

Skeptical review: Quantifying Yarkovsky-Driven Orbital Dispersion Gradients and Proxy Efficacy in Asteroid Families

Summary

This manuscript introduces the Orbital Dispersion Gradient (ODG) method to quantify Yarkovsky-driven spreading in asteroid families without defining a family center/apex (Secs. 1, 2.4–2.5). The approach bins family members by “Yarkovsky-sensitive” proxies ($\log_{10}(1/D)$, $\log_{10}(1/P)$, $\log_{10}(1/(P \cdot D))$; Sec. 2.3), computes the semimajor-axis dispersion within each bin (σ_a), and fits a (weighted) linear model $\sigma_a = G \cdot (\text{proxy}) + C$, using slope G as the ODG and R^2 to compare proxy efficacy (Secs. 2.5, 3.2–3.3). Applied to a merged dataset of 16,364 objects across six main-belt families (Eunomia, Vesta, Flora, Koronis, Eos, Maria; Secs. 2.1–2.2, 3.1), the diameter-only proxy performs best in four families (Maria, Eos, Koronis, Vesta), Proxy_PD wins only for Eunomia, and the spin-only proxy generally performs poorly (often negative R^2). Flora shows no clear linear relationship, consistent with a more complex structure (Sec. 3.2.2). A secondary Spearman test between family age and best-proxy gradient suggests a positive but statistically insignificant trend (Secs. 2.6.2, 3.4). The method is promising and conceptually simple, but the physical interpretation of σ_a -in-bins, robustness to binning/selection/membership effects, and key implementation details (binning, weights, uncertainty definitions, and R^2 under weighting) require clarification and validation for the results—especially proxy comparisons—to be fully reproducible and convincingly generalizable.

Strengths

- Introduces a conceptually simple, center-free statistic (ODG) that avoids reliance on a precisely defined family apex/center while remaining straightforward to compute (Secs. 1, 2.4–2.5).
- Applies a uniform pipeline across multiple major families with a clearly stated completeness requirement (≥ 100 members with D , P , a , and family label), enabling family-by-family comparison (Secs. 2.1–2.2, 3.1).
- Empirical findings are broadly consistent with canonical Yarkovsky scaling with size ($da/dt \propto 1/D$), and the paper usefully demonstrates the limited explanatory power of spin-period-only information in the current dataset (Secs. 3.2–3.3, Table 1).
- Clear visual juxtaposition of raw family structure and the binned regression summary (Figs. 1–6), helping readers connect qualitative V-shape-like patterns to the quantitative ODG fits.
- The manuscript explicitly acknowledges several important limitations (selection biases, missing obliquity/thermal properties, complex families), providing a good foundation for a more rigorous/robust version of the method (Secs. 3.3–3.5, 4.4).

Major issues

1. **Physical meaning of σ_a within proxy bins is not sufficiently justified, which weakens interpretation of ODG as a Yarkovsky diagnostic (Secs. 2.4–2.5, 3.2–3.3).** Unlike classic V-shape/envelope approaches, σ_a uses the full interior distribution and can be strongly influenced by initial ejection velocities, resonant sculpting/asymmetries, interlopers, and observational truncation. As a result, a trend in σ_a vs proxy may not uniquely or linearly reflect Yarkovsky drift, and failures like Flora (Sec. 3.2.2) could arise from dynamics/membership rather than proxy inadequacy.

Recommendation: Add a short analytic justification and/or a controlled numerical experiment demonstrating when $\sigma_a(\text{proxy})$ should increase monotonically (approximately linearly) with a Yarkovsky-sensitive proxy. At minimum, in Sec. 2.5 and/or Sec. 3.3: (i) state explicit assumptions (single collisional family, roughly symmetric drift about a center, limited resonance truncation, limited contamination); (ii) discuss how asymmetry/truncation changes σ_a even under normal Yarkovsky drift; and (iii) for one well-behaved family (e.g., Maria or Eos) and one problematic one (Flora), add diagnostics linking departures from linearity to known dynamical structures (e.g., proximity to resonances, one-sided truncation).

2. **Core methodological steps are underspecified, threatening numerical reproducibility and potentially biasing both G and R^2 :** (a) binning and bin-merging (Sec. 2.4), (b) definition/uncertainty of σ_a per bin and the regression weights (Sec. 2.5), and (c) the exact definition of R^2 in the weighted setting (Secs. 2.5, 3.2; Table 1). In particular, “inverse of the variance of σ_a ” is ambiguous (estimator variance vs σ_a^2), and different weighted- R^2 conventions can change values and the occurrence/interpretation of negative R^2 .

