Skepthical review: *Constraining Satellite Galaxy Radial Profiles with a Mass-Conditioned Spatial Point Process Model*

Summary

The paper proposes a likelihood-based, mass-conditioned spatial point-process framework for modeling the 3D distribution of satellite galaxies around dark-matter halos, framed as a Neyman–Scott parent–offspring process (Sec. 2.2). The mean satellite occupation is modeled with a power-law HOD in halo mass (Eq. (1)), and satellite positions are modeled with an isotropic exponential radial kernel scaled by halo virial radius (Eq. (2)), with a "concentration" parameter that is either constant or mass-dependent (Eq. (3)). Parameters are inferred via maximum likelihood on ten synthetic 3D catalogs (≈ 68k satellites in ≈ 20k halos, with $M_{\rm vir} \geq 10^{13}\ M_\odot/h$), using a hard radial cut $r < 5$ Mpc$/h$ intended to focus on the 1-halo regime (Sec. 2.1–2.3). The authors report approximate recovery of input HOD parameters with a positive bias in $M_{\rm sat}$ and $\alpha_{\rm sat}$ (Sec. 3.1), strong AIC preference (ΔAIC ≈ 136) for a weak mass dependence of concentration (Sec. 3.2), a luminosity-marked correlation signal consistent with strong 1-halo luminosity segregation and weak 2-halo dependence (Sec. 2.4, Sec. 3.3), and a large residual underprediction beyond the fitting radius (5–10 Mpc$/h$) that highlights the breakdown of an isolated 1-halo approximation (Sec. 2.5, Sec. 3.4). The methodological direction is promising and clearly motivated, but key details of the likelihood, kernel normalization/truncation, mock construction, and uncertainty quantification are currently insufficient for reproducibility and for assessing robustness (including sensitivity to profile choice, the $r < 5$ Mpc$/h$ cut, and dependence assumptions that affect AIC interpretation).

Strengths

- Compelling motivation for moving beyond summary statistics (e.g., 2PCF) to a direct likelihood for galaxy positions via a point-process/HOD formulation (Introduction, Sec. 2.2).
- Clean conceptual decomposition into (i) a mass-dependent occupation model and (ii) a mass-conditioned spatial kernel for satellite positions, which is a natural and extensible modeling template (Eqs. (1)–(3), Sec. 2.2).
- Use of controlled synthetic catalogs to test parameter recovery against known ground truth and to explore where/why the model fails (Sec. 2.1, Sec. 3.1, Sec. 3.4).
- Quantitative model comparison is attempted (AIC; Sec. 2.3, Sec. 3.2) and the nesting of models ($\beta = 0$ gives the constant-concentration model) is clear (Eq. (3)).
- The residual analysis beyond the fitting range (Sec. 2.5, Sec. 3.4) is a useful diagnostic to delineate the 1-halo regime and motivate future inclusion of inter-halo/2-halo structure.

- The manuscript is generally well organized; the main narrative from model → fit → diagnostics is easy to follow (Sec. 1–4).

Major issues

1. **The likelihood function used for inference is not written down explicitly, and critical measure/normalization details are ambiguous (Sec. 2.2–2.3). In particular, Eq. (2) specifies $f(r|M)$ only up to proportionality, without stating whether it is a 3D spatial density $f_{3D}(x|M)$ (units 1/volume) or a 1D radial pdf $p(r|M)$ (units 1/length). The stated "intensity" $\lambda(r|M,\theta) = \langle N_{sat}(M)\rangle \cdot f(r|M)$ is therefore ambiguous for a Poisson point-process likelihood. It is also unclear whether the likelihood includes a Poisson term for satellite counts per halo ($N_{sat}|M$), or whether counts are treated as fixed and only positions are used. Finally, the hard fitting cut $r < 5$ Mpc$/h$ must enter as truncation/renormalization; this is not shown, yet it directly impacts parameter estimates and the interpretation of reported biases (Sec. 3.1).**

   Recommendation: In Sec. 2.2–2.3, add the full conditional log-likelihood used for MLE. A clear option is to write, per halo h with mass M_h, a Poisson term for the count and a product over satellite locations: $\ln L = \Sigma_h[-\mu_h(\theta) + N_h \ln \mu_h(\theta) - \ln(N_h!) + \Sigma_{i=1}^{N_h} \ln f_{trunc}(x_{hi}|M_h,\theta)]$, where $\mu_h(\theta) = \langle N_{sat}(M_h;\theta)\rangle \times P(r < r_{cut}|M_h,\theta)$ if you model counts in the truncated region, and f_{trunc} is the spatial kernel normalized over the truncated domain $r < r_{cut}$. Explicitly define whether f is a 3D isotropic density (with normalization $\int f_{3D}\, d^3x = 1$) or a 1D radial pdf (with the implied $4\pi r^2$ Jacobian when mapping to 3D). State exactly how the $r < 5$ Mpc$/h$ cut is implemented (truncated kernel and/or truncated mean), and clarify independence assumptions (halos independent given the halo catalog; satellites i.i.d. within halos) that justify factorization.

