

Skeptical review: The Spatial Architecture of the Main Asteroid Belt: Size, Composition, and Dynamical Gradients

Summary

The manuscript presents a data-driven characterization of the main asteroid belt’s spatial “architecture” using a curated sample of 35,623 main-belt asteroids with diameters, orbital elements, and spectral classifications (Sec. 2.1–2.2, Sec. 3.1). The authors map size (often via \log_{10} diameter) and composition as functions of semimajor axis a , eccentricity e , and inclination i using 1D/2D binning with hypothesis tests, kernel density estimation (KDE) to reveal density structure (including Kirkwood gaps), unsupervised clustering (DBSCAN and Gaussian Mixture Models) in orbital-element space, and supervised models (XGBoost regression for $\log D$; Random Forest classification for spectral type) to assess predictability from orbits alone (Sec. 2.3–2.5, Sec. 3.2–3.4). The paper confirms well-known large-scale gradients—S-type dominance in the inner belt transitioning to C-type dominance outward—and reports an outward increase in mean size, identifies multiple compact clusters interpreted as families/dynamical regions, finds only modest orbit-only predictive power ($R^2 \approx 0.22$ for $\log D$; accuracy $\approx 53\%$ for type with low macro-F1), and reports size/composition differences inside vs. near vs. away from major resonance gaps interpreted as dynamical filtering plausibly involving Yarkovsky-driven drift (Sec. 3.2–3.4, Sec. 4). Overall, the work is timely and potentially valuable as a unified statistical/ML mapping of size–composition–dynamics, but several headline inferences are currently hard to evaluate because (i) observational selection/completeness and taxonomy/diameter provenance are not treated systematically, (ii) key methodological choices (bins/KDE/clustering/ML evaluation) are under-specified for reproducibility, and (iii) resonance and clustering interpretations would be substantially strengthened by more controlled comparisons (local conditioning, proper elements, family validation) and clearer separation of descriptive results from causal hypotheses (Sec. 2–4).

Strengths

- Clear scientific objective: jointly map size and composition across orbital-element space and relate patterns to dynamical/evolutionary processes (Sec. 1, Sec. 4).
- Large sample (35,623) combining physical and orbital information, with a coherent end-to-end analysis pipeline (Sec. 2.1–2.2, Sec. 3.1).
- Broad analytical toolkit (binned statistics + nonparametric tests, KDE, clustering, supervised ML) that provides multiple complementary views of belt structure (Sec. 2.3–2.5, Sec. 3.2–3.4).
- Empirical confirmation of strong compositional zoning and clear visualization of dynamical structure (e.g., gaps and density variations) consistent with established belt features (Sec. 3.2, Sec. 3.3.1).

- Generally careful physical discussion linking results to collisional evolution, primordial gradients, and dynamical transport, and an attempt to quantify resonance-associated differences rather than describing them qualitatively (Sec. 3.4–3.5, Sec. 4).
- Use of log-transform for diameter and standardization prior to distance-based clustering is appropriate and (mostly) consistently described (Sec. 2.3–2.4).

Major issues

1. **Selection effects, completeness, and dataset provenance are not documented or controlled, yet several central results (e.g., the outward increase in mean diameter; spectral-type fractions vs. a ; resonance-region contrasts) are sensitive to detectability and heterogeneous catalog coverage (Sec. 2.1–2.2, Sec. 3.1–3.2, Sec. 3.4).** Requiring both diameter and spectral type can impose a complex, distance- and albedo-dependent selection function that can mimic or amplify radial size/composition gradients. The manuscript also does not clearly specify the sources/versions of diameters and taxonomies, uncertainty scales, or conflict-resolution rules when multiple sources exist.

Recommendation: Add a dedicated subsection on data provenance and selection/completeness (expand Sec. 2.1–2.2 or add Sec. 2.2.1). (i) Provide a table mapping each key field (D , $a/e/i$, spectral type, H /albedo if used) to its source catalog(s), release/version, access date, and any quality-flag filters; state how conflicting values were resolved and typical uncertainties for D and taxonomy. (ii) Add an attrition/flow table (or diagram) reporting counts after each merge/filter step (initial catalogs \rightarrow merged \rightarrow after requiring D \rightarrow after requiring type \rightarrow after main-belt cuts in Sec. 2.2). (iii) Quantify how the final “ D +type” sample differs from the parent diameter-only sample in a , H (or D), and (if available) albedo; include a diagnostic plot such as D (or H) vs. a for both samples. (iv) Define at least one conservative “quasi-complete” subsample (e.g., a diameter threshold that is plausibly complete across 2.0–3.5 AU for the adopted diameter catalog) and repeat the key gradient and resonance analyses in Sec. 3.2 and Sec. 3.4 to demonstrate robustness; clearly state in Sec. 3.5/Sec. 4 which conclusions are conditional on selection effects.