Recommendation: Expand Secs. 2.4–2.5 into an algorithmic specification sufficient to reproduce Table 1 and Figs. 1–6: (i) define initial bin edges (min/max bounds, closed/open conventions) and whether bins are equal-width in proxy or otherwise; (ii) provide the exact bin-merging rule (direction, nearest-neighbor tie-breaking, iteration order), and report the final number of bins per family/proxy after merging; (iii) define σ_a precisely (sample vs population SD; proper vs osculating a —see separate issue below) and how $\text{Var}(\sigma_a)$ is estimated (analytic formula or bootstrap); (iv) state the exact weight w_i used; and (v) give the explicit weighted- R^2 formula used (weighted SSE/SST about a weighted mean, etc.). Consider adding brief pseudocode and reporting N per final bin (in captions or supplement).

3. **Linearity is assumed but not tested beyond reporting R^2 , and R^2 alone is used as “proxy efficacy” (Secs. 2.5, 3.2–3.3).** Even if $da/dt \propto 1/D$, σ_a need not be linear in $\log_{10}(1/D)$ (or any log proxy), especially under truncation/asymmetry or

heterogeneous physical properties. With only ~ 10 binned points (often fewer after merging), R^2 can be unstable and can reward overfitting/accidental linearity; Flora indicates clear model failure (Sec. 3.2.2).

Recommendation: In Sec. 2.5 and Results (Secs. 3.2–3.3), test at least two alternative model forms for representative families (e.g., Maria/Eos and Flora): (i) piecewise/segmented linear regression, and (ii) a simple non-linear alternative (e.g., quadratic term or a model in $1/D$ without log). Compare fits using AIC/BIC or cross-validated prediction error (even leave-one-bin-out). Report residual plots (or summarize curvature/heteroscedasticity). If linear ODG remains the headline statistic, justify it explicitly as a robust summary and delineate when it fails (e.g., multi-component families).

4. **Data provenance and dynamical-element choice are unclear, undermining interpretation of σ_a and susceptibility to short-period variations (Secs. 2.1–2.2).** The manuscript does not clearly state whether semimajor axes are proper or osculating, how families are assigned (which HCM/family catalog), and what (if any) interloper filtering is applied. Because σ_a is sensitive to membership contamination and to dynamical environment, these details materially affect the ODG results and proxy comparisons.

Recommendation: In Secs. 2.1–2.2: (i) explicitly state whether a is proper (preferred) or osculating and cite the source catalog; (ii) cite the family-classification source and version/date; (iii) provide a brief description of any interloper mitigation (taxonomy/albedo cuts, “core” memberships) or state explicitly that none was performed; and (iv) add a short robustness check (or an appendix) repeating ODG for at least one family using a stricter membership subset (e.g., “core” members, if available) or with a simple outlier-robust dispersion (see next issue).

5. **Robustness to binning choices, outliers, and interior contamination is not demonstrated (Sec. 2.4, Sec. 3.2–3.3).** Equal-width proxy bins plus merging can induce algorithmic dependence, particularly with skewed size distributions; σ_a is also sensitive to outliers/interlopers. Consequently, the identity of the “best proxy” (Table 1) may not be stable—especially for marginal cases like Eunomia where Proxy_PD wins but differences may be small.

Recommendation: Add a sensitivity/robustness analysis (main text or supplement): (i) equal-count (quantile) bins vs equal-width; (ii) number of bins (e.g., 8/10/12/15); (iii) minimum bin occupancy (e.g., 5 vs 10); and (iv) replace σ_a with a robust alternative such as MAD (scaled to σ) or trimmed SD. Report whether (a) slopes G and (b) the “best proxy” choice in Table 1 change under these variants. A bootstrap over the full pipeline (resample objects within family, recompute bins+fits) would also provide uncertainty on “best proxy” decisions.

6. **Proxy definitions and units/dimensionality are not fully rigorous: taking \log_{10} of dimensional quantities (D , P , $P \cdot D$) is unit-dependent, affecting intercept C and potentially confusing cross-study comparability (Sec. 2.3).**

Units for D and P are not consistently stated, and the dimensional interpretation of G is not provided (Secs. 2.3, 2.5, 3.2.2). Additionally, the physical motivation for Proxy_P and Proxy_PD as “Yarkovsky-sensitive” is only heuristic; the dependence on spin rate enters through the thermal parameter and obliquity, not simply $1/P$.

