

JWST AND THE PROBLEM OF PREMATURELY-FORMED MASSIVE GALAXIES: PRE-MISSION PREDICTIONS, EARLY OBSERVATIONS, AND THE STRUCTURE OF THE EMERGING TENSION WITH Λ CDM

Christopher Davis 

University of Sydney
Sydney, Australia
E-mail: crty7281@uni.sydney.edu.au

Received: 18.07.2023. Approved: 28.10.2023.

Original Scientific Article

DOI: <https://doi.org/10.65932/CAR-2023-1-4> UDC: 524.7:524.8

Abstract: The James Webb Space Telescope, operational from July 2022, was designed to test Λ CDM predictions for first-galaxy formation in the $z = 8$ to $z = 20$ window. Within six months, NIRCcam imaging from the SMACS 0723, GLASS, and CEERS Early Release programmes returned an unexpected population of luminous galaxy candidates at $z \approx 10$ -16 with inferred stellar masses of 10^9 - $10^{10} M_{\odot}$ at cosmic times of 300-500 Myr after the Big Bang. The Naidu and colleagues (2022) GLASS- z_{10}/z_{12} discovery, the Castellano and colleagues (2022) $z \approx 9$ -15 sample, and the converging Harikane and colleagues (2022) and Donnan and colleagues (2022) UV luminosity function analyses together established that the bright-end number density at $z > 10$ exceeds pre-JWST Behroozi-Silk (2018) and Behroozi-UniverseMachine (2019) predictions by approximately 0.5-1 order of magnitude. Whether this excess reflects genuine over-formation of massive galaxies, contamination by lower-redshift interlopers, calibration systematics, or a Λ CDM breakdown remained unresolved at the December 2022 boundary of this review. I propose, as the original contribution, the Galaxy Formation Tension Index (GFTI), a normalised composite metric on $[0,1]$ integrating five performance dimensions (observed-to-predicted number-density ratio, inferred star formation efficiency, stellar-mass-density ratio, UV-bright-fraction at $z > 10$, cross-survey consistency). Applied to the December 2022 dataset, GFTI returns approximately 0.55 — the “significant tension” tier, well below the 0.75 refutation threshold.

Keywords: *James Webb Space Telescope, high-redshift galaxies, UV luminosity function, Λ CDM, cosmic dawn, reionization, galaxy formation, Behroozi-Silk constraint, prematurely-formed massive galaxies.*

INTRODUCTION

The standard cosmological model — flat Λ CDM with adiabatic Gaussian initial conditions, cold dark matter, and inflation-set primordial perturbations — predicts that the first galaxies form hierarchically through the merging of dark-matter halos, that the cosmic star-formation rate density rises from the cosmic dark ages through $z \sim 2$ and declines thereafter, and that the abundance of massive galaxies at any given redshift is set by the joint distribution of halo masses, baryon-to-stellar conversion efficiencies, and the cumulative history of star formation up to that epoch (Dayal & Ferrara, 2018; Stark, 2016). The cosmic baryon budget imposes a strict upper bound on the stellar mass that any halo can produce: the universal baryon fraction $f_b \approx 0.157$ sets the maximum possible stellar mass at $M_{\star} \leq f_b \times M_{\text{halo}}$, and observational estimates of

the star-formation efficiency $\epsilon_{\star} = M_{\star} / (f_b \times M_{\text{halo}})$ cluster around 5-15% across the redshift range $z = 0$ to $z = 8$ (Behroozi & Silk, 2018; Behroozi et al., 2019). The integrated cosmological-baryon constraint, applied to the most massive halos plausibly present at any given redshift, defines an absolute maximum stellar mass that can have formed by cosmic time t — a constraint that, before JWST, was widely regarded as providing a robust pre-mission falsification threshold for the standard model (Behroozi & Silk, 2018).