2. **The headline “mean size increases with semimajor axis” result is not sufficiently validated against confounding by selection and compositional/family structure (Sec. 3.2.1).** Given strong type- a zoning and potential completeness variation with distance, an apparent size- a trend can arise without an intrinsic physical gradient.

Recommendation: Strengthen Sec. 3.2.1 with controlled and uncertainty-aware quantification: (i) report an effect size with uncertainty (e.g., slope of $\log_{10}(D)$ vs. a with 95% CI; not only Pearson r and $p \approx 0$). (ii) Repeat the trend within major taxonomic complexes (e.g., S-only, C-only, X-only; or “S-complex/C-complex/X-complex/V”) to check whether the trend persists after conditioning on composition. (iii) Perform

sensitivity analyses using multiple minimum-diameter cuts (e.g., $D > 5/10/20$ km) and show whether the slope and binned means remain stable. (iv) Where possible, separate family-dominated regions vs. background (or at least show the trend excluding the largest few families if family labels are available). Summarize these robustness checks in a short table/appendix and update Sec. 3.5/Sec. 4 wording to reflect what remains significant after controls.

- 3. Clustering claims are not yet physically well anchored: clustering is performed in (a, e, i) without clearly stating whether elements are proper or osculating, and the linkage to known asteroid families is asserted largely qualitatively with limited quantitative validation (Sec. 2.4.2, Sec. 3.3.2).** Because families are typically defined in proper elements (a_p, e_p, i_p) , clustering in osculating elements may mix or fragment families depending on epoch and secular evolution.

Recommendation: In Sec. 2.1–2.2 and Sec. 2.4.2, explicitly state whether $a/e/i$ are proper or osculating; if osculating, justify the choice and discuss limitations in Sec. 3.5. If proper elements are available, rerun DBSCAN/GMM in (a_p, e_p, i_p) (or provide a comparison on a subset). In Sec. 3.3.2, add quantitative cluster–family validation using published family membership labels: report purity/completeness per major family and/or an overall metric (e.g., adjusted Rand index), plus a contingency table for the main families. Also add robustness checks showing how the number of clusters, noise fraction, and the identity of major clusters change under reasonable DBSCAN ϵ/\min_s variations and GMM component-number ranges; this will allow you to narrow claims to the clusters that are stable and physically interpretable.

- 4. Resonance/Kirkwood-gap analysis is under-specified and currently risks confounding resonance effects with global radial gradients and taxonomy/family mix (Sec. 2.5, Sec. 3.4).** The definitions of “Inside Gaps”, “Adjacent to Gaps”, and “Background” are not given as explicit semimajor-axis intervals, and pooled comparisons may inadvertently compare different a -regimes rather than isolating resonance proximity. Interpreting observed differences as size-dependent dynamical filtering (e.g., Yarkovsky delivery into resonances) therefore remains ambiguous.

Recommendation: In Sec. 2.5, list the exact a -intervals used for each resonance and for each category (“inside/adjacent/background”), including window widths and boundary conventions. In Sec. 3.4, reframe the analysis to use local, controlled comparisons: for each resonance separately, compare inside vs. adjacent within a narrow a -range matched in distance from the resonance center (and ideally matched in inclination/eccentricity range), and report effect sizes per resonance (Δ mean $\log D$ with CI; Cramer’s V for type changes). Where feasible, repeat within major taxonomic complexes (S/C/X) and/or family vs. background to reduce confounding. Add a sensitivity analysis varying window widths/centers within literature-reasonable ranges and

show that conclusions persist. If you retain the Yarkovsky interpretation, explicitly label it as a consistent hypothesis and cite relevant dynamical work; otherwise soften causal language (Sec. 3.5, Sec. 4).