Recommendation: In Sec. 2.3: (i) redefine proxies in dimensionless form (e.g., $\log_{10}(D_0/D)$, $\log_{10}(P_0/P)$, $\log_{10}((P_0D_0)/(PD))$) or explicitly fix units and note that unit changes shift C but not G ; (ii) state the units used for D and P and the implied units/interpretation of G ; (iii) add 2–4 citations to Yarkovsky theory describing rotation-rate dependence via the thermal parameter and clarify that Proxy_P/Proxy_PD are heuristic. Optionally, add an exploratory alternative spin-inclusive proxy closer to theory (e.g., involving ω or $\sqrt{\omega}$) and report whether conclusions change.

7. **Selection effects from requiring both diameters and spin periods are not quantified, limiting interpretation of poor performance for spin-based proxies (Secs. 2.1–2.3, 3.1, 3.3, 3.5).** The retained subset is likely biased toward larger/brighter objects and may under-sample the smallest (most drifted) members, potentially flattening σ_a -proxy relations and distorting comparisons among proxies.

Recommendation: In Sec. 3.1 (or a new subsection): report, per family, (i) total cataloged members vs those with D , vs those with P , vs those with both (final sample); (ii) distributions of D and P for retained vs full membership (where available); and (iii) a brief discussion of how missing small objects could bias $\sigma_a(\text{proxy})$. Explicitly scope the conclusion “spin period is ineffective” to the observed subset and note that incompleteness in P may dominate Proxy_P/Proxy_PD performance.

8. **Benchmarking and claims of robustness/universality are currently stronger than what is demonstrated (Secs. 1, 3.5, 4.1, 4.4).** ODG avoids specifying a center, but it still assumes a coherent single-collision family with a monotonic dispersion-proxy relation; Flora and the excluded Nysa-Polana case suggest important limitations. The paper also does not directly compare ODG to traditional V-shape/envelope methods, so the claimed practical advantage remains largely qualitative.

Recommendation: Temper “universal/robust” phrasing in Sec. 1 and Discussion (Secs. 3.5, 4.1, 4.4), explicitly stating applicability conditions and citing Flora as a boundary case (Sec. 3.2.2). Add one direct benchmark: for at least one family, compare ODG outputs to a standard V-shape boundary fit (or literature values), and/or show via a Monte Carlo that ODG is stable under plausible center uncertainties that would affect apex-based methods.

9. **Uncertainty reporting is incomplete for the main fitted quantities and the age-gradient correlation (Secs. 2.5, 2.6.2, 3.4; Table 1).** Gradients are reported without uncertainties in Table 1 despite mentioning σ_G ; confidence in whether G

differs from zero is unclear for weak fits. The Spearman test uses $N = 6$ families and ignores uncertainties in both ages and G . Family age values are not consistently cited with uncertainties (Secs. 2.2, 3.2.1, 3.4; Table 1).

Recommendation: Add σ_G (or 95% CI) to Table 1 and describe how it is computed under the chosen weighting (Sec. 2.5). In Sec. 3.2.2, comment on which slopes are significantly non-zero. For the age- G analysis (Secs. 2.6.2, 3.4): provide explicit literature citations and uncertainty ranges for each family age in/near Table 1, and propagate uncertainties via a simple Monte Carlo sampling of ages (and optionally G) to give a confidence interval for Spearman ρ . Also clarify which proxy's G is used per family and provide a sensitivity test excluding families with very low/negative R^2 .

10. **Inconsistency in Nysa-Polana exclusion rationale between Methods and Results (Secs. 2.2, 3.1) and inconsistent Spearman reporting across sections (Abstract; Sec. 3.4; Sec. 4.3).** These inconsistencies impair trust in the selection function and secondary statistics.

Recommendation: Make Secs. 2.2 and 3.1 consistent by explicitly stating: (i) whether Nysa-Polana meets the ≥ 100 -member completeness criterion, (ii) how many objects have D and P , and (iii) whether structural complexity is the primary reason for exclusion (with 1–2 citations). Re-audit and harmonize the reported Spearman (ρ , p) values across Abstract, Sec. 3.4, and Sec. 4.3, noting any differences in sample definition if applicable.