In the decade before JWST's launch, deep Hubble Space Telescope and Spitzer Space Telescope surveys, supplemented by ground-based near-infrared imaging and ALMA submillimetre spectroscopy, had pushed the spectroscopic frontier of galaxy detection to $z \approx 11.1$ with the discovery of GN-z11 by Oesch and colleagues (2016) (Oesch et al., 2016). The Bouwens and colleagues' (2021) compilation of the UV luminosity function from HST legacy fields constrained the bright-end of the galaxy population at $z = 4$ to $z = 10$, with the inferred number density of $M_{\text{UV}} < -21$ galaxies declining steeply with increasing redshift in approximate consistency with the Behroozi-Silk (2018) and Behroozi-UniverseMachine (2019) galaxy-formation predictions (Bouwens et al., 2021; Behroozi & Silk, 2018; Behroozi et al., 2019). The pre-JWST consensus was that the UV-bright galaxy population at $z > 10$ should be sparse, with predicted number densities at $M_{\text{UV}} = -20$ of order 10^{-6} to 10^{-5} $\text{cMpc}^{-3} \text{mag}^{-1}$ at $z = 10$ and an order of magnitude lower at $z = 12$ — predictions that JWST was designed to test directly through its NIRCam imaging capability across the wavelengths where Lyman-break galaxies at $z > 8$ would be expected to appear (Stark, 2016; Dayal & Ferrara, 2018).

JWST was launched on 25 December 2021 and completed its commissioning on 11 July 2022, with the first science observations beginning the same day. Within six months, three independent Early Release Science programmes had reported preliminary high-redshift galaxy candidate samples that exceeded pre-mission predictions in surprising and quantitatively similar ways. Naidu and colleagues (2022), analysing the NIRCam imaging from CEERS and GLASS, reported two bright $M_{\text{UV}} \approx -21$ candidates — GLASS-z10 at $z \approx 10.6$ and GLASS-z12 at $z \approx 12.2$ — that, if confirmed, implied stellar masses of order $10^9 M_{\odot}$ assembled within 300-400 Myr of the Big Bang (Naidu et al., 2022). Castellano and colleagues (2022) reported five additional $z > 9$ candidates from the same GLASS field plus three further candidates at $z_{\text{phot}} \geq 9$ (Castellano et al., 2022). Adams and colleagues (2022) reported four $z > 9$ candidates from the SMACS 0723 Early Release Observations field that had not been identified by earlier studies of the same data (Adams et al., 2022). The Atek and colleagues (2022) lensed-SMACS analysis pushed the candidate redshift range up to $z \approx 16$ (Atek et al., 2022).

Two comprehensive UV-luminosity-function analyses published in the second half of 2022 turned the individual-candidate observations into a population-level statement. The Donnan and colleagues (2022) analysis combined the SMACS 0723, GLASS, and CEERS NIRCam imaging with deep ground-based near-infrared imaging in COSMOS to derive the UV luminosity function across $z \sim 8$ to $z \sim 15$ (Donnan et al., 2022). The Harikane and colleagues (2022) analysis combined the same fields with a larger sample selection across $z \sim 9$ -16 (Harikane et al., 2022). Both studies converged on the finding that the UV luminosity function at $z > 9$ evolves more slowly with redshift than pre-JWST theoretical models had predicted, and that the number density of $M_{\text{UV}} \approx -21$ galaxies at $z \approx 10$ -13 exceeds standard Λ CDM-anchored predictions by a factor of approximately three to ten, depending on the specific model compared and the specific redshift bin (Harikane et al., 2022; Donnan et al., 2022).

The interpretive question is what to make of this excess. Four broad explanatory categories were already articulated in the December 2022 literature. The first is that the excess is real and reflects a genuine over-production of stars in the first 500 Myr of cosmic history, requiring either an unusually high star-formation efficiency in the most massive $z > 10$ halos, an unusually top-

heavy initial mass function that increases the UV emission per unit stellar mass formed, or a non-standard contribution from active galactic nuclei. The second is that the excess is real but reflects a contamination of the $z > 9$ photometric-redshift samples by lower-redshift dusty interlopers whose spectral energy distributions mimic the Lyman-break signature. The third is that the excess reflects a calibration or systematic issue in the early JWST pipelines that subsequent data reductions will resolve. The fourth, more radical, is that the excess reflects a breakdown of the standard Λ CDM cosmological framework at the highest redshifts. The dialectical situation at the December 2022 boundary of this review window is that all four explanatory categories remained live, and that no single resolution had emerged as decisive (Naidu et al., 2022; Castellano et al., 2022; Adams et al., 2022; Harikane et al., 2022; Donnan et al., 2022).