2. **Mock-catalog construction and "ground truth" are under-specified, limiting interpretability of recovery/bias and preventing reproducibility (Sec. 2.1, Sec. 3.1, Sec. 3.3). The paper does not clearly document how halos are generated (N-body vs analytic; box size; cosmology; mass definition; halo finder), the full HOD used to populate galaxies (central occupation, scatter, any non-Poisson satellite dispersion), the true satellite radial profile used in the mock (and whether it matches Eq. (2)), the true $\alpha_c(M)$ (constant vs mass-dependent and its parameters), and how luminosities/marks are assigned (including scatter and whether centrals/satellites differ).**

   Recommendation: Expand Sec. 2.1 (or add an Appendix) to fully specify the mock pipeline: simulation/cosmology details; halo definition (e.g., M_{200c} vs M_{vir}), halo selection ($M_{vir} \geq 10^{13}\, M_\odot/h$), and any boundary conditions; exact central+satellite HOD functional forms and stochasticity (Poisson or otherwise); the injected satellite radial profile and its parameter values (including whether α_c is mass-dependent in the truth); and the luminosity assignment/mark model (functional form, scatter, and whether it is radius-dependent). Provide enough detail that an independent reader can regenerate the catalogs and reproduce Figures in Sec. 3.

3. **Uncertainty quantification, parameter degeneracies, and validation of inference are missing, making it hard to assess the practical significance of the reported trends (Sec. 3.1–3.2, Sec. 4). The paper reports point estimates and an AIC difference, but gives no confidence intervals on ($M_{\rm sat}$, $\alpha_{\rm sat}$, α_0, β), no correlation/degeneracy analysis (e.g., between β and HOD parameters), and no coverage or mock-to-mock variability assessment despite having ten catalogs. This is especially important because the reported HOD bias (~ 0.25 dex in $M_{\rm sat}$) cannot be judged without uncertainties, and because ΔAIC can appear large for large N even if effect sizes are small per satellite.**

   Recommendation: In Sec. 3.1–3.2, report uncertainties and covariances for all fitted parameters (e.g., via the observed Fisher/Hessian at the MLE, bootstrap over halos, or per-mock fits with scatter across the ten catalogs). Add at least one diagnostic of degeneracies (a covariance matrix table or 2D contours in an appendix). For the mass-dependent concentration result, report $\beta \pm \sigma_\beta$ and show 68% bands on $\alpha_c(M)$ vs M in the relevant plot. Consider a simulation-based calibration / coverage check on the synthetic setup (even a lightweight version): generate multiple realizations at the fitted parameters and verify that the true parameters are recovered with correct coverage under the assumed likelihood.

4. **The interpretation and use of AIC as "decisive evidence" for mass-dependent concentration need additional context and robustness checks (Sec. 2.3, Sec. 3.2, Sec. 4). With ≈ 68k satellites, small per-point likelihood improvements can yield very large ΔAIC. Additionally, AIC assumes an effectively independent sample; conditional independence given halos may be violated by substructure, exclusion, or mock-generation correlations, reducing the effective sample size and potentially inflating ΔAIC. Since the models are nested ($\beta = 0$), complementary diagnostics (likelihood-ratio behavior, predictive performance) would strengthen the claim.**

   Recommendation: In Sec. 3.2, report the per-satellite (or per-halo) log-likelihood improvement between models (e.g., $\Delta \ln L/N$) alongside ΔAIC to communicate effect size. Add at least one robustness check: (i) cross-validation across halos (fit on a subset of halos, evaluate predictive log-likelihood on held-out halos), or (ii) out-of-sample validation across mocks (fit on some catalogs, score on others). Optionally report BIC as a sensitivity check (not as a replacement). In Sec. 4, soften language around "decisive" thresholds and explicitly note the dependence of AIC strength on N and on independence assumptions.