Minor issues

1. Figures 1–6 are difficult to read (fonts/axis labels/ticks), and statistical information needed to interpret fits is often missing (error bars on binned σ_a , confidence bands, N per bin, final bin counts after merging). Overplotting in raw scatter panels obscures structure (Figs. 1–6).

Recommendation: Increase figure and font sizes and/or split multi-panel figures. Put explicit proxy formulas and units on axes (e.g., $\log_{10}(D_0/D [\text{km}])$). Add error bars for binned σ_a (SE or bootstrap CI), show 95% fit bands, and annotate bin counts (N_i) and number of merged bins in captions. Use alpha blending or hexbin/density plots to reduce overplotting.

2. Notation is not fully standardized across text/figures (Proxy_D vs ProxyD; $R2$ vs R^2 ; occasional σ_e vs σ_a references), and regression details (weighting, intercept handling) are sometimes omitted in figure captions (Secs. 2.3–3.5; Fig. captions).

Recommendation: Adopt a single notation convention (Proxy_D, Proxy_P, Proxy_PD; σ_a ; R^2) throughout. Ensure every figure caption states: element type (proper a), binning method, weighting scheme, and the exact statistics shown (G , R^2 , p -value if used). Correct any σ_e occurrences if they are typos.

3. R^2 interpretation (including negative values) is not explained where first used, which may confuse readers (Sec. 2.5; Table 1; Sec. 3.2.2).

Recommendation: Add a short note in Sec. 2.5 (and/or first Results paragraph) defining the chosen weighted- R^2 formula and explicitly stating that negative R^2 can occur and indicates worse performance than a constant-mean model under that definition.

4. Using the “best proxy per family” gradient in the age- G Spearman test mixes proxies (Proxy_D and Proxy_PD), complicating physical interpretation of an across-family trend (Secs. 2.6.2, 3.4).

Recommendation: In Sec. 3.4, explicitly list which proxy is used for each family and add a companion analysis using a single consistent proxy across all families (e.g., Proxy_D) to see if the qualitative trend persists. If retained as-is, frame the correlation as exploratory and method-dependent.

5. Limited context relative to prior family-age/Yarkovsky techniques (V-shape envelope fitting, apex-based drift estimation) makes it harder to evaluate novelty and regimes of advantage (Sec. 1; Sec. 3.5; Sec. 4.4).

Recommendation: Add a concise paragraph in Sec. 1 and/or Sec. 4.4 summarizing key prior approaches with representative citations, and explicitly position ODG as complementary (dispersion-based, center-free) rather than a universal replacement.

6. Some value/reporting inconsistencies remain (e.g., Koronis Fig. 3 caption threshold ‘ ≤ -0.11 ’ vs Table 1 value -0.1065 ; placeholder ‘Table ??.’) (Fig. 3 caption; Sec. 2.2).

Recommendation: Audit and reconcile all numeric thresholds and cross-references: update Fig. 3 caption to match Table 1 with consistent rounding, and resolve ‘Table ??.’ to the correct table number.

Very minor issues

1. Minor formatting/typographic issues reduce polish: stray orphan heading “###” (Sec. 2.2), inconsistent heading formatting (e.g., a Markdown-like “# 3.3 ...”), inconsistent capitalization/spacing (“Semimajor Axis”, “Spin-{}Period”), and inconsistent rendering of R^2 (Secs. 2.2–4.4).

Recommendation: Perform a final style pass: remove/complete the orphan heading, normalize heading syntax to the journal template, standardize capitalization (“semi-major axis”, “spin period”) and LaTeX spacing, and render R^2 consistently.

2. Author/affiliation block is non-standard (“Denario [Anthropic, Gemini & OpenAI servers. Planet Earth.]”) and may not comply with venue policies (title page).

Recommendation: Replace with a conventional author/affiliation format consistent with the target venue, or confirm that pseudonymous/non-institutional attribution is acceptable and format accordingly.

3. Several long/repetitive sentences, especially repeating that ODG avoids a family center, slightly reduce readability (Secs. 1, 3.5, 4.1).

Recommendation: Tighten language in Introduction/Discussion by consolidating repeated claims and breaking long sentences into shorter, more direct statements.