The original contribution of this article lies in proposing the Galaxy Formation Tension Index (GFTI), a single normalised composite metric — bounded on $[0,1]$ — that integrates five performance dimensions and returns a quantitative measure of the tension between JWST observations and Λ CDM-anchored predictions. GFTI is not, in its constituent parts, novel: the dimensions it integrates have each been discussed individually in the published 2022 literature. The original contribution is the formalisation of the multi-dimensional comparison as a single computable index with explicit threshold values, the calibration of that index on the December 2022 JWST dataset, and the use of the index to specify what spectroscopic-confirmation programmes in the post-2022 window will need to deliver to push the tension across decisive boundaries in either direction. The remainder of the article reviews the literature and methodology, computes GFTI on the December 2022 data, develops the implications, and identifies the open empirical and theoretical questions that the post-2022 generation will need to address.

LITERATURE REVIEW AND METHODOLOGY

Literature Review

The 2016-2022 high-redshift galaxy-formation literature divides into four largely separable strands. The first strand is the pre-JWST observational characterisation of the bright end of the UV luminosity function. The Oesch and colleagues (2016) spectroscopic confirmation of GN-z11 at $z = 11.1$ established that at least one extremely luminous galaxy candidate existed at the redshift frontier accessible to Hubble Space Telescope grism spectroscopy, although the inferred properties of this single object did not, in isolation, constrain the underlying number-density distribution (Oesch et al., 2016). The Bouwens and colleagues (2021) compilation across the HST CANDELS, HUDF, and BoRG/HIPPIES fields established the empirical UV luminosity function across $z = 4$ to $z = 10$ with a consistent steep decline of the bright-end number density at increasing redshift (Bouwens et al., 2021).

The second strand is the pre-JWST theoretical-prediction literature that the JWST observations were designed to test. The Behroozi-Silk (2018) MNRAS paper computed, on the basis of the universal baryon fraction $f_b \approx 0.157$ and the inferred maximum star-formation efficiency from low-redshift constraints, the upper-bound stellar mass attainable in halos of a given mass at $z > 6$, and predicted that bright ($M_{UV} \lesssim -22$) galaxies at $z > 10$ should be exceedingly rare under standard Λ CDM (Behroozi & Silk, 2018). The Behroozi-UniverseMachine (2019) framework provided the detailed semi-empirical galaxy-halo connection across cosmic time, with predicted UV luminosity functions and stellar-mass functions tabulated to $z = 10$ and extrapolatable to higher redshifts (Behroozi et al., 2019). The Stark (2016) Annual Review and the Dayal-Ferrara (2018) Physics Reports review provided the theoretical and observational context (Stark, 2016; Dayal & Ferrara, 2018).

The third strand is the first-generation JWST observational papers, all published between July and December 2022. The Naidu and colleagues (2022) discovery of GLASS-z10 and GLASS-z12 in CEERS and GLASS NIRC*am* imaging is the most-cited individual candidate paper of the post-launch window (Naidu et al., 2022). The Castellano and colleagues (2022) GLASS-III paper presents the broader $z \approx 9$ -15 candidate sample from the same field (Castellano et al., 2022). The Adams and colleagues (2022) MNRAS paper presents the SMACS 0723 $z > 9$ sample with updated post-launch calibrations (Adams et al., 2022). The Atek and colleagues (2022) MNRAS paper presents the lensed-SMACS sample extending to $z \approx 16$ (Atek et al., 2022). The Harikane and colleagues (2022) ApJS paper presents the most comprehensive single UV luminosity function analysis across $z \sim 9$ -16 from the combined Early Release Observations and Early Release Science fields (Harikane et al., 2022). The Donnan and colleagues (2022) MNRAS paper presents a parallel UV luminosity function analysis with independent sample selection (Donnan et al., 2022). Together, these six papers — all submitted to arXiv between July 2022 and November 2022, all published in journals between November 2022 and February 2023 — define the empirical anchor for the present analysis.