5. **Predictive modeling (XGBoost/RF) is reported at a high level without sufficient evaluation protocol detail, baselines, leakage controls, or uncertainty estimates, limiting interpretability of “limited predictive power” and its physical implications (Sec. 2.4.3, Sec. 3.3.3).** Random train/test splits can leak family/cluster structure across splits (near-duplicates in orbital space), inflating performance; accuracy $\approx 53\%$ alongside macro-F1 ≈ 0.16 suggests strong class-imbalance or majority-class dominance that is not discussed.

Recommendation: Expand Sec. 2.4.3 and Sec. 3.3.3 to include: (i) explicit sample sizes for each task and per class; (ii) the exact split/CV scheme (fold count, stratification, repeats, random seeds) and whether hyperparameter tuning is nested; (iii) class-imbalance handling (class weights/resampling) and full metric reporting as mean \pm SD (regression: R^2 /MAE/RMSE; classification: accuracy, macro- and weighted-F1, per-class precision/recall/F1) plus confusion matrices (appendix). Add simple baselines (e.g., predict mean $\log D$; predict majority type; or a -only logistic baseline) and compare against them. To mitigate leakage, consider blocked CV (e.g., by semimajor-axis bins) and/or group CV by family/cluster if labels exist; report how performance changes. Update interpretation in Sec. 3.5/Sec. 4 to distinguish “orbit-only features are insufficient” from stronger claims about stochasticity.

6. **Core methodological choices needed for reproducibility are missing or scattered, particularly numerical specifications for binning, KDE bandwidths, DBSCAN/GMM settings, and resonance-window definitions; additionally, statistical reporting focuses on tiny p -values with limited effect sizes/uncertainty, making practical significance hard to judge (Sec. 2.3–2.5, Sec. 3.2–3.4).**

Recommendation: Consolidate and specify all analysis parameters in Sec. 2.3–2.5 (and/or an appendix): (i) explicit bin edges (or start/end + number of bins) for $a/e/i$ and boundary conventions; (ii) KDE kernel type, dimensionality (2D/3D), bandwidth selection procedure and final bandwidths, feature scaling, and boundary treatment; (iii) DBSCAN $\epsilon/\text{min}_s\text{amples}$ used for final results and the exact k -distance configuration; (iv) GMM covariance type, initialization, convergence criteria, and component-number search range for AIC/BIC; (v) resonance intervals as per the resonance issue above. In Results (Sec. 3.2–3.4), supplement p -values with effect sizes and uncertainty (e.g., η^2 or rank-based analogs for Kruskal–Wallis; Cramer’s V for chi-squared; confidence intervals for correlations and mean differences) and briefly address multiple-testing control if many binwise tests are performed.

Minor issues

1. Taxonomy harmonization is unclear: the manuscript does not provide a mapping from original taxonomic labels/schemes to the final reduced set (e.g., S/C/X/V/Other), and merging many rare classes into “Other” can affect chi-squared tests and obscure meaningful gradients (Sec. 2.2, Sec. 3.2.2).

Recommendation: Add a short mapping table in Sec. 2.2 (or appendix) listing original taxonomy scheme(s) and how each class/subclass is mapped into the final labels (ideally by complexes: S-complex, C-complex, X-complex, V-type, etc.). Report class counts before/after merging and ensure chi-squared expected counts are adequate; if not, use a likelihood-ratio test or further aggregation to complexes.

2. Missing-data handling is listwise deletion for key variables, but the impact on representativeness is not quantified (Sec. 2.2, Sec. 3.1).

Recommendation: Report how many objects are removed for each missing field (D , type, $a/e/i$) and compare retained vs. removed distributions in a , e , i , and H (or D) where available. A brief table/figure will help readers assess missingness bias.

3. KDE and clustering comparability is unclear: it is not explicit whether KDE and clustering are performed in the same (possibly standardized) feature space, and how smoothing choices affect visibility of features like Kirkwood gaps (Sec. 2.4.1, Sec. 3.3.1–3.3.2).

Recommendation: Clarify in Sec. 2.4.1 and Sec. 2.4.2 the exact features and scaling used for each method, and add a brief bandwidth sensitivity statement (e.g., show one alternative bandwidth in an appendix) to demonstrate that key density structures are not artifacts of smoothing.

