	0.217
Portugal	Europe	8	+0.251	0.439
Slavkov Forest (CZ)	Europe	14	+0.148	0.308
Paprok (AF)	C. Asia	10	−0.039	0.383
Spain	Europe	9	−0.191	0.228
Mushiston (TJ)	C. Asia	28	−0.270	0.129
Dara-i-Pech (AF)	C. Asia	22	−0.864	1.226
Kunar (AF)	C. Asia	9	−1.038	1.011

Batch-level model (model B). Smelted ingots are homogenized averages over kilograms of ore, cast in batches; the unit that samples a source's specimen distribution is the *batch*, not the ingot. With Powell's eight named groups (plus the two ungrouped ingots as singletons) standing in for batches g:

$$m_g \sim \textstyle\sum_k \pi_k \, \mathcal{N}\big(\mu_k + \delta,\ \sigma_k^2\big), \qquad y_i \sim \mathcal{N}\Big(m_{g(i)},\ \tau_{g(i)}^2\Big), \qquad \tau_g \sim \mathrm{HalfNormal}(0.2).$$

Using the full specimen sd for batch means is deliberately conservative for the exclusions — and anti-conservative for the mutual unidentifiability, which wider components make easier; the aggregation curve of §5.1 answers that asymmetry. Groups-as-batches is the conservative extreme of a bracket whose other end is model A; reality sits between.

Diagnostics. Posterior predictive checks on within-group dispersion; fractionation-prior sensitivity (both priors above, plus a fixed-δ profile over $[0, 0.9]$); simulation-based calibration of the inference machinery (1,000 iterations, no failures; rank-uniformity χ^2 $p = 0.19$–1.00 across parameters; Talts et al., 2018); source-library perturbation (3, 6, 11, then 12 sources — §5.2 — plus an explicit broad "none of the above" component, $\mathcal{N}(0.3, 1.5^2)$, not δ-shifted); and PSIS-LOO for the one model comparison it can address (the convex-mixing alternative; inconclusive by construction, as reported in the Limitations; Vehtari et al., 2017). All models were fit with NUTS (PyMC) on the default transformed parameterization (stick-breaking simplex for π, log-transformed scales); $\hat{R} \leq 1.01$ throughout unless noted. Residual divergences in the hierarchical batch fit (79 of 8,000 draws at target accept 0.995) left results stable to the third decimal across sampler settings; part of that residue is a (m_g, τ_g) funnel on the smallest batches; the naive non-centering $\tau_g \propto |z_g|$ was tried and trades the funnel for a reflection degeneracy NUTS cannot traverse ($\hat{R} = 1.9$), so hierarchical pooling of τ remains the principled cure — deferred as model development, and flagged where the funnel bites (the two singleton batches, whose τ_g is prior-dominated; Fig. 6 caption). The mixture-over-well-separated-components geometry is intrinsically divergence-prone, and prior families with extreme concentration (§6) used an adaptive retry: any fit showing pathology was rerun at stricter settings (target accept 0.999, doubled tuning) until converged, with per-fit settings recorded in the repository.

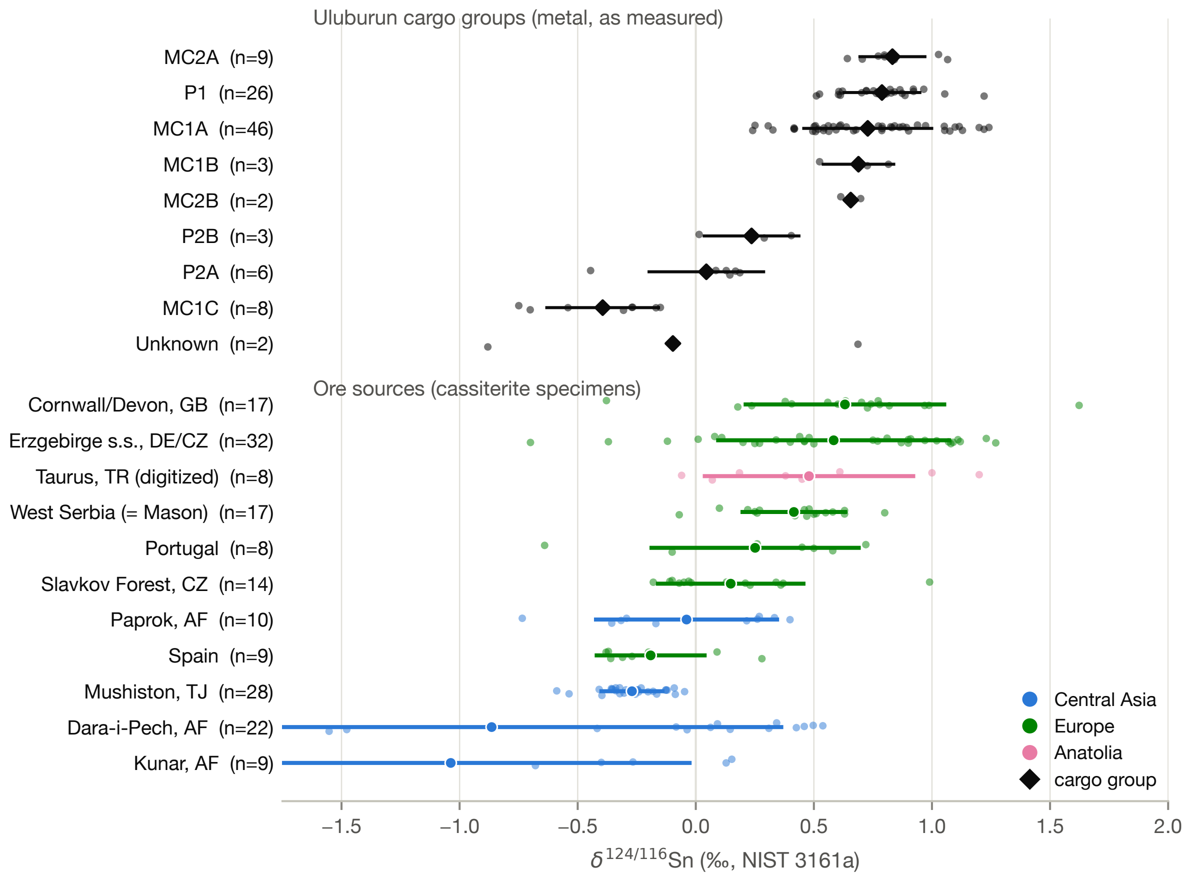

Figure 2: The identifiability problem in one picture. Top: the nine Uluburun cargo groups (metal, as measured), tightly clustered. Bottom: the eleven-source cassiterite reference library (specimen values; interval = mean ± 1 sd). The positive-$\delta^{124/116}$Sn European and Anatolian sources overlap each other almost completely; within corrected Spain, the Penouta mine alone (n=6) is a tighter negative signature than Mushiston; and every cargo group is far tighter than any source. Ore and metal scales differ by the smelting-fractionation shift δ (contested; +0.09–0.18‰ under complete reduction, up to +0.88‰ under incomplete reduction, Berger et al., 2018).

4 The per-ingot model fails three independent diagnostics

Fitted at face value, model A produces results of the same form as the dispute's published attributions: with a three-source library (Mushiston, Cornwall, Erzgebirge), Powell's "Central Asia" groups P1+P2A/B classify 34/35 to Cornwall ($\pi_{\mathrm{ENG}} = 0.65$); with Taurus added, "all 26 P1 ingots classify to Cornwall." Three diagnostics reject this.