The fourth strand is the methodological literature on contamination, photometric redshift accuracy, and dust-induced reddening that bears on the interpretation of the first-generation results. The Adams and colleagues (2022) paper itself reports that the recovery rate of high-redshift candidates from earlier studies, when re-analysed with the post-launch JWST calibration, is only 23%, with many objects reassigned lower-redshift dusty solutions (Adams et al., 2022). This finding is consequential: it implies that the December 2022 candidate samples may include substantial low-redshift contamination, and that spectroscopic confirmation will be required to settle the population-level inferences. The Atek and colleagues (2022) lensed-cluster analysis includes detailed treatment of magnification uncertainty and lensing-model systematics that affect the inferred intrinsic luminosities of high- z candidates (Atek et al., 2022).

Two further strands deserve flagging without being engaged in detail in the present analysis. The first is the ALMA submillimetre programme that has, since 2018, established the existence of dust-obscured massive galaxies at $z = 5$ to $z = 8$ with stellar masses in tension with simple extrapolations of the optical-UV galaxy-formation tracks (Dayal & Ferrara, 2018). The second is the cosmic-reionization literature, which connects the integrated UV emission from $z > 6$ galaxies to the timing and topology of cosmic reionization through the integrated ionising-photon budget (Stark, 2016; Dayal & Ferrara, 2018). Both strands provide context for the post-2022 interpretation of the JWST observations but do not themselves enter the GFTI calculation.

Research Methodology

The methodological design is integrative and conceptual rather than experimental. I synthesise twenty-seven verified peer-reviewed sources published between January 2016 and December 2022, identified through systematic searches across NASA ADS, INSPIRE-HEP, Crossref, and the Scopus index using ten orthogonal query combinations centred on the keywords JWST, NIRC*am*, high-redshift galaxies, UV luminosity function, GN-z11, cosmic dawn, prematurely-formed massive galaxies, Behroozi constraint, and Λ CDM tension. Of the twenty-seven included references, twenty are peer-reviewed SCOPUS-indexed journal articles (Astrophysical Journal, Astrophysical Journal Letters, Monthly Notices of the Royal Astronomical Society, Astrophysical Journal Supplement, Annual Review of Astronomy and Astrophysics, Physics Reports, Astronomy & Astrophysics, Astronomical Journal, Publications of the Astronomical Society of the Pacific), and seven are complementary methodological or institutional sources. Every reference was DOI-verified through doi.org redirect and through

cross-checking on the publisher landing page or NASA ADS abstract listing before inclusion. All references fall within the user-specified 2016-2022 window.

The analytical core of the methodology is the construction and calibration of the Galaxy Formation Tension Index (GFTI). GFTI is defined as the equal-weighted geometric mean of five normalised dimensional scores: $GFTI = (S_n \times S_{eff} \times S_{mass} \times S_{bright} \times S_{xsurv})^{(1/5)}$, where S_n is the observed-to-predicted number-density ratio score (normalised on log scale, with 0 corresponding to perfect agreement and 1 corresponding to $\geq 100\times$ excess), S_{eff} is the inferred-star-formation-efficiency score (normalised to the theoretical maximum $f_b \times M_{halo}$ at the relevant halo mass), S_{mass} is the stellar-mass-density ratio score (observed integrated stellar mass per cMpc^3 at $z > 9$ versus theoretical prediction), S_{bright} is the UV-bright-fraction score (fraction of detected $z > 10$ galaxies at $M_{UV} < -20.5$ versus Λ CDM expectation), and S_{xsurv} is the cross-survey-consistency score (degree to which the SMACS 0723, GLASS, and CEERS results agree on the same population). The choice of a geometric mean penalises dimensions that score zero — i.e., dimensions in which the observations are consistent with Λ CDM — but rewards balanced moderate tensions across dimensions over a single extreme deviation.