4. Figure interpretability: multiple figures are crowded or ambiguous (units, panel labels, sample sizes, and whether statistics are on D vs $\log_{10}(D)$); overplotting obscures dense regions (Figures 1–4, 6–11; also the mean-size plot discussion in Sec. 3.2.1).

Recommendation: Revise captions/axes to explicitly state units (AU/deg/km), whether plotted values are $E[\log_{10}(D)]$ vs $E[D]$ (or geometric mean), and include sample sizes per bin/group. Use density-preserving encodings (hexbin/KDE/alpha) where needed and annotate resonance locations on relevant panels.

5. Potential inconsistency/ambiguity in main-belt orbital bounds stated in different locations (Methods vs Results), flagged by an internal consistency check (Sec. 2.2 vs Sec. 3.1).

Recommendation: Re-audit and standardize the stated main-belt cuts ($a_{\min}/a_{\max}/e_{\max}/i_{\max}$) wherever repeated; consider defining them once in Sec. 2.2 and referring back to that definition elsewhere.

6. Statistical test reporting is often limited to p -values (sometimes effectively $p \approx 0$) without test statistics/DOF and with limited discussion of assumptions or post-hoc comparisons (Sec. 2.3.1–2.3.2, Sec. 3.2, Sec. 3.4).

Recommendation: Report the test statistic and degrees of freedom for each omnibus test; for Kruskal–Wallis across many bins, consider post-hoc pairwise tests (with correction) or report a global effect size plus selected planned comparisons.

7. The role of GAMs is introduced but not clearly reflected in Results (Sec. 2.4.3 vs Sec. 3.3.3).

Recommendation: Either summarize GAM performance and any partial-dependence insights in Sec. 3.3.3 or remove/relocate the GAM description to an appendix to streamline Methods.

8. Reproducibility and research artifacts: the workflow mentions saved outputs, but code/data availability is not clearly stated (Sec. 2.6, Sec. 4).

Recommendation: Add a clear code/data availability statement (repository/DOI if possible) including the processed analysis table, cluster labels, and scripts/notebooks; if restrictions apply, state how materials can be accessed.

9. Interpretation occasionally moves from descriptive associations to causal explanations (e.g., composition-dependent Yarkovsky clearing near resonances; low ML predictability implying stochastic history) without sufficient caveats (Sec. 3.4–3.5, Sec. 4).

Recommendation: Tighten language to distinguish observations from hypotheses; explicitly discuss alternative explanations (family mix, taxonomy bias, catalog heterogeneity) and indicate what additional modeling/data would be required to discriminate mechanisms.

Very minor issues

1. Affiliation line appears non-standard for journal submission (Unstructured report notes “Anthropic, Gemini & OpenAI servers. Planet Earth.”).

Recommendation: Replace with a conventional institutional affiliation (or omit if the venue allows anonymous/independent submission), following journal formatting requirements.

2. Minor typographical/formatting issues: word breaks (e.g., “forma\n\ntion”), inconsistent quotes for code-like names, HTML entities (e.g., “<”), and inconsistent heading styles (Sec. 1, Sec. 2.6).

Recommendation: Proofread LaTeX/source to remove OCR artifacts, standardize typography, replace HTML entities with proper math symbols, and harmonize subsection formatting.

3. Terminology inconsistencies: e.g., “siliceous” vs “siliceous”; mixed notation (ϵ vs eps ; $\log_diameter$ vs $\log_{10}(\text{diameter}_{\text{km}})$) (Sec. 3 and figure captions).

Recommendation: Standardize terminology and notation; add a short notation table or define the chosen conventions once and use them consistently.

4. Figure accessibility: some color choices may be inconsistent across panels or not color-blind-safe; duplicate colorbars and inconsistent normalization are mentioned (Figures 6–11).

Recommendation: Adopt a consistent, colorblind-safe palette across related figures, share colorbars where appropriate, and state normalization/binning choices in captions.

5. Reporting of extremely small p -values as “ $p \approx 0$ ” is imprecise (Sec. 3.1–3.4).

Recommendation: Report as $p < 10^{-k}$ (or “below machine precision”) and include the corresponding test statistic to avoid ambiguity.