4.1 The dispersion check: ingots are not ore specimens

If each ingot were an independent specimen draw, within-group spread should approximate source spread (0.38–1.2‰). Observed group spreads are 0.06–0.27‰ (Fig. 2). The posterior predictive check (Fig. 3) rejects the generative assumption outright for the three largest positive-$\delta^{124/116}$Sn groups (P1, MC1A, MC2A; 81 ingots): zero of 2,000 replicates are as tight as reality ($p < 5 \times 10^{-4}$ under even the best-fitting single source). Physically unsurprising — smelting

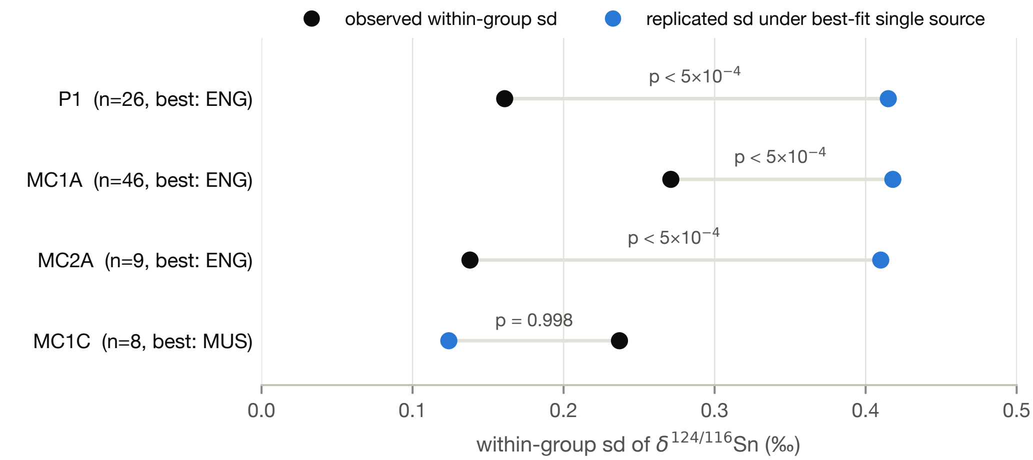

Figure 3: Posterior predictive check on within-group dispersion. Bayesian $p = P(\mathrm{sd}_{\mathrm{rep}} \le \mathrm{sd}_{\mathrm{obs}})$ under the group's best-fit *single* source. The three largest positive-$\delta^{124/116}$Sn groups (P1, MC1A, MC2A; 81 ingots) are 1.5–3× tighter than even their best-fitting single source generates ($p < 5 \times 10^{-4}$, i.e. zero of 2,000 replicates; per-group factors 2.6, 1.5, 3.0); MC1C is *wider* than pure Mushiston generates ($p = 0.998$).

homogenizes — but statistically consequential: *the effective unit of independent source evidence is the batch, not the ingot*, so "n=26 ingots classify to Cornwall" carries something close to one batch-mean's worth of evidence, not twenty-six. The check also discriminates: MC1C fails in the opposite direction ($p = 0.998$; too wide for pure Mushiston, too tight for the European sources), the signature of a mixed or multi-batch origin.

A circularity in this diagnostic must be stated at full width, because it touches the batch model too. Powell's groups were themselves defined partly by clustering the same isotope values modeled here, so part of the within-group tightness is selection, not smelting: the dispersion check partly rejects an assumption the grouping built in, and it cannot apportion the observed compression between the two mechanisms. This does not rescue per-ingot classification — selection-induced tightness and casting-batch homogenization both violate the iid-specimen assumption, and either alone deflates "n=26 independent draws" — but it does mean the check's quantitative reading (evidence ≈ one batch-mean per group) is a bound, not a measurement, and that model B, which adopts Powell's partition as its batch structure, is not an independent choice of unit. Three things limit the damage. First, the other two diagnostics below are computed on *ungrouped* fits over all 105 ingots — the fractionation profile and the library perturbation are partition-free, and each independently rejects the per-ingot unit. Second, the exclusion results of §5 ride on where the group means sit, not on how tightly ingots hug them: no re-partition of the cargo moves 86 positive-$\delta^{124/116}$Sn ingots onto a negative-$\delta^{124/116}$Sn source. Third, the partition is not merely carving tight bins — MC1C's failure in the *too-wide* direction is exactly what selection alone would not produce. This is why models A and B are presented as ends of a bracket rather than alternatives: A conditions on no grouping, B takes the grouping at face value, reality — some unknown batch structure — lies between, and both ends agree on the exclusions.

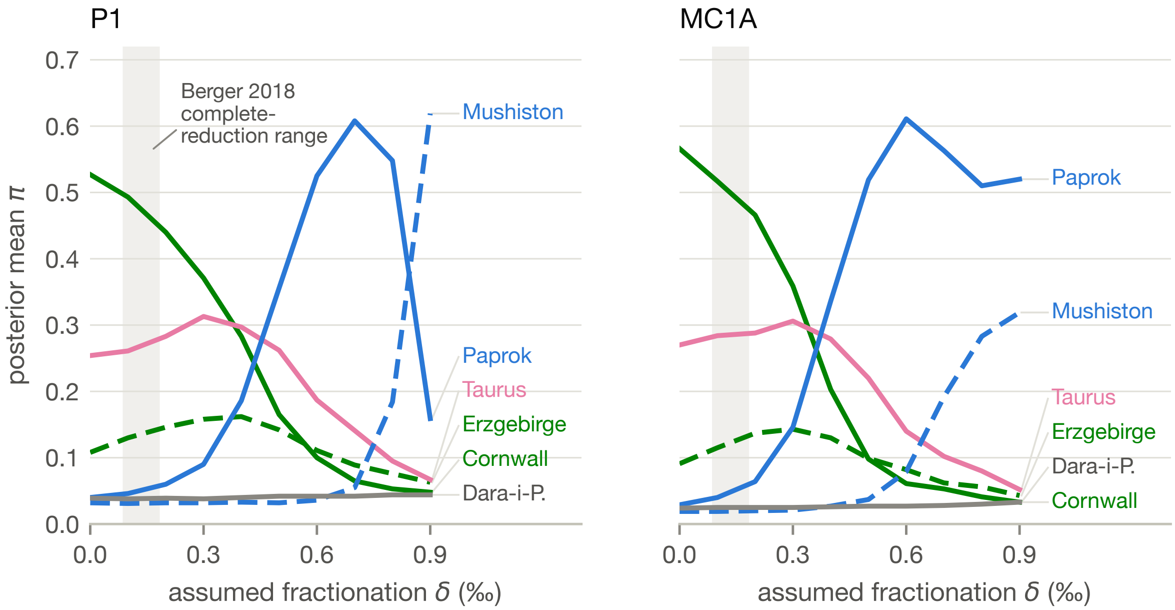

Figure 4: Posterior-mean π as a function of the assumed fractionation δ (model A, six-source library). The attribution for the cargo bulk is a function of an assumed smelting parameter: Cornwall requires $\delta \lesssim 0.25$; Paprok owns 0.5–0.8; Mushiston takes over by 0.9. The shaded band is the complete-reduction range that the baseline prior encodes.

4.2 The fractionation prior was carrying the headline

Under the heavy-tailed prior that reduction chemistry implies, model A's answer *inverts*: P1 flips to Mushiston ($\pi = 0.72$) at $\delta \approx 1.0$; MC1A to Paprok ($\pi = 0.47$) at $\delta \approx 0.67$. The fixed-δ profile (Fig. 4) shows why: δ shifts all ore means together, so it trades directly against which source sits nearest the tight ingot clusters, with the decision boundary at $\delta \approx 0.25$–0.35 — well inside the contested range. The mechanism is quantitative: a tight source matched to a tight cluster beats wide Cornwall by ∼0.45 nats *per ingot* under the iid likelihood (∼12 nats over P1), overwhelming the ∼5.6-nat prior penalty for extreme δ. That is, *the flip is powered by the exact assumption the dispersion check rejected*. Counterweights exist — $\delta \approx 1.0$ exceeds the experimental maximum and implies implausibly uniform extreme fractionation; MC1C independently pins δ low if the cargo shares one smelting regime; and Mushiston specifically is excluded on trace-element chemistry (Berger et al., 2023b) — but the isotope layer itself, with fractionation honestly uncertain, cannot decide Cornwall-vs-Central-Asia for the cargo bulk.