5. **The radial kernel choice (exponential) and halo-centering assumptions are not sufficiently justified, and robustness to profile misspecification is not demonstrated (Sec. 2.2, Eq. (2); Sec. 3.2). An exponential satellite profile is non-standard relative to common NFW/gNFW satellite prescriptions in HOD work. If the mocks were generated with the same exponential, the strong AIC preference for β may be partly a "matched-model" effect; if not, kernel mismatch can spuriously induce mass dependence in $\alpha_c(M)$.**

   Recommendation: Clarify in Sec. 2.1–2.2 whether the mock truth uses the same exponential form as Eq. (2). Then add a robustness experiment in Sec. 3.2: either (a) fit an alternative, more standard profile family (e.g., NFW with a concentration–mass relation, or a gNFW) and check whether β remains > 0, or (b) generate additional mocks with an NFW-like truth and fit the exponential model to test whether β is spuriously inferred. Explicitly discuss the mapping between α_c and a physical scale radius (scale length $= R_{\text{vir}}/\alpha_c$) to avoid confusion with standard NFW concentration conventions.

6. **Conditioning vs full Neyman–Scott generative story is conceptually mixed, affecting interpretation of the residual analysis (Sec. 2.2–2.5, Sec. 3.4). The text alternates between (i) conditioning on an observed halo catalog (centers/masses treated as fixed covariates for the satellite likelihood) and (ii) describing halos as a homogeneous Poisson parent process whose failure is diagnosed by residuals at 5–10 Mpc/h. However, if halos are conditioned upon, no parent-process likelihood is being fit, and the enormous underprediction at 5–10 Mpc/h is largely an expected consequence of fitting only a 1-halo component (and/or truncating at 5 Mpc/h), not uniquely a diagnosis of "parent Poisson" failure.**

   Recommendation: Revise Sec. 2.2–2.5 to clearly separate: (A) the conditional satellite model given the halo catalog (the likelihood you optimize), versus (B) any model for the halo (parent) process or 2-halo component (not currently included). Reframe Sec. 2.5 and Sec. 3.4 to state explicitly that the 5–10 Mpc/h residuals demonstrate the breakdown of a pure 1-halo conditional model and motivate adding an explicit 2-halo term and/or a clustered parent process, rather than diagnosing a parent Poisson assumption unless that term is truly included in the likelihood. If you do include a parent term, write it down and justify the homogenous Poisson assumption.

Minor issues

1. The rationale and implications of the $r < 5$ Mpc/h fitting cut are not quantified (Sec. 2.3, Sec. 3.1, Sec. 3.4). Because the kernel is scaled by R_{vir} but the truncation is in fixed Mpc/h, the effective truncation in r/R_{vir} is mass-dependent; this can induce or mask apparent mass trends in $\alpha_c(M)$ and can contribute to HOD bias.

Recommendation: In Sec. 2.3, justify $r < 5$ Mpc$/h$ by summarizing the $R_{\rm vir}$ distribution for halos in the sample and reporting the implied $r_{\rm cut}/R_{\rm vir}$ range across halo masses. Add a brief sensitivity test (e.g., $r_{\rm cut} = 3$ and 7 Mpc$/h$) and report changes in ($M_{\rm sat}$, $\alpha_{\rm sat}$, α_0, β) and in ΔAIC. Explicitly discuss the mass-dependent truncation effect from mixing a fixed $r_{\rm cut}$ with an $R_{\rm vir}$-scaled kernel.

2. The reported HOD parameter bias explanation is plausible but remains qualitative (Sec. 3.1). Alternative contributors (kernel mismatch, truncation handling, selection, finite-volume effects) are not disentangled, and without uncertainties it is unclear whether the bias is statistically significant or practically large.

   Recommendation: Strengthen Sec. 3.1 with at least one quantitative diagnostic: show likelihood slices around the truth, or a small table of recovered parameters versus $r_{\rm cut}$ and/or versus restricted mass ranges. Separate biases that are expected from known selection/truncation from any residual bias that might indicate model misspecification.

3. Marked correlation function definition/estimation is not specified precisely and the interpretation may largely reflect built-in features of the mock HOD (Sec. 2.4, Sec. 3.3). Different marked-correlation estimators exist; without an explicit formula it is hard to verify that $M(r)/M_{\rm shuffled}(r) = 1$ under the null of mark–position independence. Additionally, the observed small-scale luminosity segregation may be a direct consequence of the mock design (e.g., bright centrals at halo centers) rather than a discriminating test of the method.