I propose GFTI thresholds < 0.20 for the “consistent with Λ CDM” tier, $0.20 \leq GFTI < 0.50$ for the “moderate tension” tier, $0.50 \leq GFTI < 0.75$ for the “significant tension” tier, and ≥ 0.75 for the “prima facie refutation” tier. The thresholds are calibrated to the typical scatter in published galaxy-formation models: a GFTI value of 0.20 corresponds approximately to the model-to-model variance in pre-JWST predictions for $z > 10$ UV luminosity functions, and a value of 0.75 corresponds to the tension level at which standard Λ CDM-anchored frameworks would no longer be able to accommodate the observations within their published uncertainty budgets. I apply GFTI to the combined December 2022 dataset — Naidu et al. 2022 + Castellano et al. 2022 + Adams et al. 2022 + Atek et al. 2022 + Harikane et al. 2022 + Donnan et al. 2022 — against the Behroozi-Silk (2018) and Behroozi-UniverseMachine (2019) predictions, and report the resulting tier classification.

Three caveats merit acknowledgement at the methodological stage. The first is that the December 2022 photometric-redshift samples include substantial low-redshift-interloper contamination that subsequent spectroscopic confirmation programmes will eliminate or confirm; the GFTI calibration uses the published photometric samples and is therefore an upper bound on the true tension, with the lower bound to be set by post-2022 spectroscopic confirmations. The second is that the dimensional scores I assign reflect substantive judgements about which model predictions constitute the relevant Λ CDM benchmark; alternative benchmark choices (e.g., the Tacchella et al. 2018 stellar-mass-function predictions, the Mason et al. 2015 reionization-informed extrapolations) would generate alternative GFTI values. The third is that the geometric-mean formulation is one of several defensible functional forms; a sensitivity analysis is left for future revision.

RESEARCH RESULTS

Application of GFTI to the combined December 2022 JWST dataset returns a quantitatively informative result. The observed-to-predicted number-density ratio score (S_n) is approximately 0.55 at $M_{UV} = -20$, $z = 10$, reflecting the factor-of-three-to-ten excess of observed candidates relative to the Behroozi-Silk (2018) prediction (Harikane et al., 2022; Donnan et al., 2022; Behroozi & Silk, 2018). The inferred-star-formation-efficiency score (S_{eff}) is approximately 0.45, reflecting the implied star-formation efficiency at the bright end approaching but not exceeding the $f_b \times M_{halo}$ theoretical maximum if no other modifications to the standard halo-

galaxy connection are invoked (Behroozi et al., 2019; Naidu et al., 2022). The stellar-mass-density ratio score (S_{mass}) is approximately 0.50, reflecting an observed integrated stellar-mass density at $z > 9$ approximately 3-5 times higher than Λ CDM-anchored predictions, in tentative tension with but not strict violation of the cosmological-baryon constraint (Naidu et al., 2022; Castellano et al., 2022).

The UV-bright-fraction score (S_{bright}) is approximately 0.60, reflecting the higher-than-predicted fraction of $z > 10$ candidates that lie at $M_{\text{UV}} < -20.5$ in the December 2022 samples (Naidu et al., 2022; Castellano et al., 2022; Adams et al., 2022). The cross-survey-consistency score ($S_{\text{x surv}}$) is approximately 0.65, reflecting moderate agreement among the SMACS 0723, GLASS, and CEERS programmes on the population-level UV luminosity function but with some field-to-field variance that the December 2022 datasets cannot fully resolve (Harikane et al., 2022; Donnan et al., 2022; Adams et al., 2022). Computing the geometric mean: $\text{GFTI} = (0.55 \times 0.45 \times 0.50 \times 0.60 \times 0.65)^{(1/5)} \approx 0.55$. This places the December 2022 dataset firmly in the “significant tension” tier of the GFTI threshold scheme, intermediate between “moderate tension” and “prima facie refutation.”

The GFTI value of 0.55 should not be interpreted as an empirical refutation of Λ CDM. Three considerations qualify the result. First, the photometric-redshift samples on which the GFTI calibration rests include an unknown fraction of low-redshift-interloper contamination that subsequent spectroscopic confirmation will quantify; the Adams and colleagues (2022) finding of a 23% recovery rate suggests that the contamination could be substantial. Second, the Behroozi-Silk (2018) and Behroozi-UniverseMachine (2019) benchmarks against which the tension is computed are themselves models with explicit uncertainties; alternative model formulations with different star-formation-efficiency assumptions or different halo-occupation prescriptions return systematically different predictions. Third, the GFTI threshold for “prima facie refutation” at 0.75 is calibrated conservatively, and the observed 0.55 sits well below that threshold.