4.3 The library perturbation: the tightest source wins

Adding the remaining real sources (Table 2) plus an explicit outlier component produces the clearest demonstration of the pathology (Fig. 5): at the per-ingot unit, *Serbia* absorbs $\pi = 0.54\ [0.35, 0.70]$ of the cargo, with Cornwall collapsing to 0.06. No one proposes majority-Balkan tin at Uluburun; Serbia simply became the tightest positive-$\delta^{124/116}$Sn source in the library (sd 0.217). This retroactively explains the earlier Cornwall headlines as *tightest-source-then-available* artifacts. Per-ingot membership probabilities tell the same story from below: every P1 ingot individually carries ≈ 0.5/0.3/0.15 across the top three candidates — "all 26 classify to Cornwall"

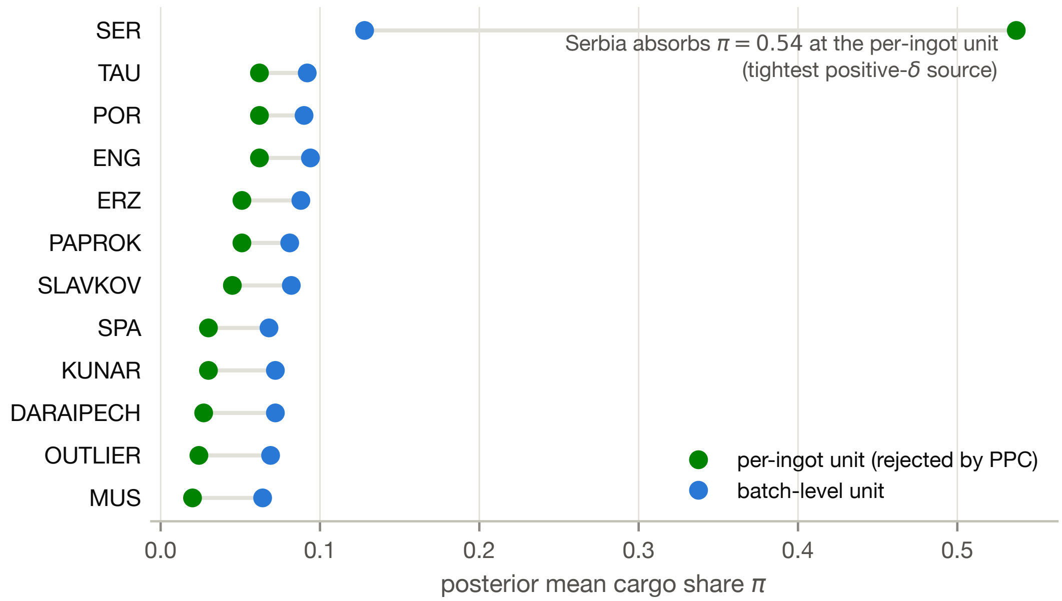

Figure 5: Cargo share π per component under the 11-source + outlier library, per-ingot vs batch-level unit. At the per-ingot unit Serbia — the tightest positive-$\delta^{124/116}$Sn source — absorbs $\pi = 0.54$ of the cargo. At the batch level no component exceeds 0.13.

restates the same three-way uncertainty twenty-six times, near-identical across ingots because the model's discrimination happens entirely at the group-mean level.

5 What survives: the batch-level model

Model B, run under both fractionation priors, resolves both pathologies at once. The δ dependence vanishes — under the same heavy-tailed prior that flipped model A, the batch model stays at complete reduction ($\delta = 0.164$ $[0.07, 0.49]$, $P(\delta > 0.35) = 0.07$) and every π_k moves by ≤ 0.013 — exactly as the nats arithmetic predicts, since the tight-source bonus now accrues once per batch rather than once per ingot. And the library perturbation no longer flips leaders: no component exceeds $\pi = 0.13$ (Fig. 5).

What remains (Fig. 6; mapped in Fig. 7) is a structure of *exclusions*:

- **The cargo bulk (P1, MC1A, MC1B, MC2A, MC2B; 86 ingots) excludes every negative-$\delta^{124/116}$Sn source, with unequal strength.** Mushiston sits 6–7σ from the bulk means and its membership (≈ 0.000) is robust to the fractionation prior, which it previously was not; Kunar and Dara-i-Pech are likewise far. Spain is different in kind: at the posterior fractionation it sits $\approx$ 3–3.8σ from the bulk (membership ≤ 0.005; the §2.2 correction strengthened this exclusion, since the misfiled Pyrite Belt sample had been inflating Spain's dispersion), but near the top of δ's credible interval (0.49) the gap still narrows to $\approx 2\sigma$ — Spain's exclusion remains δ-conditional in a way Mushiston's is not. This is the strongest surviving statement against the Central-Asia attribution of Powell et al. (2022) for these groups, and it is exclusion-shaped, not attribution-shaped: Cornwall, the Erzgebirge, the Taurus, Serbia, and Portugal remain mutually unidentifiable

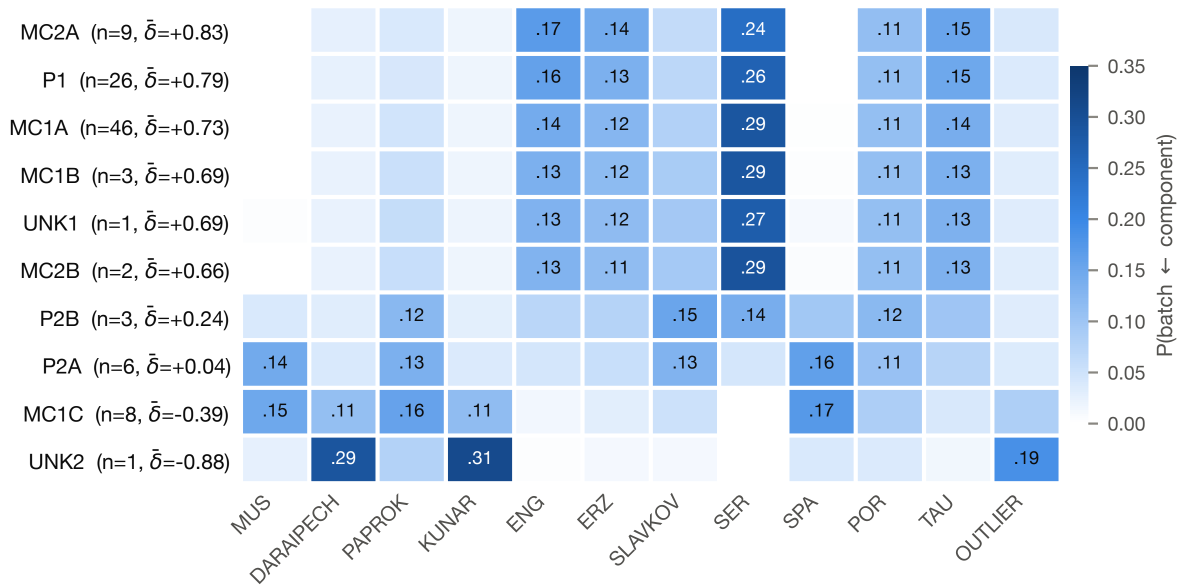

Figure 6: Batch-level source membership, 11-source + outlier library (rows ordered by mean $\delta^{124/116}$Sn). Exclusions are crisp (white cells); positive attribution is diffuse everywhere (no cell exceeds 0.30). Serbia's consistent lead across the bulk rows is the residual tightest-source artifact of §4 in softened form — it dissolves as batch aggregation grows (§5.1) and should not be read as an attribution. The two UNK rows are single ingots whose within-batch scale τ_g is prior-dominated; their memberships are shown for completeness and should be read accordingly. The outlier component attracts no batch except, mildly, the single -0.88‰ ingot.

(membership 0.11–0.29 each).