   Recommendation: In Sec. 2.4, provide the explicit estimator used for $M(r)$ and $M_{\rm shuffled}(r)$: pair weighting, normalization (including whether marks are normalized by $\langle m \rangle$), binning, and number of shuffles. State whether the statistic uses all galaxies or satellites only, and how edges/finite-volume effects are handled. To make Sec. 3.3 more compelling, add a controlled mock or relabeling test (e.g., shuffle luminosities within halos vs across halos; impose/remove a satellite luminosity–radius gradient) to demonstrate that the statistic can distinguish alternative segregation mechanisms beyond the baseline HOD construction.

4. Residual construction and goodness-of-fit reporting are underspecified (Sec. 2.5, Sec. 3.4). It is unclear whether O and E are totals summed over all halos or per-halo averages, what uncertainties are used (Poisson vs bootstrap), and how well the model fits within the fitted domain $r < 5$ Mpc$/h$.

   Recommendation: In Sec. 2.5 and Sec. 3.4, explicitly define O and E (total vs per-halo), provide uncertainty estimates (Poisson and/or bootstrap over halos), and plot standardized residuals across 0–10 Mpc$/h$ with the fitted region marked. Optionally report a simple goodness-of-fit metric (e.g., deviance or χ^2) for $r < 5$ Mpc$/h$ to document whether any systematic deviations exist even inside the fitted regime.

5. The paper's "survey readiness" framing is stronger than warranted by the current idealized setup (Abstract, Sec. 1, Sec. 4). The analysis assumes perfect 3D positions, known halo centers/masses, and no observational systematics (RSD, masks, incompleteness, miscentering, mass-proxy scatter), all of which can strongly affect sensitivity to a small β and to luminosity segregation.

   Recommendation: Temper claims in the Abstract and Sec. 4 and add a focused limitations/future-work paragraph listing key observational effects and how they would enter the likelihood (e.g., redshift-space kernels, survey selection function, marginalizing miscentering, halo-mass uncertainty). Briefly outline a concrete path to incorporate a 2-halo term and/or a clustered parent process for application to real surveys.

Very minor issues

1. Presentation/notation issues reduce clarity in places: inconsistent unit spacing ("Mpc/h" vs "Mpc / h"), HTML entities for inequalities ("<", ">"), small rounding inconsistencies for ΔAIC across text/captions, and occasional stray formatting (e.g., markdown-like headings) (various locations; noted around Sec. 3.2–3.3 and figure captions).

   Recommendation: Proofread for consistency: standardize unit formatting, ensure inequality symbols render correctly, harmonize ΔAIC rounding across all mentions, remove stray formatting artifacts, and ensure parameter notation (M_{sat}, α_{sat}, α_c, α_0, β) is consistent between text and figures.

2. Terminology: calling α_c a "concentration" may confuse readers because in the exponential model it is an inverse scale length (scale radius $= R_{\mathrm{vir}}/\alpha_c$), not an NFW concentration parameter (Sec. 2.2, Eq. (2)).

   Recommendation: Add one sentence near Eq. (2) clarifying the mapping: larger α_c implies a smaller scale length $R_{\mathrm{vir}}/\alpha_c$ and therefore a more centrally concentrated satellite distribution; optionally note how (or whether) this relates to standard NFW concentration.

Mathematical consistency audit

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

Maths relevance: light

The paper contains a small set of core parametric equations (HOD mean occupation, an exponential radial kernel, a mass-dependent concentration scaling, and AIC) and uses them conceptually to motivate an MLE/AIC comparison and diagnostic statistics. The main internal-consistency risk is not algebraic manipulation, but incomplete/ambiguous mathematical specification of the point-process kernel/intensity and the implied likelihood measure/normalization in 3D.

Checked items

1. ✔ **HOD mean satellite occupation power law** (Eq. (1), Sec. 2.2, p.3)
   - **Claim:** The mean number of satellites in a halo of mass M is $\langle N_{\text{sat}}(M)\rangle = (M/M_{\text{sat}})^{\alpha_{\text{sat}}}$.
   - **Checks:** dimensional consistency, notation consistency
   - **Verdict:** PASS; confidence: high; impact: moderate
   - **Assumptions/inputs:** M and M_{sat} have the same mass units, $\langle N_{\text{sat}}(M)\rangle$ is dimensionless and nonnegative for $M > 0$
   - **Notes:** The expression is dimensionally consistent and uses defined parameters M_{sat} and α_{sat}. Distributional assumption for N_{sat} (e.g., Poisson) is implied later but not part of Eq. (1) itself.