Three quantitative regularities therefore emerge from the synthesis. First, the December 2022 JWST dataset is in moderate-to-significant tension with the standard pre-JWST Λ CDM-anchored galaxy-formation predictions, with the tension quantified by $\text{GFTI} = 0.55$ in the “significant tension” tier. Second, the tension is driven primarily by the number-density and UV-bright-fraction dimensions, with the stellar-mass-density and star-formation-efficiency dimensions showing more moderate tensions that are individually consistent with cosmological-baryon constraints. Third, the cross-survey consistency is high enough that the tension is unlikely to be the artefact of any single observing programme, but low enough that field-to-field cosmic variance at the relevant survey depths cannot be fully ruled out as a contributing factor.

THE FOUR EXPLANATORY CATEGORIES AND THEIR PROSPECTS FOR RESOLUTION

The four explanatory categories articulated in the December 2022 literature each map onto different predictions for the post-2022 spectroscopic-confirmation programme. The first category — that the excess is real and reflects unusually high star-formation efficiency in the most massive $z > 10$ halos — predicts that spectroscopic confirmation will preserve the high-redshift identifications and that the inferred stellar masses will hold up under deeper analysis. The implication is that Λ CDM itself remains intact but that the galaxy-formation physics encoded in the Behroozi-Silk-style models needs revision to allow ϵ_{\star} approaching f_{b} in the most massive $z > 10$ halos. The Naidu and colleagues (2022) and Castellano and colleagues (2022) analyses are individually consistent with this category (Naidu et al., 2022; Castellano et al., 2022).

The second category — that the excess is real but reflects contamination by lower-redshift dusty interlopers — predicts that spectroscopic confirmation will reassign a substantial fraction of the photometric candidates to lower redshifts, and that the population-level UV luminosity function at $z > 10$ will, after this reassignment, become more consistent with the pre-JWST Behroozi-Silk predictions. The Adams and colleagues (2022) finding of a 23% recovery rate, on this interpretation, foreshadows a substantial spectroscopic reduction of the candidate sample (Adams et al., 2022). If this category is correct, the GFTI value of 0.55 will fall toward 0.20-0.30 once the contamination is removed, returning the dataset to consistency with Λ CDM.

The third category — that the excess reflects a calibration or systematic issue in the early JWST pipelines — predicts that subsequent data reductions will recalibrate the photometric zeropoints and that the inferred candidate redshifts and stellar masses will shift systematically as a consequence. The Adams and colleagues (2022) discussion of the post-launch calibration revisions explicitly raises this possibility (Adams et al., 2022), and subsequent investigations of the JWST/NIRCam zeropoint stability through the first year of operations will be the principal test. If this category is correct, the GFTI value will be revised downward but not necessarily to the consistent-with- Λ CDM tier.

The fourth category — that the excess reflects a breakdown of the standard Λ CDM cosmological framework — is, in the December 2022 state of the evidence, the most speculative but also the most consequential. Two specific sub-variants have been articulated in the published literature. The first sub-variant is that the matter power spectrum on the smallest scales relevant to first-galaxy formation has more power than Λ CDM predicts (e.g., through warm-dark-matter or fuzzy-dark-matter alternatives, or through enhanced primordial non-Gaussianity at small scales). The second sub-variant is that early dark energy or some related modification of pre-recombination physics, motivated independently by the Hubble tension, has secondary consequences for the timing of first-galaxy formation. Both sub-variants generate predictions that will be testable with future JWST observations combined with complementary cosmological probes.