- **MC1C (8 ingots, mean -0.39‰) excludes Serbia** (0.000) **and Cornwall** (0.013) **at baseline dispersions — §5.1 grades how these hold under the sweeps; Portugal** (0.070) **and the Slavkov Forest** (0.046) **are disfavored, not excluded** — and with Iberia in the library its positive support splits: Spain, Paprok, and Mushiston are indistinguishable at 0.15–0.17 (Kunar 0.11), a spread the paper's own stress grid says not to rank. Summed Central-Asian support ($\approx$ 0.53) leads Iberia ($\approx$ 0.26), the inference "negative $\delta^{124/116}$Sn $\Rightarrow$ Central Asia" is unsafe, and the trace-element argument that excludes Mushiston-smelted tin (Berger et al., 2023b) does not transfer to Spanish cassiterite. MC1C is the cargo's most genuinely open question; the -0.88‰ ungrouped ingot shows the same exclusion pattern. (The sole published Bolkardağ Sn value — a stannite at -1.94‰, in-text in Powell et al., 2022 — sits 1.5‰ from the group Powell attributes to Bolkardağ; being a different mineral phase it is excluded from our library and carries no evidential weight in our verdicts, so we note the distance as context only.)
- **P2A and P2B (9 ingots, means +0.05 and +0.24‰) receive no exclusion at all.** They are the cargo's least decided batches (posterior perplexity $\approx$ 10 of 12 components); on P2A, Spain, Mushiston, Paprok, and Slavkov are indistinguishable at 0.13–0.16, and P2B's leaders (Slavkov, Serbia) sit in an equally tight band. If any part of Powell et al. (2022)'s Central-Asia attribution survives on the isotopic evidence, it is here — as one undecided possibility among several near-equals, not a positive finding.

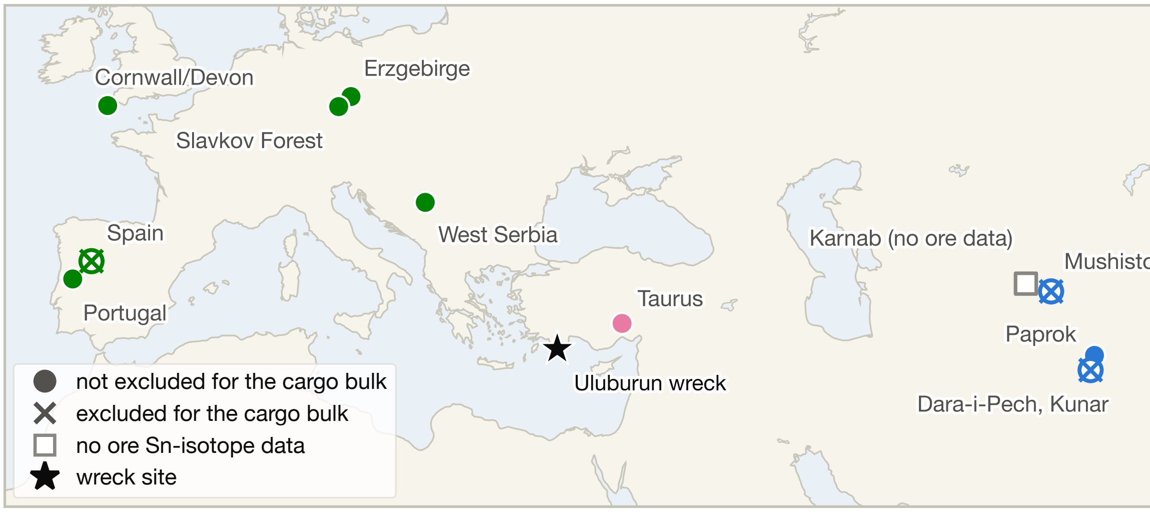

Figure 7: The batch-level verdict on the map. Marker color is region family; marker form is the verdict for the cargo bulk — filled circles are not excluded, × marks are excluded, and the open square is Karnab, for which no ore Sn-isotope data exists at all (§7). Locations are approximate district centroids.

- **The outlier component goes unused** ($\pi = 0.068$; maximum batch share 0.15, on the single −0.88‰ ungrouped ingot). Given an honest escape hatch, the data decline it: the expanded library plausibly spans the cargo.

5.1 Stress tests: what the exclusions survive, and what they do not

Twelve adversarial checks: eight tabulated so their symmetry is structural rather than vigilant (Table 3), four below that do not fit the table's format. All reported fits converged ($\hat{R} \leq 1.01$) and live in the repository.

Reading the table: the bulk exclusions survive everything, and the two real degradations are exactly where arithmetic predicts them. Per-batch δ_g — each batch independently free to reach the incomplete-reduction regime, the archaeologically natural model for a multi-consignment cargo — lifts Mushiston to 0.017 on the smallest low-mean batch (MC2B, n=2) and Spain to 0.034; MC1C is untouched. Joint 2× dispersion inflation degrades MC1C's exclusions to marginal grade (Serbia 0.041, Cornwall 0.051), the same grading Spain's bulk exclusion carries, and §5 uses that language for both. (The Student-t row is the principled propagation of σ sampling error; the 2× row is a deliberately blunt adversarial bound, indiscriminate about which entries — Serbia's carefully sampled n=17 among them — deserve it.) The tail-shape row closes a refuge in both directions: Mushiston's own 28 samples fit *left*-skewed ($\alpha = -1.8$, the opposite of the shape that would threaten the exclusion), and even the imposed adversarial right skew leaves its bulk membership at baseline levels.

Four checks do not fit the table's format. *Partition.* Everything above inherits Powell's ten batches; two repartitions free it. Seeded 1-D k-means on the raw ingot values — owing nothing to the published clustering — at k=6 and k=15 gives Mushiston 0.0000 on every positive cluster, Spain at most 0.012, and no positive-$\delta^{124/116}$Sn cell above 0.30. Since that still partitions the modeled variable, we also cluster the ingots on their *trace elements* (log In/Sb/Pb/Bi/Cu from Powell's own supplement, standardized; k=8 and k=12) and refit the isotope model on those

Table 3: Stress tests of the exclusion structure (corrected library; posterior membership maxima). "Unident. max" is the largest membership any positive-$\delta^{124/116}$Sn source achieves on any bulk batch — identifiability among the five never emerges under any check. The shared-offset row shown is the widest prior swept ($b \sim \mathcal{N}(0, 0.20)$, against an observed cross-orbit anchor of 0.034‰); b is not identified separately from δ, so that row reads as prior-robustness, not estimation. The per-orbit row is the *differential* version — an offset applied only to the CEZA-measured references (Mushiston, the Afghan deposits, and Williams's Cornwall) against the Mathur/Powell-orbit references and ingots, the asymmetry that most exposes Mushiston; its posterior is −0.007 [−0.29, +0.28] — the prior, as expected for an unidentified offset — so the row is profiled to breakdown instead: Mushiston's bulk membership does not reach 0.05 until the CEZA offset reaches ≈ 0.45–0.5‰, thirteen times the observed cross-orbit anchor, a magnitude no one will defend for laboratories sharing the NIST 3161a standard. The per-source δ_k row lets each source carry its own fractionation, the variant that would break the one-dimensional profile of Fig. 4; the exclusions degrade only to the per-batch-δ_g level.

Check	MUS (bulk)	SPA (bulk)	MC1C: SER	MC1C: ENG	Unident. max
baseline	0.000	0.005	0.000	0.013	0.29
per-batch δ_g, heavy prior	0.017	0.034	0.001	0.011	0.24
$\sigma \times 2$, all sources	0.006	0.063	0.041	0.051	0.26
$\sigma_{\mathrm{MUS}} \times 2$ only	0.004	0.006	0.000	0.012	0.29
shared offset $b \sim \mathcal{N}(0, 0.20)$	0.001	0.014	0.001	0.011	0.28
Student-t predictives ($\nu = n_k - 1$)	0.001	0.013	0.003	0.020	0.29
adversarial skew ($\alpha_{\mathrm{MUS}} = +5$)	0.001	0.008	0.000	0.014	0.28
per-orbit offset $b_{\mathrm{CEZA}} \sim \mathcal{N}(0, 0.15)$	0.001	0.006	0.000	0.014	0.29
per-source δ_k, heavy prior	0.023	0.040	0.000	0.014	0.26