Key statements and references

- ✓ **The locations of major mean-motion resonances with Jupiter that define the Kirkwood gaps (e.g., the 3:1, 5:2, 7:3, and 2:1 resonances) are adopted from values reported in the astrophysical literature and used to define semimajor-axis windows for targeted analysis of resonant regions.**
- *Reference(s):* Denario [11]
- *Justification:* No valid PDFs found; assumed supported.
- ✓ **The interpretation that the observed transition from S-type dominated populations in the inner main belt to C-type dominated populations in the outer belt reflects the protoplanetary disk’s thermal structure, with rocky S-types forming inside the snow line and primitive carbonaceous C-types forming beyond it, follows the widely accepted explanation in prior work.**
- *Reference(s):* Denario [11]
- *Justification:* No valid PDFs found; assumed supported.
- ✓ **The conclusion that smaller asteroids are preferentially cleared from mean-motion resonance regions, and that this size-dependent filtering is consistent with the Yarkovsky effect causing smaller bodies to drift more rapidly into unstable resonant orbits, is based on earlier studies of size-dependent non-gravitational forces on asteroids.**
- *Reference(s):* Denario [11]
- *Justification:* No valid PDFs found; assumed supported.

- ✓ The inference that dynamical clearing near resonances is not compositionally neutral, and may depend on spectral-type-dependent physical properties such as thermal inertia that modulate the response to non-gravitational forces like the Yarkovsky effect, builds on prior literature linking composition, thermal properties, and Yarkovsky-driven orbital evolution.
- *Reference(s)*: Denario [11]
- *Justification*: No valid PDFs found; assumed supported.

Mathematical consistency audit

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

Maths relevance: light

The paper is primarily methodological/descriptive, using standard statistical constructs (binning, KDE, clustering, regression/classification) and a single explicit transformation $\log_{\text{diameter}} = \log_{10}(\text{diameter}_{\text{km}})$. There are no extended algebraic derivations or equation chains to audit. Internal consistency mainly hinges on consistent definition/usage of transformed variables, clarity on which summary statistics are plotted (raw vs log space), and consistent specification of algorithm parameters (e.g., DBSCAN k -distance/min_samples relationship).

Checked items

1. ✓ **Log-diameter definition** (Sec. 2.2 (p.3): ' $\log_{\text{diameter}} = \log_{10}(\text{diameter}_{\text{km}})$ ')
 - **Claim:** Asteroid diameter is log-transformed using base-10 logarithm for analyses involving size.
 - **Checks:** definition consistency, notation consistency
 - **Verdict:** PASS; confidence: high; impact: moderate
 - **Assumptions/inputs:** $\{\text{diameter}\} \{ \text{km} \} > 0$ so \log is defined, $\log_{\text{base}} 10$ as written
 - **Notes:** Definition is explicit and is referenced consistently later as $\log_{10}(\text{diameter}_{\text{km}})$.
2. ✓ **Consistency of 'log_diameter' usage as main size variable** (Secs. 2.3.1, 3.1–3.4 (pp.3–9))
 - **Claim:** Subsequent analyses summarize or model asteroid size primarily via $\log_{10}(\text{diameter}_{\text{km}})$.
 - **Checks:** definition consistency, symbol consistency
 - **Verdict:** PASS; confidence: high; impact: moderate
 - **Assumptions/inputs:** When reporting 'mean $\log_{10}(\{\text{diameter}\} \{ \text{km} \})$ ' they are working in log space, *Regression target is \log* $\{\text{diameter}\} \{ \text{km} \}$