2. ✔ **Exponential radial kernel exponent is dimensionless** (Eq. (2), Sec. 2.2, p.3)
   - **Claim:** Satellite radial kernel satisfies $f(r|M) \propto \exp(-\alpha_c(M)\, r/R_{\text{vir}})$.
   - **Checks:** dimensional consistency, sanity/limiting behavior
   - **Verdict:** PASS; confidence: high; impact: minor
   - **Assumptions/inputs:** r and R_{vir} have the same length units, $\alpha_c(M)$ is dimensionless
   - **Notes:** The exponent is dimensionless. Larger α_c increases central concentration as intended. However, normalization/support are not specified (handled in separate items).

3. ⚠ **Kernel normalization and 3D vs 1D radial density ambiguity** (Eq. (2) and surrounding text, Sec. 2.2, p.3)
   - **Claim:** $f(r|M)$ is a 'radial probability density function' used as the spatial kernel for 3D satellite positions.
   - **Checks:** definition consistency, measure/normalization consistency
   - **Verdict:** UNCERTAIN; confidence: medium; impact: critical
   - **Assumptions/inputs:** Satellites are a 3D point process around each halo center, Likelihood is constructed from 3D satellite positions
   - **Notes:** As written, $f(r|M)$ depends only on radius and is given only up to proportionality. For 3D isotropic positions, one must specify whether f is (i) a 3D density per unit volume (requiring normalization under d^3x, and if expressed via r, accounting for the $4\pi r^2$ Jacobian), or (ii) a 1D radial pdf per unit r (normalized under dr). The paper does not specify which. Without this, the implied likelihood and the meaning of $\lambda(r|M)$ are not mathematically well-defined.

4. ⚠ **Intensity function definition** $\lambda(r|M,\theta)$ (Sec. 2.2, after Eq. (2), p.3)

   - **Claim:** The full intensity function is $\lambda(r|M,\theta) = \langle N_{\text{sat}}(M)\rangle \cdot f(r|M)$.
   - **Checks:** dimensional/units consistency, definition consistency
   - **Verdict:** UNCERTAIN; confidence: medium; impact: critical
   - **Assumptions/inputs:** λ is the Poisson point-process intensity for 3D satellite positions, $\langle N_{\text{sat}}(M)\rangle$ is the expected total number of satellites per halo
   - **Notes:** For a 3D Poisson process, intensity has units 'number per volume'. If f is a 1D radial pdf 'per length', then λ as written would be 'number per length' and not a 3D intensity. If f is intended as a 3D density per volume, it must be normalized under d^3x and explicitly stated. This affects the analytic form of the likelihood (product of intensities evaluated at points times $\exp(-\int \lambda\, d^3x)$).

5. ✔ **Mass-dependent concentration scaling** (Eq. (3), Sec. 2.3, p.4)

   - **Claim:** Concentration varies with mass as $\alpha_c(M) = \alpha_0 \left(\frac{M}{M_{\text{pivot}}}\right)^{\beta}$.
   - **Checks:** dimensional consistency, model nesting check
   - **Verdict:** PASS; confidence: high; impact: moderate
   - **Assumptions/inputs:** α_0 is dimensionless, M_{pivot} has mass units, β is dimensionless exponent
   - **Notes:** Dimensionless and reduces to the constant-concentration model when $\beta = 0$, as required for a nested comparison.

6. ✔ **AIC definition and algebraic consistency with reported values** (Eq. (4), Sec. 2.3 p.4; calculations in Sec. 3.2 p.6)

   - **Claim:** $\text{AIC} = 2k - 2\ln(L)$; reported AIC values follow from reported $\ln(L)$ for the two models.
   - **Checks:** algebra check, definition consistency
   - **Verdict:** PASS; confidence: high; impact: minor
   - **Assumptions/inputs:** Constant-α_c model has $k = 3$ parameters (M_{sat}, α_{sat}, α_0), Mass-dependent model has $k = 4$ parameters (M_{sat}, α_{sat}, α_0, β), $\ln(L)$ values are the maximized log-likelihoods
   - **Notes:** Using Eq. (4): for $\ln(L) = 340132.53$ and $k = 3$, $\text{AIC} = 6 - 680265.06 = -680259.06$ (matches). For $\ln(L) = 340201.60$ and $k = 4$, $\text{AIC} = 8 - 680403.20 = -680395.20$ (matches). ΔAIC computed as 136.14 is consistent with these AIC values.