batches — a partition owing nothing to the modeled variable at all: Mushiston ≤ 0.0035 on every positive cluster, Spain ≤ 0.026, largest positive cell 0.40. The circularity of §4 is thereby closed, not merely limited: the verdict structure is a property of where the cargo's values sit, not of who drew the group boundaries or of which variable drew them. *Admixture.* The exclusions are *single-source* statements, and in one isotopic dimension they weaken under mixing: modeling each bulk batch as a convex blend of a positive-$\delta^{124/116}$Sn draw and a Mushiston fraction f_g bounds f_g below only 0.48–0.59 (95%) under the complete-reduction prior — Mushiston as *majority* component is excluded, minority admixture up to roughly half is not — and under the heavy-tailed δ prior the isotope channel cannot bound admixture at all (95% bounds reach 0.85–0.96). A real admixture bound requires the chemical channel (Berger et al., 2023b); Sn isotopes alone cannot supply one. *Aggregation.* Model B's specimen-level batch variance ("deliberately conservative") is also the assumption that could be accused of manufacturing the unidentifiability, so we turned it into a curve: drawing batch means from $\sigma_k/\sqrt{m}$ for $m \in \{2, 4, \ldots, 128\}$ lowers bulk perplexity from 7.7 only to ≈ 4.5 effective components (5.0 at m=16, 4.7 at 32, 4.6 at 64, 4.5 at 128), with no membership cell above 0.39 anywhere on the curve. The flattening has a mechanism: once $\sigma_k/\sqrt{m}$ is small, the fractionation confound binds for the family as a whole — at high m the surviving unidentifiability is δ's, not overlap's. For the specific Cornwall–Erzgebirge pair the mechanism must differ, since δ shifts both together: their means sit 0.048‰ apart, 0.36 reference standard errors, so what binds there is the reference contrast itself (§8), and the two trade the lead along the plateau accordingly. Instructively, the leading candidate flips from Serbia (tightest source) to the nearest-mean sources as m grows: the tightest-source artifact dissolves exactly as §4 diagnosed. *Positive controls.* Synthetic cargoes generated under the model's own assumptions are recovered correctly: a Mushiston cargo yields ≥ 0.54 Mushiston membership on every batch (mean 0.60) — the machine says *yes* when the source is separated — while a Cornwall cargo puts Cornwall first on every batch but hedged (mean membership 0.20), the honest answer inside the overlap band. The pipeline's "no" for the real cargo is a property of the data geometry, not of the pipeline.

Independent evidence is consistent with this picture where it exists. Trace-element chemistry — the one discriminator with demonstrated power in this dataset (87% Cornwall-vs-Erzgebirge accuracy, vs ∼68–74% for isotopes) — favors Cornwall over the Erzgebirge specifically for the cargo bulk, but has no Taurus or Serbian reference data to complete the comparison. That figure earns no free pass from our own discipline: elemental partition during smelting is the trace-element analog of δ, and a serious joint model would carry partition-coefficient uncertainty the way ours carries the fractionation prior. Lead isotopes were evaluated and set aside: ingot Pb matches no candidate ore and is dominated by external contamination, consistent with Powell et al. (2024a); in our runs Pb's only effect was to reward whichever source had the thinnest Pb reference sample.

5.2 The granularity demonstration: Penouta

The granularity confound of §8 admits a demonstration rather than only a description. Penouta — the single Galician mine that constitutes two-thirds of the corrected Spain entry — enters the model as its own twelfth component (n=6, -0.285 ± 0.075‰): nominally $\sim 12\sigma$ from the cargo bulk, nearly double Mushiston's distance, from a mine no one has ever proposed for this cargo. Fitted, it takes essentially no bulk membership (≤ 0.0002, converged at strict settings) — a "stronger exclusion" than the paper's flagship, manufactured entirely by the choice of an entry's aggregation level. One accounting caveat is stated rather than hidden: as fitted, Penouta and Spain are *nested* components — the same six specimens inform both — so their shares are not disjoint and only Penouta's column bears interpretation. A disjoint refit with Spain-minus-Penouta (n=3, poorly estimated and said so) leaves the demonstration's headline unchanged, as it must, since the bulk exclusion depends only on Penouta's own location and scale. (In the disjoint fit: Penouta 0.0000 on the bulk; on MC1C, Penouta 0.095 with the Spain remainder at 0.114.)

The demonstration also bounds itself, in both directions. At n=6 the sd's own 95% interval is $[0.047, 0.184]$, so the nominal 12σ is anywhere from ~ 5 to $\sim 19\sigma$ — which is equally the honest limit of the Student-t reassurance above, powerful for well-sampled entries and unable to rescue thinly sampled ones. The flagship survives the same arithmetic where Penouta cannot: at Mushiston's n=28 the sd's 95% interval is $[0.102, 0.176]$, so its 6.6σ spans only ≈ 4.9–8.4σ — a stable σ-multiple, which is what distinguishes a real exclusion from a manufacturable one. Constructively, Penouta behaves as a genuinely tighter nested signature should, absorbing part of Spain's MC1C support (0.088, with Spain at 0.159 and Mushiston at 0.143 — indistinguishable, per our own rule). The moral for any provenance library: a σ-multiple quoted without its entry's n and aggregation level is not comparable across the library, and Table 2 must be read under that rule.

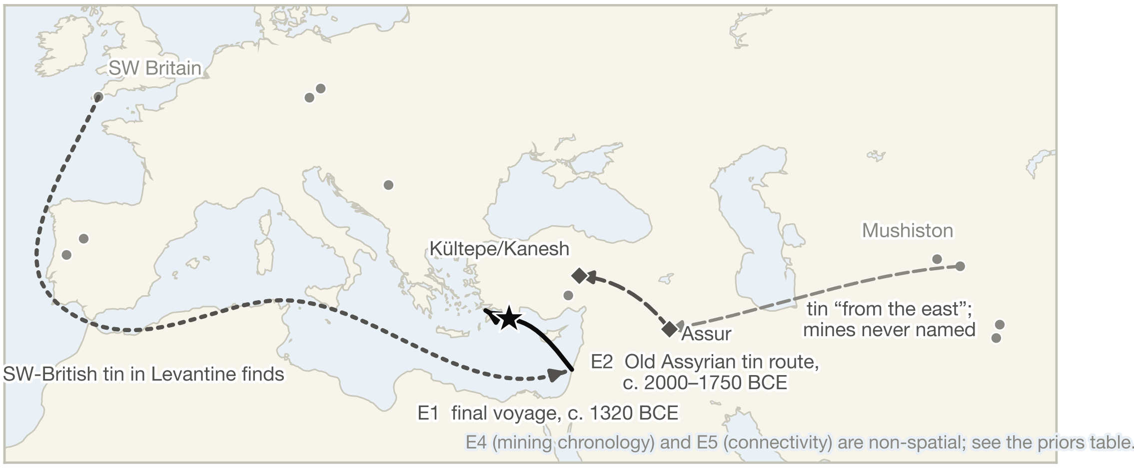

Figure 8: The spatial evidence lines behind the trade-route prior. E1: the ship's reconstructed final voyage (Levant → Cyprus → wreck site, heading west; Pulak, 1998). E2: the Old Assyrian tin trade attested at Kültepe/Kanesh — ∼500 years before the wreck, moving tin westward from Assur, with an eastern supply implied but no mine ever named. E3: SW-British tin provenanced in Levantine finds (Williams et al., 2025). Two further evidence lines (E4 mining chronology, E5 connectivity) are non-spatial; the full evidence-to-weight table is in the repository.

6 Trade-route priors: the apportionment is prior-driven; the exclusions are not

Everything so far used a flat Dirichlet($\mathbf{1}$) prior on the mixture weights. That is itself a substantive assumption, and not even a neutral one: with eleven components spread unevenly over regions, a flat prior over *components* implies region shares proportional to component counts — Central Asia, holding four library entries, gets $4/12 \approx 0.3$ of the cargo before any isotope is measured. Whatever the flat-prior posterior says about positive apportionment inherits that arbitrariness. We therefore built informative priors from the archaeology and asked what they change.