Key statements and references

- • The Flora family is known for its complex and dynamically diffuse structure, often considered to be a superposition of multiple, overlapping collisional events rather than a single, distinct family, and although it has a relatively young assigned age of 0.5 Gyr, other studies suggest a much older and more complex history, which likely obscures any clear Yarkovsky-driven signature in its present-day orbital distribution.
- *Reference(s):* (none)
- • The Yarkovsky effect's dependence on an asteroid's spin is complex and critically relies on its spin-axis orientation (obliquity), which dictates the direction and magnitude of the thermal force, and the Yarkovsky–O'Keefe–Radzievskii–Paddack (YORP) effect can significantly alter both spin rates and obliquities over time, further complicating the observed spin-state distribution within a family.
- *Reference(s):* (none)
- • Asteroid family ages are derived from various dynamical models and inherently carry significant uncertainties, which were not explicitly propagated in this analysis and can easily obscure genuine physical correlations between family age and Yarkovsky-driven dispersion gradients.
- *Reference(s):* (none)
- • The observed orbital dispersion in asteroid families is not solely a function of Yarkovsky-driven drift and age, but is also influenced by confounding variables such as the initial velocity dispersion from the family-forming impact, subsequent secondary collisional events, and compositional variations that affect thermal properties, all of which contribute to the final observed family structure.
- *Reference(s):* (none)
- • The analysis of the correlation between family age and orbital dispersion gradient is severely limited by the small sample size of only six asteroid families, so detecting a statistically robust trend with such a limited number of data points would require an exceptionally strong underlying correlation, which is not observed here.

- *Reference(s)*: (none)

Mathematical consistency audit

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

Maths relevance: light

The paper’s analytic content is primarily definitional and statistical: it defines three logarithmic ‘Yarkovsky-sensitive’ proxies, bins data by proxy, computes within-bin semimajor-axis dispersion σ_a , and fits a (weighted) linear model $\sigma_a = G \cdot (\text{Meanproxy}) + C$, using R^2 to compare proxy efficacy and Spearman ρ to relate gradients to family ages. There are no multi-step symbolic derivations beyond these definitions.

Checked items

1. **✘ ProxyD definition (diameter-only)** (Sec. 2.3, p. 3 (also reiterated Sec. 3.2, p. 4))
 - **Claim:** Defines diameter proxy as $\text{ProxyD} = \log_{10}(1/\text{Diameter})$.
 - **Checks:** symbol/definition consistency, dimensional/unit consistency
 - **Verdict:** FAIL; confidence: high; impact: critical
 - **Assumptions/inputs:** Diameter is a positive real-valued physical quantity with units (not specified in the paper)., \log_{10} is applied directly to the reciprocal.
 - **Notes:** A logarithm requires a dimensionless argument; $1/\text{Diameter}$ has physical units. Unless a reference scale or unit convention is explicitly built into the definition, **ProxyD** is unit-dependent up to an additive constant, undermining the mathematical well-definedness and interpretability of regression coefficients.
2. **✘ ProxyP definition (spin-period-only)** (Sec. 2.3, p. 3 (also reiterated Sec. 3.2, p. 4))
 - **Claim:** Defines spin-period proxy as $\text{ProxyP} = \log_{10}(1/\text{Spin Period})$.
 - **Checks:** symbol/definition consistency, dimensional/unit consistency
 - **Verdict:** FAIL; confidence: high; impact: critical
 - **Assumptions/inputs:** Spin Period is positive and has time units (not specified).
 - **Notes:** Same issue as ProxyD: \log_{10} is applied to a dimensional quantity ($1/\text{time}$). Needs explicit normalization (e.g., P_0/P) or a stated fixed unit convention with discussion of unit dependence.
3. **✘ ProxyPD definition (combined) and algebraic compatibility** (Sec. 2.3, p. 3 (also reiterated Sec. 3.2, p. 4))

- **Claim:** Defines combined proxy as $\text{ProxyPD} = \log_{10}(1/(\text{Spin Period} \times \text{Diameter}))$.
- **Checks:** algebraic consistency, symbol/definition consistency, dimensional/unit consistency
- **Verdict:** FAIL; confidence: high; impact: critical
- **Assumptions/inputs:** Spin Period and Diameter are positive., Same unit conventions are used wherever the proxies are computed.
- **Notes:** Algebra is fine: $\log_{10}(1/(PD)) = -\log_{10}(PD)$ and equals $\log_{10}(1/P) + \log_{10}(1/D)$ if the same base and compatible unit conventions are used. However, the argument is again dimensional ($1/(\text{time} \cdot \text{length})$), so the proxy is not mathematically well-defined without normalization.