7. ⚠ **Hard radial cut and likelihood truncation handling** (Sec. 2.3, p.3–4)

   - **Claim:** Likelihood includes only satellites with $r < 5\ \text{Mpc}/h$ to isolate 1-halo term.

- **Checks:** likelihood specification completeness, normalization under truncation
- **Verdict:** UNCERTAIN; confidence: low; impact: moderate
- **Assumptions/inputs:** A Poisson point-process likelihood is used, Data are truncated/censored by a selection $r < 5\ \mathrm{Mpc}/h$
- **Notes:** Analytically, truncating observed points generally requires either truncating the observation window in the integral term and/or renormalizing the kernel if interpreting it as a conditional density within the window. The paper does not show the likelihood, so it is not possible to verify that the model intensity and $\langle N_{\mathrm{sat}}(M)\rangle$ are treated consistently under the imposed cut.

8. ⚠ **Marked correlation function ratio behavior under shuffling** (Sec. 2.4, p.4; interpretation in Sec. 3.3, p.7–8)

    - **Claim:** $M(r)/M_{\mathrm{shuffled}}(r) > 1$ indicates luminosity segregation; ≈ 1 indicates no segregation.
    - **Checks:** definition consistency, sanity check (null expectation)
    - **Verdict:** UNCERTAIN; confidence: medium; impact: minor
    - **Assumptions/inputs:** $M(r)$ is computed with a consistent pair-selection/binned-separation rule in both observed and shuffled catalogs, Shuffling preserves the marginal mark distribution while breaking mark-position dependence
    - **Notes:** The qualitative logic is consistent, but without an explicit estimator for $M(r)$ (including whether it is normalized by pair counts and/or mean mark), it is not possible to guarantee analytically that the shuffled ratio equals 1 under the null for the chosen estimator.

9. ✔ **Fractional residual definition** (Sec. 2.5, p.4; Sec. 3.4, p.8)

    - **Claim:** Fractional residual is $(O - E)/E$ comparing observed vs expected counts in $5 < r < 10\ \mathrm{Mpc}/h$.
    - **Checks:** algebra/definition check
    - **Verdict:** PASS; confidence: high; impact: minor
    - **Assumptions/inputs:** $E > 0$
    - **Notes:** Residual definition is standard and internally consistent with the later narrative (large positive residual when $O \gg E$).

Limitations

- The audit is restricted to the provided 9-page text extraction; no additional appendices, supplementary material, or full likelihood expressions are available to verify omitted derivation steps.

- Figures are referenced but their underlying mathematical definitions/fit functions are not fully specified in the text; only equations explicitly present in the extracted pages were audited.
- Numerical values (e.g., best-fit parameters, log-likelihood magnitudes, residual counts) were not validated beyond simple algebraic consistency with stated formulas, per scope.

Numerical results audit

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

10 numeric checks were run: 9 PASS and 1 UNCERTAIN. Key model-selection quantities (AIC, ΔAIC, $\ln(L)$ differences) are arithmetically consistent, including implied integer parameter counts for the compared models. One cross-document consistency check (pivot mass usage) could not be completed from the available extracted text.

Checked items

1. ✔ **C1** (Page 6, Section 3.2 (model selection paragraph))
    - **Claim:** Constant-α_c model: $\ln(L) = 340,132.53$, corresponding to an AIC of $-680,259.06$.
    - **Checks:** AIC recomputation from log-likelihood and parameter count
    - **Verdict:** PASS
    - **Notes:** AIC matches recomputation; implied parameter count $k = 3$ is exactly integer and as expected.
2. ✔ **C2** (Page 6, Section 3.2 (model selection paragraph))
    - **Claim:** Mass-dependent model: $\ln(L) = 340,201.60$, resulting in a lower AIC of $-680,395.20$.
    - **Checks:** AIC recomputation from log-likelihood and parameter count
    - **Verdict:** PASS
    - **Notes:** AIC matches recomputation; implied parameter count $k = 4$ is exactly integer and as expected.
3. ✔ **C3** (Page 6, Section 3.2 (model selection paragraph))
    - **Claim:** The difference, ΔAIC = AIC$\mathrm{const}-AIC\mathrm{mass-dep}=136.14$.
    - **Checks:** difference of reported AICs
    - **Verdict:** PASS
    - **Notes:** Computed ΔAIC from the two AIC values equals 136.14 up to floating-point rounding.
4. ✔ **C4** (Page 6, Section 3.2 (model selection paragraph))