Five evidence lines (Fig. 8) — the wreck's westward itinerary (E1), the Old Assyrian textual vector, explicitly time-discounted (E2), provenanced comparanda (E3), mining-in-operation chronology at 1320 BC (E4), and connectivity (E5) — were scored as per-region log-odds weights in a table that was *frozen by pre-registration*: the weight-table commit precedes every posterior computed under it, and each weight is a single attackable cell, not a diffuse judgment call. One scoping choice deserves flagging: E3's comparanda (the SW-British provenance of Levantine ingots, Williams et al., 2025) are genuine likelihood-eligible evidence *in a joint model over multiple assemblages*; because this paper's likelihood covers one cargo, they can enter only through the prior, and readers who would rather see them in a multi-assemblage likelihood are pointed to that extension explicitly rather than told the evidence is "just prior." Two limits of the pre-registration itself: the freeze binds the analyst against tuning weights to a desired posterior, but it cannot blind the elicitor, who had already seen §4–5 and therefore knew the likelihood would push back weakly; and single-analyst SHELF is SHELF without the group-elicitation discipline that gives the framework its usual warrant. The table's value is its cell-level attackability, not analyst neutrality; a third-party-timestamped registration would be stronger and is the obvious upgrade. The construction follows single-analyst SHELF elicitation (Gosling, 2018);

Table 4: Batch-model posterior region shares under the five prior families, with the Mushiston-membership maximum over the five bulk groups and a prior-predictive conflict check (two-sided p on the mean and sd of the ten batch means; small values would indicate prior–data conflict). Sampler diagnostics clean ($\hat{R} \leq 1.01$) for every row (under the corrected library all five families converged at default settings).

Prior family	SW Brit.	Erzgeb.	Iberia	Balkans	Taurus	C. Asia	max MUS (bulk)	conflict p
flat (baseline)	0.094	0.170	0.158	0.128	0.092	0.288	0.0000	0.60 / 0.57
route-informed	0.193	0.123	0.088	0.161	0.100	0.225	0.0003	0.89 / 0.56
Powell-aligned	0.019	0.019	0.018	0.025	0.597	0.245	0.0002	0.97 / 0.63
Berger-aligned	0.447	0.176	0.096	0.054	0.046	0.096	0.0008	0.47 / 0.98
mixture (ψ-robust)	0.143	0.147	0.124	0.127	0.094	0.267	0.0002	0.72 / 0.52

region masses are the normalized exponentiated column sums, scaled by $\alpha_0 = 6$ (the whole prior carries roughly six batches' worth of weight against ten real batches) and spread equally within regions. Around the elicited ("route") prior we bracket with four more families: the flat baseline; a *Powell-aligned* enthusiast prior (region mass 0.62 Taurus + 0.30 Central Asia, $\alpha_0 = 10$); a *Berger-aligned* prior (0.50 SW Britain + 0.20 Erzgebirge); and a robustness mixture $\pi = \psi\, q_{\text{route}} + (1-\psi)\, p_{\text{flat}}$, $\psi \sim \text{Beta}(2, 2)$, which lets the data down-weight the informative component if it misfits.

The section's headline number comes from a concentration sweep. Defining $M(\alpha_0)$ as the L_1 distance between prior and posterior region shares, refits at $\alpha_0 \in \{1, 3, 6, 10, 20, 50, 100, 300\}$ trace a curve that fits $M(\alpha_0) \propto N_{\text{eff}}/(N_{\text{eff}} + \alpha_0)$ almost exactly, with $N_{\text{eff}} \approx 3$: *the ten batch means carry about three pseudo-batches of information about the positive apportionment* — the rest of their informativeness is spent on the exclusions — which is the paper's data/prior division of labor as a single number. Everything else in Table 4 follows from it. Pass-through at $\alpha_0 = 6$–10 (Powell's 0.62 Taurus mass emerging as 0.597; Berger's 0.50 SW Britain as 0.447) is what $N_{\text{eff}} \approx 3$ predicts, not a separate discovery; and the robustness mixture's ψ posterior (0.51, its $\text{Beta}(2, 2)$ prior mean) is the same fact in a second metric — with four pseudo-observations in ψ's prior against three in the likelihood, an unmoved ψ is *predicted*, and it duly arrives. No family shows prior–data conflict ($p \geq 0.48$ throughout; check in the spirit of Evans and Moshonov, 2006), though at $N_{\text{eff}} \approx 3$ that check has little power and confirms consistency rather than sanity. Meanwhile *every exclusion is prior-invariant*: Mushiston's membership for the cargo bulk stays below 10^{-3} under all five families, including the prior built to favor it. A knockout sweep closes the loop on the elicitation itself: refitting the route-informed family twenty times with each nonzero weight cell zeroed in turn moves no region's share by more than 0.063 — the largest swings belong, as they should, to the three largest weights — and leaves Mushiston's bulk membership below 10^{-3} in every refit. No single piece of archaeological evidence is load-bearing.

This is the paper's thesis measured from a second direction, and its scope must be stated with it. §4–5 showed this likelihood cannot pick among the positive-$\delta^{124/116}$Sn sources; this section shows the corollary: any positive apportionment of the Uluburun cargo that rests on the Sn-isotope channel is, *on that channel*, a restatement of its authors' priors. Both camps also argue from Pb, trace elements, and cross-assemblage comparanda; those channels are not adjudicated here — they are currently uncomputable for this question (contamination-dominated, reference-poor, and single-cargo-inaccessible respectively), which is an emphatically different status from "prior-driven." The exclusions — and only the exclusions — are what the isotope data themselves insist on.

7 Three structural data gaps

The analysis is capped by the reference data, in three documented ways. **Taurus:** no tabulated per-sample cassiterite Sn-isotope dataset exists. The nine Taurus-region Sn values reported to date reconcile as follows: eight cassiterite values, all plotted in Powell et al., 2024b, Fig. 6 (two of which — the new Hisarcık samples — are also stated exactly in-text, anchoring our digitization), plus one Bolkardağ *stannite* value given in-text only, which Table 2 excludes as a different mineral phase; hence n=8 in the library. The 2022 paper's supplement and both of its declared data deposits (GitHub and Zenodo) contain only the 105-ingot artifact table, the 2024 paper has no retrievable supplementary files, and the 2026 supplement is artefact-only. Our digitized parameters cross-validate against the two stated values and the stated nine-sample range. The granularity confound (§8) cuts here too: the entry pools Kestel, Bolkardağ, and Hisarcık into n=8, so deposit-specific claims — MC1C-from-Bolkardağ above all — cannot be addressed by this library entry even in principle. **Karnab:** the Uzbek half of "Central Asia (Uzbekistan and Tajikistan)" has *no ore Sn-isotope data at all*: per Berger et al. (2023a), Zirabulak-district ore was sampled but proved too low-grade for cassiterite separation, leaving three samples from neighboring Lapas — whose values (1.64–1.71‰) are heavier than every ingot in the cargo. The Uzbek attribution is currently unfalsifiable and unconfirmable. **The undeposited compilation:** Powell et al. (2026) describes "a global dataset of 857 cassiterite Sn isotope analyses"; the deposited subset (Mathur et al., 2025) contains 286 physical rows (one author-side duplicate and six non-cassiterite rows among them — our parser-level reconciliation is in the repository). The remaining ∼570 analyses have never been deposited *as data*, so the compilation cannot be audited, reharmonized, or reused as a dataset by anyone outside the producing group. Some unknown fraction of those rows are previously published values scattered through the older literature and are individually inspectable there; the never-published fraction is, by construction, unknowable from outside. Either way, the analyses underpinning this dispute's largest claimed evidence base cannot currently be checked against the claims made from them. A formal request for the compilation was sent to the corresponding authors of Mathur et al. (2025) and Powell et al. (2026) on 1 August 2026; the outcome, including any non-response, will be reported in any submitted version. The highest-value next step in this field requires no new laboratory work: it is a data release.