4. ✓ **Binned x-variable definition (Mean_Proxy)** (Sec. 2.4, p. 3; Sec. 2.5, p. 3–4)

- **Claim:** Uses the arithmetic mean of the proxy values within each bin as the representative x -coordinate for regression.
- **Checks:** definition consistency, logic of construction
- **Verdict:** PASS; confidence: high; impact: minor
- **Assumptions/inputs:** Proxy values are numeric scalars (after the log transform)., Bins may be merged; after merging, mean is recomputed.
- **Notes:** Given proxy is defined per asteroid, taking the arithmetic mean within a bin is internally consistent with the described regression pipeline, including after bin merges.

5. △ **Dispersion metric σ_a definition** (Sec. 2.4, p. 3; referenced throughout Sec. 3.2, pp. 4–5)

- **Claim:** Defines σ_a as the standard deviation of semimajor axis values within each proxy bin.
- **Checks:** symbol/definition consistency, dimensional/unit consistency
- **Verdict:** UNCERTAIN; confidence: medium; impact: moderate
- **Assumptions/inputs:** Semimajor axis a is measured in a consistent unit (likely AU, not explicitly stated)., Standard deviation computation convention (N vs $N - 1$) is not specified.
- **Notes:** Conceptually consistent (σ_a has units of semimajor axis). But the estimator convention is unspecified, which matters for any claim about the variance/uncertainty of σ_a used in weighting.

6. △ **Linear model and ODG definition** (Sec. 2.5, p. 3–4; reiterated Sec. 3.2, p. 5)

- **Claim:** Fits $\sigma_a = G \cdot (\text{Meanproxy}) + C$, interpreting G as the Orbital Dispersion Gradient (ODG).
- **Checks:** algebraic/notation consistency, dimensional/unit consistency
- **Verdict:** UNCERTAIN; confidence: medium; impact: moderate

- **Assumptions/inputs:** σ_a is treated as the dependent variable; **Meanproxy** as independent variable., Linear relationship is used as a first-order approximation.
- **Notes:** The linear form is internally consistent. However, because **Meanproxy** is defined via log of dimensional quantities, the implied units/meaning of G and C are unit-convention-dependent, which compromises the analytic interpretability of ODG as a physical gradient.

7. \triangle **Regression weighting by inverse variance of σ_a** (Sec. 2.5, p. 4)

- **Claim:** Weights are set as the inverse of the variance of σ_a in each bin to emphasize more precise σ_a estimates.
- **Checks:** definition consistency, logic of derivation
- **Verdict:** UNCERTAIN; confidence: high; impact: critical
- **Assumptions/inputs:** A computable per-bin variance associated with σ_a exists and is used., The intended meaning is estimator uncertainty of σ_a .
- **Notes:** Key formula for weights is missing/ambiguous. ‘Variance of σ_a ’ could mean $\text{Var}(\sigma_a)$ (estimator uncertainty) or σ_a^2 (the within-bin variance of a). The stated rationale (‘more precise estimates of σ_a ’) suggests $\text{Var}(\sigma_a)$, but no method to compute it is provided, preventing verification of the regression setup central to reported G and R^2 .

8. \triangle **Use and interpretation of negative R^2** (Sec. 3.2.2, p. 5; Table 1, p. 6)

- **Claim:** Negative R^2 values indicate no linear relationship and performance worse than predicting by the mean σ_a .
- **Checks:** logical consistency, definition consistency
- **Verdict:** UNCERTAIN; confidence: medium; impact: minor
- **Assumptions/inputs:** R^2 is computed in a way that permits negative values (e.g., $1 - \text{SSE}/\text{SST}$ under some weighting conventions).
- **Notes:** The interpretation is consistent with a common R^2 definition where SSE can exceed SST. But the exact R^2 formula under weighting is not given, so it is not possible to verify that negative values are mathematically expected under their chosen computation.

9. \checkmark **Best-proxy selection criterion** (Sec. 2.6.1, p. 4; Sec. 3.2.1–3.3, pp. 5–8; Table 1, p. 6)

- **Claim:** Selects the ‘best proxy’ for each family as the one with highest R^2 .
- **Checks:** logic of comparison, definition consistency
- **Verdict:** PASS; confidence: medium; impact: minor
- **Assumptions/inputs:** R^2 values are comparable across the three regressions for a given family.

- **Notes:** As a purely internal decision rule, ‘highest R^2 wins’ is consistent. Comparability can be affected by differing weighting conventions, but that is a methodological choice rather than an internal algebraic contradiction.