- **Notes:** Text repeatedly references $\log_{10}(\text{diameter}_{\text{km}})$ as the transformed size measure; no contradictory alternative definition appears.
3. ✓ **Geometric-mean interpretation of mean log diameter** (Sec. 3.2.1 (p.6): 'mean $\log_{10}(\text{diameter}_{\text{km}})$... (~ 3.6 km geometric mean diameter)')
- **Claim:** A mean in $\log_{10}(\text{diameter})$ corresponds to a geometric-mean diameter after exponentiation base 10.
 - **Checks:** algebraic relationship, interpretation consistency
 - **Verdict:** PASS; confidence: medium; impact: minor
 - **Assumptions/inputs:** Geometric mean of D is $10^{E[\log_{10}(D)]}$ when logs are base 10, They intend 'geometric mean' rather than arithmetic mean
 - **Notes:** The mapping from mean $\log_{10}(D)$ to geometric mean diameter is algebraically correct; numeric approximations are not audited.
4. △ **Figure 5: mean diameter vs mean log diameter ambiguity** (Sec. 3.2.1 (p.6) and Figure 5 caption (p.7))
- **Claim:** Figure 5 top panel shows the mean asteroid diameter increasing with semimajor axis.
 - **Checks:** notation/definition consistency, interpretation consistency
 - **Verdict:** UNCERTAIN; confidence: medium; impact: moderate
 - **Assumptions/inputs:** The paper computes both raw-diameter and log-diameter means (Sec. 2.3.1), Figure 5 could plot either mean(D) or mean($\log_{10}(D)$)
 - **Notes:** Main text emphasizes mean $\log_{10}(\text{diameter}_{\text{km}})$, but the Figure 5 caption says 'mean asteroid diameter'. Without explicit axis labels/units or a statement of back-transform, it is not verifiable which statistic is plotted.
5. ✓ **Binning scheme specification** (Sec. 2.3 (p.3))
- **Claim:** Orbital space is discretized with widths: $\Delta a = 0.1$ AU, $\Delta e = 0.05$, $\Delta i = 2.5$ degrees.
 - **Checks:** unit/dimension consistency, definition consistency
 - **Verdict:** PASS; confidence: high; impact: minor
 - **Assumptions/inputs:** a in AU, e dimensionless, i in degrees
 - **Notes:** Units/dimensions are consistent: a has length units, e is dimensionless, i is angular.
6. ✓ **Standardization prior to clustering** (Sec. 2.4.2 (p.4))
- **Claim:** Before DBSCAN/GMM, (a, e, i) are standardized to mean 0 and variance 1 to prevent scale dominance in distance computations.
 - **Checks:** methodological math consistency, unit/dimension handling
 - **Verdict:** PASS; confidence: high; impact: minor

- **Assumptions/inputs:** Standardization means $z = (x - \text{mean})/\text{std}$ applied per feature
 - **Notes:** This is internally consistent and addresses mixed units/scales appropriately for distance-based methods.
7. **△ DBSCAN eps selection via k-distance plot** (Sec. 3.3.2 and Figure 8 caption (p.8))
- **Claim:** ϵ (ϵ) is chosen from an elbow in the k -distance plot using the distance to the 5th nearest neighbor.
 - **Checks:** parameter-definition consistency, method specification completeness
 - **Verdict:** UNCERTAIN; confidence: medium; impact: minor
 - **Assumptions/inputs:** k -distance plot uses $k = \text{min}_s\text{amples}$ in standard DBSCAN heuristics, They used $k = 5$ (5th nearest neighbor)
 - **Notes:** The paper states $\text{min}_s\text{amples}$ was explored (Sec. 2.4.2) but the plot/caption fixes $k = 5$. If the final model uses $\text{min}_s\text{amples} \neq 5$, the ϵ -choice heuristic as described would be inconsistent; the final $\text{min}_s\text{amples}$ is not explicitly stated.
8. **✓ Regression target definition** (Sec. 2.4.3 (p.4) and Sec. 3.3.3 (p.8))
- **Claim:** Regression models predict $\log_{10}(\text{diameter}_{\text{km}})$ from (a, e, i) .
 - **Checks:** definition consistency, notation consistency
 - **Verdict:** PASS; confidence: high; impact: minor
 - **Assumptions/inputs:** Target variable is the previously defined base-10 log diameter
 - **Notes:** Consistent with the defined transformation and avoids mixing raw vs log targets in stated modeling objective.
9. **✓ Use of Kruskal–Wallis for across-bin size differences** (Secs. 2.3.1 and 3.2.1 (pp.3,6))
- **Claim:** Kruskal–Wallis is used to assess whether size distributions differ across semimajor-axis bins.
 - **Checks:** statistical-logic consistency
 - **Verdict:** PASS; confidence: medium; impact: minor
 - **Assumptions/inputs:** Bins define groups; observations are treated as independent within/among groups
 - **Notes:** Test choice is consistent with the stated goal of comparing distributions across multiple groups; detailed assumptions (independence) are not discussed but no internal contradiction is present.
10. **✓ Chi-squared test for spectral-type variation across bins** (Sec. 3.2.1 (p.6))
- **Claim:** A chi-squared test evaluates whether spectral-type proportions vary significantly across semimajor-axis bins.