8 Discussion

The answer to the title question is: *identifiable at the family level, not at the deposit level.* Everything that survives our diagnostics is an exclusion — negative-$\delta^{124/116}$Sn sources excluded for the cargo bulk; Serbia and Cornwall excluded for MC1C at baseline dispersions, with Portugal and Slavkov disfavored and the exclusions grading toward Spain-strength under §5.1's sweeps — and every deposit-level attribution, from either camp or from our own intermediate runs, failed at least one diagnostic: the dispersion check, the fractionation sensitivity, or the library perturbation. The prior-family analysis (§6) states the same conclusion in the language of priors: what the data cannot decide, the prior decides, and both camps' positive apportionments pass through this likelihood nearly untouched — a statement scoped to the Sn-isotope channel, with the chemical and comparanda channels open rather than adjudicated. Table 5 maps the dispute's principal published claims onto the three distinct verdicts this analysis can return; the distinction between *contradicted*, *unsupported*, and *untestable* does much of the paper's work, and conflating

Table 5: The dispute's principal claims, each mapped to its status under this analysis (Sn-isotope channel, batch level). "Unsupported" means the data cannot distinguish the claim from its rivals, not that it is wrong.

Published claim	Status here
Powell et al. (2022): one third of the cargo is Central Asian, Mushiston central	Of the ∼35 ingots this claim covers: 26 (P1) **contradicted** as sole-source (Mushiston 6–7σ, robust across §5.1's sweeps; admixture bounded only weakly by isotopes alone) and 9 (P2A/B) **unidentifiable**, with Mushiston among four indistinguishable leaders on P2A (0.13–0.16)
Powell et al. (2022): the Uzbek (Karnab) component of Central Asia	**Untestable** — no ore Sn-isotope data exists (§7)
Powell et al. (2022): two thirds of the cargo is Taurus tin	**Unsupported** — Taurus is one of five mutually unidentifiable positive-$\delta^{124/116}$Sn sources, on a figure-digitized reference
Powell et al. (2022): MC1C derives from Bolkardağ	**Unidentifiable** on library evidence — the sole published Bolkardağ Sn value is a stannite, excluded from the library as a different mineral phase and given no evidential weight here. Collaterally, MC1C's negative-$\delta^{124/116}$Sn affinity (Central Asia jointly leading; Spain, Paprok, and Mushiston indistinguishable at the top) counts *against* a Taurus derivation without adjudicating what it was instead
Berger et al. (2023b): Mushiston is excluded for the cargo bulk	**Supported as sole batch source**, and strengthened: that exclusion survives fractionation-prior sensitivity, §5.1's sweeps, and all five trade-route prior families. The stronger published form ("not a source" at all) exceeds what isotopes alone establish — admixture is bounded only weakly (§5.1)
Williams et al. (2025): the cargo bulk is SW-British tin	**Unsupported on isotopes** — exactly as unidentifiable against the Erzgebirge, Taurus, Serbia, and Portugal; its real support is chemical, outside this likelihood

them would overstate every result in it.

This is quantitative support for the exclusion-over-inclusion doctrine long argued on geochemical grounds (Pernicka, 2014), now demonstrated as a statistical property of this specific dataset. It cuts against both camps' headline modes: Powell et al. (2022)'s positive Central-Asia/Taurus apportionment is unsupported at any unit of analysis we can defend (and its Uzbek component untestable), while the counter-attribution of the cargo bulk to Britain specifically

(Williams et al., 2025; Berger et al., 2023b) is exactly as unidentifiable against the Erzgebirge, the Taurus, Serbia, and Portugal on the isotopic evidence — its real support is chemical, and the chemical reference library covers too few candidates.

Four findings generalize beyond Uluburun. First, the *unit-of-evidence* problem: smelted metal is homogenized, so per-artifact mixture classifiers systematically overcount independent evidence by roughly the batch size, and their positive attributions track the smallest-variance reference source — a bias that grows, perversely, as reference libraries improve unevenly. Second, the *fractionation confound*: any process offset with contested magnitude, shared across all sources, trades one-for-one against source location; priors on it should be stress-tested with profile plots like Fig. 4 before an attribution is announced. Third, the *closed-world default*: without an explicit outlier component, misfit probability is silently reassigned to the least-bad candidate; with one, "no library source explains this" becomes an observable outcome. Fourth, the *granularity confound*: in a provenance mixture, component dispersion is the discriminator, so a library that mixes deposit-, district-, and country-level entries produces verdict strengths that track sampling granularity rather than geology — our own flagship contrast illustrates it: Mushiston and Spain sit 0.86 and 0.78‰ from the cargo bulk — comparable distances — while the difference between "6–7σ" and "$\approx 3\sigma$" is carried by the 1.8× ratio of their reference dispersions, i.e. by how each entry was aggregated and sampled. The dispersion sweeps of §5.1 exist because this dependence is real. A quantitative coda locates the ceiling: deciding Cornwall against the Erzgebirge (means 0.048‰ apart) at 10:1 odds would take ≈420 independent batch means against the ten this cargo affords — and the deeper cap is on the reference side, where the library's own contrast estimate sits 0.36 standard errors from zero (the probability its *sign* is correct is 0.64), so no quantity of cargo data can decide the pair until the reference library itself does, at roughly 700 samples per source against the 17 and 32 it holds.

Limitations. Groups-as-batches is the conservative end of a bracket, and the group labels themselves originate in Powell et al. (2022)'s isotope clustering — the circularity discussed in §4; MC1A's internal spread suggests multiple batches, so real evidence sits between our batch and per-ingot extremes (both ends agree on the exclusions). Source scales are estimated from small samples (n = 8–28); §5.1's Student-t refit propagates their sampling error, with no material change to any verdict. Serbia's tightness rests on n=17 non-independent samples. Two shared-offset risks act in the same slot as δ: inter-laboratory and standard-lot bias — the Central Asian references, European compilation, and ingot measurements come from different laboratory orbits, and any systematic offset is confounded with exactly the source-family contrasts the exclusions ride on — and three millennia of seafloor corrosion. One empirical anchor exists: Cornwall is measured in both laboratory orbits, and the observed cross-orbit difference of means is 0.034‰ (Mathur n=15, +0.666 vs. Williams n=17, +0.632) — small against every exclusion margin in the paper, though its own standard error (≈ 0.15‰) is wide at these sample sizes, which is why the composed-offset check below uses priors matched to that uncertainty rather than to the point estimate. Williams's seventeen serve as the Cornwall entry rather than a pool of both sets because Mathur's Cornwall rows are compilation reprints of earlier published values with unknown overlap against other library rows — the same non-pooling discipline applied to Mason/Mathur for Serbia — and because keeping the sets separate is what makes the cross-orbit comparison possible at all. Composing an explicit shared offset with δ leaves every verdict unchanged (§5.1); the ingot data's $\delta^{124/116}$Sn convention was verified against the source's own methods text. Taurus is figure-digitized; refitting model B with Taurus removed changes no exclusion verdict, and the freed mass redistributes diffusely — the largest single movement is Serbia's membership on the bulk groups (+0.053), and no positive attribution emerges — so the reconstruction bears weight only in the claim that the Taurus *itself* remains unidentifiable, the direction of caution. Source shapes are now tested against scale inflation, t tails, and skew (§5.1); multimodality is confirmed rather than suspected for the two widest entries — Hartigan dip tests reject unimodality for Dara-i-Pech ($p < 0.001$, n=22) and Kunar ($p = 0.008$, n=9), so those single Gaussians certainly blend mineralization styles. Splitting them requires sub-deposit metadata the source table does not carry; both sit far below the bulk on $\delta^{124/116}$Sn, so the bulk verdicts are insulated, but their MC1C and UNK2 memberships should be read as family-level only; the convex-mixing (recycling) alternative is just-identified at best with one observed dimension per ingot — its admixture bounds are now quantified (§5.1) and are the paper's weakest exclusion-adjacent result; and the analysis is Sn-only where it matters most — the trace-element check that discriminates Cornwall from the Erzgebirge lacks reference data for the other positive-$\delta^{124/116}$Sn candidates.