10. ✓ **Spearman correlation setup (age vs gradient)** (Sec. 2.6.2, p. 4; Sec. 3.4, p. 9)

- **Claim:** Computes Spearman ρ between family age and G from the best-performing proxy per family.
- **Checks:** definition consistency, logic of construction
- **Verdict:** PASS; confidence: high; impact: minor
- **Assumptions/inputs:** Each family contributes one (age, G) pair., Ranking is well-defined even with small $n = 6$.
- **Notes:** Statistical construction is internally consistent. (Numeric correctness of ρ and p -value is out of scope.)

Limitations

- The PDF contains no explicit formulas for (i) the variance/uncertainty of σ_a used for weights, (ii) the exact weighted-regression procedure, or (iii) the exact R^2 definition under weighting; these missing definitions prevent a complete analytic verification of the regression mathematics.
- Units for Diameter, Spin Period, and semimajor axis are not explicitly stated in the mathematical definitions, which is central to assessing dimensional consistency of the log proxies and interpretability of ODG.
- No intermediate algebraic steps are provided for regression parameter estimation (G , C , σ_G), so only definitional/consistency checks are possible.

Numerical results audit

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

Out of 14 numeric checks, 12 passed and 2 failed. Passes include repeated dataset size consistency, multiple cross-section value matches (Abstract/Table 1), Table 1 internal argmax/lookup consistency, and caption-to-table rounding checks for several figures. Failures involve (1) inconsistent repeated reporting for Spearman ρ and p across sections and (2) an inequality/threshold mismatch in the Koronis Figure 3 caption relative to Table 1.

Checked items

1. ✓ **C1_dataset_count_consistency** (p.1 Abstract; p.2 §2.1; p.4 §3.1; p.10 §4.2)
 - **Claim:** Paper repeatedly states the cleaned master dataset contains **16,364** asteroids with complete records.
 - **Checks:** repeated-constant-equality
 - **Verdict:** PASS
 - **Notes:** All repeated values match expected **16364**.

2. ✓ **C2_table1_best_proxy_is_argmax_r2** (p.6 Table 1)
 - **Claim:** For each family, 'Best Proxy' should correspond to the maximum R^2 among ProxyD/ProxyP/ProxyPD, and 'Best R2' should equal that maximum.
 - **Checks:** argmax-consistency
 - **Verdict:** PASS
 - **Notes:** Checked BestProxy corresponds to argmax R^2 and BestR2 equals that max (ties allowed).

3. ✓ **C3_table1_best_gradient_matches_best_proxy** (p.6 Table 1)
 - **Claim:** For each family, 'Best Gradient' should equal the gradient G corresponding to the selected 'Best Proxy'.
 - **Checks:** lookup-consistency
 - **Verdict:** PASS
 - **Notes:** Checked BestGradient equals gradient corresponding to BestProxy.

4. ✓ **C4_abstract_maria_r2_matches_table** (p.1 Abstract vs p.6 Table 1)
 - **Claim:** Abstract claims Maria family has R^2 up to **0.9353** for the diameter-only proxy; Table 1 lists Maria ProxyD $R^2 = 0.9353$.
 - **Checks:** cross-section-value-match
 - **Verdict:** PASS
 - **Notes:** Compared two values across sections within tolerance.

5. ✓ **C5_abstract_eunomia_r2_matches_table** (p.1 Abstract vs p.6 Table 1)
 - **Claim:** Abstract reports Eunomia combined proxy effectiveness $R^2 = 0.4366$; Table 1 Eunomia ProxyPD $R^2 = 0.4366$.
 - **Checks:** cross-section-value-match
 - **Verdict:** PASS
 - **Notes:** Compared two values across sections within tolerance.

6. ✗ **C6_spearman_rho_p_repeated_consistency** (p.1 Abstract; p.9 §3.4; p.11 §4.3)
 - **Claim:** Spearman correlation is reported as $\rho = 0.3714$ and p -value = 0.4685 in multiple sections; values should match exactly across the paper.
 - **Checks:** repeated-constant-equality
 - **Verdict:** FAIL
 - **Notes:** Repeated values are not consistent and/or do not match expected constant.