- **Claim:** Log-likelihood improvement from constant to mass-dependent model is implied by the two $\ln(L)$ values.
- **Checks:** log-likelihood difference and its consistency with AIC change given Δk
- **Verdict:** PASS
- **Notes:** With $\Delta k = 1$, $d \ln L = 69.07$ implies ΔAIC$= 136.14$, matching the reported value.

5. ✔ **C5** (Page 8, Section 3.4 (breakdown of 1-halo model paragraph))

   - **Claim:** Observed number of galaxies (362) far exceeds the model prediction (5.6) in $5 < r < 10$ Mpc$/h$.
   - **Checks:** ratio and fractional residual recomputation
   - **Verdict:** PASS
   - **Notes:** Quantification from the provided numbers: $O/E = 64.642857...$, fractional residual $(O - E)/E = 63.642857...$. No explicit target value was stated to compare against.

6. ⚠ **C6** (Page 4, Section 2.3 (pivot mass definition) and Eq. (3))

   - **Claim:** Pivot mass fixed to $M_{\rm pivot} = 10^{13}\ M_{\odot}/h$.
   - **Checks:** constant consistency across document
   - **Verdict:** UNCERTAIN
   - **Notes:** Text-wide consistency cannot be checked from the provided payload because only one $M_{\rm pivot}$ occurrence/value was available.

7. ✔ **C7** (Page 3, Section 2.1 (dataset description))

   - **Claim:** Aggregated dataset comprises over $68,000$ satellite galaxies hosted by approximately $20,000$ halos.
   - **Checks:** sanity ratio computation (satellites per halo) from explicit counts
   - **Verdict:** PASS
   - **Notes:** Implied average satellites per halo ≈ 3.4 using $68,000/20,000$; no additional totals were provided for cross-validation.

8. ✔ **C8** (Page 6, Section 3.1 (HOD parameter recovery) vs Page 6, Section 3.2 (best-fit preferred model))

   - **Claim:** Recovered $M_{\rm sat}$ differs slightly between baseline and preferred model: 13.253 vs 13.252 ($\log_{10}$ units).
   - **Checks:** cross-section numeric consistency / rounding check
   - **Verdict:** PASS
   - **Notes:** $\Delta \log_{10} = 0.001$; linear ratio $= 10^{0.001} = 1.002305...$, i.e., $\approx 0.2305\%$ difference.

9. ✔ **C9** (Page 6, Section 3.1 (HOD parameter recovery) vs Page 6, Section 3.2 (best-fit preferred model))
    - **Claim:** Recovered $\alpha_{\rm sat}$ differs slightly between baseline and preferred model: $\mathbf{1.103}$ vs $\mathbf{1.102}$.
    - **Checks:** cross-section numeric consistency / rounding check
    - **Verdict:** PASS
    - **Notes:** $\Delta\alpha_{\rm sat} = \mathbf{0.001}$, consistent with a rounding-level difference.
10. ✔ **C10** (Page 6 (text) vs Page 9 (Conclusions) and Page 7 (Figure 2 caption))
    - **Claim:** ΔAIC is reported as $\mathbf{136.14}$ (Section 3.2), $\approx \mathbf{136}$ (Figure 2 caption), and $\mathbf{136.13}$ (Conclusions).
    - **Checks:** repeated-constant consistency
    - **Verdict:** PASS
    - **Notes:** ΔAIC computed from AICs is $\mathbf{136.14}$; $\mathbf{136.13}$ is consistent within rounding, and "$\approx \mathbf{136}$" is a coarse approximation.

Limitations

- Only parsed PDF text was available; no access to underlying synthetic catalogs, likelihood code, or tabulated numeric outputs.
- No numeric values were extracted from figures/plots beyond what is explicitly stated in the text/captions, per instruction to avoid plot-pixel extraction.
- Some quantities are approximate qualifiers ("over", "approximately", "≈"), limiting strict equality checks; only coarse sanity checks are possible for those.
- Several claims require external data/code to validate (catalog counts and selection thresholds; correct application of radial cuts), and figure-based significance/consistency claims cannot be checked without numeric series/uncertainties.