- **Checks:** statistical-logic consistency, normalization/constraints
- **Verdict:** PASS; confidence: medium; impact: minor
- **Assumptions/inputs:** Counts per type per bin form a contingency table
- **Notes:** Internally consistent: proportions arise from counts; chi-squared is applied to count tables, not proportions directly (paper implies this workflow).

Limitations

- The paper contains very few explicit equations and no multi-step derivations; therefore the audit is largely limited to checking definition/notation consistency and whether stated statistical constructs align logically with the described workflow.
- Several key quantitative objects are presented only in figures (e.g., what exactly is plotted in Figure 5); without explicit axis labels in the provided text, some checks are necessarily marked UNCERTAIN.
- Numeric claims (p -values, correlation coefficients, means) are not audited per instructions; only the symbolic/interpretive relationships around those quantities are assessed.

Numerical results audit

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

Out of 16 candidate numeric checks, 14 passed, 1 failed (orbital-bounds consistency), and 1 was uncertain (binning-width repetition could not be verified). Key arithmetic/log transformations (log10-to-km conversions; percentage totals; percent-to-count) were internally consistent within stated tolerances.

Checked items

1. ✓ **C1** (Abstract (page 1) and Results §3.1 (page 5))
 - **Claim:** Dataset size is reported as **35,623** main-belt asteroids (with measured/complete properties).
 - **Checks:** repeated_constant_match
 - **Verdict:** PASS
 - **Notes:** Reported values matched exactly (**35,623** vs **35,623**).
2. ✗ **C2** (Results §3.1 (page 5))
 - **Claim:** Main-belt orbital bounds are stated as ($2.0 < a < 3.5$ AU, $e < 0.4$, $i < 30^\circ$).
 - **Checks:** repeated_constant_match
 - **Verdict:** FAIL

- **Notes:** Repeated-constant match failed with max deviation 28.0 when comparing the listed bounds as a repeated-constant set (per check implementation notes).
3. **⚠ C3** (Methods §2.3 (page 3))
- **Claim:** Binning widths: semimajor axis 0.1 AU, eccentricity 0.05, inclination 2.5 degrees.
 - **Checks:** unit_consistency_and_repeated_constants
 - **Verdict:** UNCERTAIN
 - **Notes:** Could not search/confirm other mentions of bin widths; only the provided bin-width values were available.
4. **✓ C4** (Results §3.1 (page 5))
- **Claim:** Spectral-type composition: S-type 44.8%, C-type 17.8%, X-type 12.0%, V-type 7.0%, B-type 4.6%.
 - **Checks:** percentage_sum_leq_100
 - **Verdict:** PASS
 - **Notes:** Sum of listed percentages = 86.2%, within the $\leq 100\%$ requirement.
5. **✓ C5** (Results §3.2.1 (page 6))
- **Claim:** Mean $\log_{10}(\text{diameter}_{\text{km}}) \sim 0.56$ corresponds to ~ 3.6 km geometric mean diameter.
 - **Checks:** log10_inverse_transform
 - **Verdict:** PASS
 - **Notes:** $10^{0.56} = 3.630780547701014$ km, consistent with the reported ~ 3.6 km.
6. **✓ C6** (Results §3.2.1 (page 6))
- **Claim:** Mean $\log_{10}(\text{diameter}_{\text{km}}) > 0.88$ corresponds to ~ 7.6 km in outer belt (3.1–3.2 AU).
 - **Checks:** log10_inverse_transform
 - **Verdict:** PASS
 - **Notes:** $10^{0.88} = 7.5857757502918375$ km, consistent with ~ 7.6 km; text notes '> 0.88' so the computed value is a lower bound.
7. **✓ C7** (Results §3.2.1 (page 6))
- **Claim:** Inner belt S-types are $\sim 75\%$ at 2.2 AU and decline to $< 10\%$ by 3.2 AU; C-types are $< 5\%$ at 2.2 AU and $\sim 47\%$ at 3.2 AU.
 - **Checks:** percentage_range_sanity
 - **Verdict:** PASS
 - **Notes:** All percentages lie in $[0, 100]$ and inequalities are logically consistent ($75 > 10$; $47 > 5$).