7. ✓ **C7_four_of_six_best_proxyD_count** (p.8 §3.3; corroborate with p.6 Table 1)

- **Claim:** Text claims ProxyD is best in four out of six families (Maria, Eos, Koronis, Vesta). Verify count and membership list from Table 1 'Best Proxy'.
 - **Checks:** count-from-table
 - **Verdict:** PASS
 - **Notes:** Counted families with BestProxy==ProxyD and compared to claimed count and named set.
8. ✓ **C8_maria_variance_explained_percent** (p.5 §3.2.2 (Maria bullet); p.6 Table 1)
- **Claim:** Maria bullet says $R^2 = 0.9353$ implies $\sim 93.5\%$ variance explained.
 - **Checks:** percentage-from-r2
 - **Verdict:** PASS
 - **Notes:** Computed $100 \times R^2$ and compared to claimed percent within tolerance.
9. ✓ **C9_figures_r2_rounding_matches_table_maria** (p.6 Figure 1 caption vs p.6 Table 1)
- **Claim:** Figure 1 caption reports Maria R^2 values rounded (ProxyD= 0.94, ProxyP= 0.27, ProxyPD= 0.04); Table 1 provides 0.9353, 0.2716, 0.0419.
 - **Checks:** rounding-consistency
 - **Verdict:** PASS
 - **Notes:** Checked caption equals table value rounded to 2 decimals (exact match).
10. ✓ **C10_figures_r2_rounding_matches_table_eos** (p.6 Figure 2 caption vs p.6 Table 1)
- **Claim:** Figure 2 caption reports Eos R^2 values rounded (ProxyD= 0.80, ProxyP= -0.09, ProxyPD= 0.26); Table 1 provides 0.7988, -0.0916, 0.2645.
 - **Checks:** rounding-consistency
 - **Verdict:** PASS
 - **Notes:** Checked caption equals table value rounded to 2 decimals (exact match).
11. ✗ **C11_figures_r2_rounding_matches_table_koronis** (p.7 Figure 3 caption vs p.6 Table 1)
- **Claim:** Figure 3 caption reports Koronis R^2 rounded (ProxyD= 0.61; ProxyP/ProxyPD ≤ -0.11). Table 1 gives ProxyD= 0.6115, ProxyP= -0.2130, ProxyPD= -0.1065.
 - **Checks:** rounding-and-inequality-consistency
 - **Verdict:** FAIL

- **Notes:** Rounding for ProxyD matched, and ProxyP met the ≤ -0.11 threshold, but ProxyPD did not: -0.1065 is greater than -0.11 .
12. ✓ **C12_figures_r2_rounding_matches_table_eunomia** (p.7 Figure 4 caption vs p.6 Table 1)
- **Claim:** Figure 4 caption reports Eunomia ProxyPD $R^2 = 0.44$; Table 1 gives 0.4366 .
 - **Checks:** rounding-consistency
 - **Verdict:** PASS
 - **Notes:** Checked caption equals table value rounded to 2 decimals (exact match).
13. ✓ **C13_figures_r2_rounding_matches_table_vesta** (p.8 Figure 5 caption vs p.6 Table 1)
- **Claim:** Figure 5 caption reports Vesta ProxyD $R^2 = 0.21$; Table 1 gives 0.2122 .
 - **Checks:** rounding-consistency
 - **Verdict:** PASS
 - **Notes:** Checked caption equals table value rounded to 2 decimals (exact match).
14. ✓ **C14_figures_flora_negative_r2_all_proxies** (p.9 Figure 6 caption vs p.6 Table 1)
- **Claim:** Figure 6 caption says Flora has negative R^2 values for all proxies; Table 1 lists -0.1313 , -0.0380 , -0.0031 .
 - **Checks:** sign-check
 - **Verdict:** PASS
 - **Notes:** Checked all provided values are strictly negative.

Limitations

- Audit uses only numeric values explicitly present in the provided PDF text/tables/captions; no underlying datasets are available to recompute regressions or bin statistics.
- Figure-based checks are limited to numbers written in figure captions; no values are extracted from plot graphics/pixels.
- Some statements are qualitative (e.g., 'largest gradient among all families') but can be checked only when all needed numeric comparators are explicitly tabulated.
- Spearman correlation (ρ , p) was not recomputed from primary data vectors; only internal repeated-value consistency was checked, and that consistency check failed (p.1 Abstract; p.9 §3.4; p.11 §4.3).

- Regression-weighting and binning/merging numeric implications could not be validated because necessary bin-level variances and counts are not explicitly provided (p.4 §2.5; p.3 §2.4).