What would change the answer. A value-of-information analysis in the repository quantifies the baseline (posterior perplexity ≈ 8 of 12 components per bulk batch — more undecided, numerically, than the five-source prose conveys) and tests interventions against it: adding more Taurus Sn samples on the current protocol changes entropy by ≈ 0 *at the current dispersion estimate* — itself from eight digitized points, with a 95% interval spanning roughly 0.29–0.90‰; a substantially tighter true dispersion would revise this conclusion (the problem, at present knowledge, is population overlap, not mean uncertainty), while real trace-element data — no simulation; the Cornwall/Erzgebirge reference sheets already in hand — cuts the isotopically coin-flip ENG-vs-ERZ question by 18–69% depending on group. In value order, then: trace-element ore reference libraries for Taurus, Serbia, and Portugal; release of the ∼857-analysis cassiterite compilation (or the nine Taurus values as a table); any Karnab ore measurement (a sample-preparation problem, not a fieldwork one); and paired multi-isotope measurements that would make per-ingot mixing models identifiable.

Reproducibility and data availability

Every number and figure in this draft is generated by a script in the analysis repository (Reid, 2026), from the published datasets in Table 1; each result section corresponds to a dated writeup and commit. Reference datasets are redistributed only per their licenses (Powell 2022 data are CC BY-NC and are fetched, not re-hosted). Sampler settings, convergence diagnostics, the simulation-based-calibration run, and the parser-fidelity checks (including the Mathur row-count reconciliation) are documented alongside the code.

Acknowledgments

The analysis pipeline — data parsing, model implementation, diagnostics, and drafting — was built and executed with substantial agentic AI assistance (Claude, Anthropic), with all extractions cross-validated against the published sources and all modeling decisions and their failures documented in the repository's phase writeups. Errors remain the author's.

References

Michael J. Baxter. On the multivariate normality of data arising from lead isotope fields. *Journal of Archaeological Science*, 26(1):117–124, 1999.

Michael J. Baxter, Christian C. Beardah, and S. Westwood. Sample size and related issues in the analysis of lead isotope data. *Journal of Archaeological Science*, 27(10):973–980, 2000.

Daniel Berger, Elin Figueiredo, Gerhard Brügmann, and Ernst Pernicka. Tin isotope fractionation during experimental cassiterite smelting and its implication for the isotope composition of ancient tin. *Journal of Archaeological Science*, 92:73–86, 2018. doi: 10.1016/j.jas.2018.02.006.

Daniel Berger, Gerhard Brügmann, Ronny Friedrich, Joachim Lutz, Hans-Peter Meyer, and Ernst Pernicka. Shiny bronze in glassy matter: an inconspicuous piece of slag from the Bronze Age mining site of Muštiston (Tajikistan) and its significance for the development of tin metallurgy in Central Asia. *Archaeological and Anthropological Sciences*, 14:150, 2022.

Daniel Berger, Kai Kaniuth, Nikolaus Boroffka, Gerhard Brügmann, Sabine Kraus, Joachim Lutz, Mike Teufer, Andreas Wittke, and Ernst Pernicka. The rise of bronze in Central Asia: new evidence for the origin of Bronze Age tin and copper from multi-analytical research. *Frontiers in Earth Science*, 11:1224873, 2023a. doi: 10.3389/feart.2023.1224873.

Daniel Berger, Kai Kaniuth, Gerhard Brügmann, and Ernst Pernicka. Why Central Asia's Mushiston is not a source for the Late Bronze Age tin ingots from the Uluburun shipwreck. *Frontiers in Earth Science*, 11:1211478, 2023b. doi: 10.3389/feart.2023.1211478.

Michael Evans and Hadas Moshonov. Checking for prior-data conflict. *Bayesian Analysis*, 1(4): 893–914, 2006.

John Paul Gosling. SHELF: The Sheffield elicitation framework. In Luis C. Dias, Alec Morton, and John Quigley, editors, *Elicitation: The Science and Art of Structuring Judgement*, pages 61–93. Springer, Cham, 2018.

Andrea H. Mason, Wayne Powell, H. Arthur Bankoff, Ryan Mathur, et al. Provenance of tin in the Late Bronze Age Balkans based on probabilistic and spatial analysis of Sn isotopes. *Journal of Archaeological Science*, 122:105181, 2020. doi: 10.1016/j.jas.2020.105181.

Ryan Mathur, Wayne Powell, Junming Yao, Frederico Guimaraes, Yanbo Cheng, Linda Godfrey, Fernando Tornos, David Killick, Jay Stephens, Jingwen Mao, Mingguang Sun, and Bernd Lehmann. Global Sn isotope compositions of cassiterite identify the magmatic–hydrothermal evolution of tin ore systems. *Geosciences*, 15(1):28, 2025. doi: 10.3390/geosciences15010028.

Ernst Pernicka. Provenance determination of archaeological metal objects. In Benjamin W. Roberts and Christopher P. Thornton, editors, *Archaeometallurgy in Global Perspective: Methods and Syntheses*, pages 239–268. Springer, New York, 2014.

Wayne Powell, Michael Frachetti, Cemal Pulak, et al. Tin from Uluburun shipwreck shows small-scale commodity exchange fueled continental tin supply across Late Bronze Age Eurasia. *Science Advances*, 8(30):eabq3766, 2022. doi: 10.1126/sciadv.abq3766.

Wayne Powell, Ryan Mathur, George Kamenov, Jay A. Stephens, and David J. Killick. Constraints on the application of lead isotopes to the study of tin artifacts due to mixed Pb sources. *Journal of Archaeological Science*, 170:106045, 2024a. doi: 10.1016/j.jas.2024.106045.

Wayne Powell, K. Aslıhan Yener, Gojko Barjamovic, Fikri Kulakoğlu, Evren Yazgan, and Ryan Mathur. Tin isotopes reveal changing patterns of tin trade, connectivity and consumption from Anatolia and Central Asia at Kültepe. *Journal of Archaeological Science*, 162:105917, 2024b. doi: 10.1016/j.jas.2023.105917.

Wayne Powell, Ryan Mathur, K. Aslıhan Yener, and Michael Johnson. Mixed sources of Pb and Sn in late Bronze Age European tin ingots provide insights into production processes. *Journal of Archaeological Science*, 174:106135, 2025. doi: 10.1016/j.jas.2024.106135.

Wayne Powell et al. Shifting tin imports into the broader Eastern Mediterranean region (ca. 2000–900 BC). *Journal of Archaeological Science*, 2026. Article S0305440326000440, in press.

Cemal Pulak. The Uluburun shipwreck: an overview. *International Journal of Nautical Archaeology*, 27(3):188–224, 1998.

Conor Reid. iddin-annakam: a Bayesian mixture-model and identifiability reanalysis of the Uluburun tin cargo. Analysis repository: `https://git.sr.ht/~calgacus/iddin-annakam`, 2026. All model code, parsers, and per-result writeups referenced in this draft; archival DOI to be minted at submission.

Brian C. Stock, Andrew L. Jackson, Eric J. Ward, Andrew C. Parnell, Donald L. Phillips, and Brice X. Semmens. Analyzing mixing systems using a new generation of Bayesian tracer mixing models. *PeerJ*, 6:e5096, 2018. doi: 10.7717/peerj.5096.

Sean Talts, Michael Betancourt, Daniel Simpson, Aki Vehtari, and Andrew Gelman. Validating Bayesian inference algorithms with simulation-based calibration. *arXiv preprint*, arXiv:1804.06788, 2018.

Cezary Tomczyk. Tracing copper mixing in the Atlantic Bronze Age: A Bayesian interpretation of lead isotope data. *Archaeological and Anthropological Sciences*, 17:208, 2025.

Aki Vehtari, Andrew Gelman, and Jonah Gabry. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, 27:1413–1432, 2017. doi: 10.1007/s11222-016-9696-4.

Katrin J. Westner, Thomas Rose, Sabine Klein, et al. GlobaLID — global lead isotope database. GFZ Data Services, version 08/2025, 2025.

Alan Williams, Daniel Berger, et al. From Land's End to the Levant: did Britain's tin sources transform the Bronze Age in Europe and the Mediterranean? *Antiquity*, 99(405):708–726, 2025. doi: 10.15184/aqy.2025.41.