8. ✓ **C8** (Results §3.3.2 (page 8))
- **Claim:** DBSCAN identified 38 clusters, with 14.25% of asteroids classified as noise.
 - **Checks:** percent_to_count
 - **Verdict:** PASS
 - **Notes:** Computed $\text{noise_count} = 35,623 \times 0.1425 = 5076.2775$ (rounded 5076); $\text{clustered_count} = 30547$, positive.
9. ✓ **C9** (Results §3.3.2 (page 8))
- **Claim:** GMM optimal number of components is 10; clusters include examples with stated means and compositions.
 - **Checks:** percentage_bounds_and_consistency
 - **Verdict:** PASS
 - **Notes:** Checked stated cluster percentages are within $[0, 100]$ and mean semimajor axes (2.30, 3.14, 3.15 AU) fall within $[2.0, 3.5]$ AU.
10. ✓ **C10** (Results §3.3.3 (page 8-9))
- **Claim:** XGBoost regression achieved $R^2 = 0.22$; Random Forest classification accuracy 53%; S-type F1 = 0.70; macro-averaged F1 = 0.16.
 - **Checks:** metric_range_sanity
 - **Verdict:** PASS
 - **Notes:** All metrics fall in their valid ranges as checked (R^2 in $[0, 1]$ for this claim; accuracy in $[0, 100]$; F1 in $[0, 1]$).
11. ✓ **C11** (Results §3.4 (page 9))
- **Claim:** Kruskal-Wallis test $p \approx 10^{-174}$ for size distribution differences near/within gaps.
 - **Checks:** p_value_range_sanity
 - **Verdict:** PASS
 - **Notes:** Parsed $p = 1 \times 10^{-174}$ and confirmed it lies in $(0, 1]$.
12. ✓ **C12** (Results §3.4 (page 9))
- **Claim:** Mean $\log_{10}(\text{diameter}_{\text{km}})$ values: Adjacent to Gaps 0.739, Inside Gaps 0.722, Background 0.670.
 - **Checks:** ordering_consistency
 - **Verdict:** PASS
 - **Notes:** Verified $0.739 > 0.670$ and $0.722 > 0.670$.
13. ✓ **C13** (Results §3.4 (page 9))
- **Claim:** Convert mean $\log_{10}(\text{diameter}_{\text{km}})$ differences into diameter ratios for interpretability (implied by log scale).

- **Checks:** log10_ratio_recompute
 - **Verdict:** PASS
 - **Notes:** Computed ratios: adjacent/background = 1.1721953655481303; inside/background = 1.1271974561755103; both > 1 as implied.
14. ✓ **C14** (Results §3.4 (page 9))
- **Claim:** Chi-squared test $p \approx 10^{-129}$ for compositional mix changes near resonances.
 - **Checks:** p_value_range_sanity
 - **Verdict:** PASS
 - **Notes:** Parsed $p = 1 \times 10^{-129}$ and confirmed it lies in $(0, 1]$.
15. ✓ **C15** (Results §3.4 (page 9))
- **Claim:** X-type proportion: Inside Gaps 18.2% vs background 9.8% (enhanced inside gaps).
 - **Checks:** percentage_difference_and_ratio
 - **Verdict:** PASS
 - **Notes:** Confirmed 18.2% > 9.8%; computed absolute difference = 8.4 percentage points and enhancement factor = 1.857142857142857.
16. ✓ **C16** (Results §3.1 (page 6))
- **Claim:** Pearson correlations: $r = 0.427$ ($\log_{10}(\text{diameter}_{\text{km}})$ vs semimajor axis) and $r = 0.251$ (semimajor axis vs inclination).
 - **Checks:** correlation_range_sanity
 - **Verdict:** PASS
 - **Notes:** Both correlation coefficients are within $[-1, 1]$.

Limitations

- Only the provided parsed PDF text was available; no underlying dataset, tables of bin counts, or model outputs were included, preventing recomputation of statistical tests, correlations, KDEs, clustering metrics, or ML performance.
- Figure-based numeric claims were not audited via pixel/plot extraction per instructions; checks were limited to explicit numeric values written in the text.
- Several statements use approximate qualifiers (e.g., '~', 'over', 'less than'); tolerances were suggested, but exact verification is inherently limited without source data.