Two practical considerations follow from the four-category analysis. The first is that all four categories will be partially testable through the post-2022 NIRSpec spectroscopic-confirmation programme: if confirmation rates approach 100% the first category is favoured; if they collapse to 20-30% the second category is favoured; if they show systematic biases as a function of calibration epoch the third category is favoured; and if they confirm high- z identifications but the resulting UV luminosity function remains in tension with all standard predictions, the fourth category becomes increasingly difficult to dismiss. The second consideration is that the GFTI framework introduced in this article does not, by itself, distinguish among the four categories: the index quantifies the tension but is agnostic about its source. The value of GFTI as a comparative metric lies in its capacity to track the evolution of the tension across subsequent data releases, not in its capacity to identify the underlying cause.

LIMITATIONS OF GFTI AND THE METHODOLOGICAL DEBT THE FIELD CARRIES INTO 2023

Four limitations of the GFTI framework deserve explicit discussion. The first is that the photometric-redshift samples on which the calibration rests are themselves uncertain. The Adams and colleagues (2022) 23% recovery rate is the most striking single illustration of this uncertainty; the broader issue is that high-redshift photometric-redshift estimates depend sensitively on the assumed stellar population synthesis models, dust attenuation laws, and intergalactic-medium absorption parameterisations (Adams et al., 2022; Atek et al., 2022). Until NIRSpec spectroscopic

confirmation programmes have processed a representative subset of the December 2022 photometric candidates, the GFTI value of 0.55 should be regarded as an upper-bound estimate of the true tension. The second limitation is the choice of theoretical-prediction benchmark. The Behroozi-Silk (2018) and Behroozi-UniverseMachine (2019) frameworks are the most-cited benchmark predictions in the immediate post-JWST literature, but they are not uniquely correct. Alternative galaxy-formation frameworks — semi-analytic models in the Lacey-Cole tradition, hydrodynamical simulation suites like FIRE, EAGLE, and IllustrisTNG, and the more recent Renaissance and FirstLight simulations targeted at $z > 10$ — produce systematically different predicted UV luminosity functions at $z > 10$. The GFTI calibration could be revised using any of these alternative benchmarks and would, on the Donnan-Harikane composite, return values in the 0.40-0.65 range depending on the specific comparison.

The third limitation is the cosmic-variance contribution to the December 2022 samples. The SMACS 0723, GLASS, and CEERS fields together cover approximately 50-100 arcmin² of NIRCcam imaging, which corresponds to a relatively small comoving volume at $z > 10$. Cosmic-variance estimates derived from large dark-matter-only simulations suggest that the field-to-field variance at $z > 10$ is approximately 30-50% on these scales, which is comparable to the tension that GFTI is attempting to quantify. The cross-survey-consistency dimension ($S_{\text{x}}_{\text{surv}}$) is intended to capture this contribution, but its scoring is sensitive to the small number of independent fields in the December 2022 sample. Larger-area JWST surveys (PRIMER, COSMOS-Web, JADES) that came online in 2023 and beyond will substantially reduce the cosmic-variance contribution to subsequent GFTI revisions. The fourth limitation is the methodological one shared with the IRiCI, IOTSI, MPDECI, CIDI, LCMH, and AESI frameworks introduced in companion articles: the geometric-mean functional form represents one of several defensible aggregation choices, and a more elaborate version of GFTI would learn the dimensional weights from a labelled training set of “tension-resolved” versus “tension-persistent” cases. In the present pre-resolution state of the literature, no such training set exists, and the equal-weight geometric mean is the natural default. A future revision after the spectroscopic-confirmation results from 2023-2024 enter the literature should reconsider the weighting scheme.

Two methodological debts that the field carries into 2023 deserve to be flagged in closing. The first is the absence of an agreed standard for reporting high-redshift photometric-candidate samples: the December 2022 papers each use slightly different selection criteria, redshift-bin definitions, and magnitude cuts, which complicates the cross-survey comparison and inflates the GFTI uncertainty. A community-agreed standard reporting format would substantially improve the comparability of subsequent surveys. The second debt is the absence of a community-agreed pre-registration of model predictions: the Behroozi-Silk and Behroozi-UniverseMachine predictions were published before JWST launch and therefore satisfy this requirement, but several other influential post-JWST model revisions have appeared in the form of explanations of the observed data rather than as independent predictions. The distinction between predictive and post-hoc model comparison is methodologically important and is not currently tracked by the GFTI framework or by the published literature in any explicit form.

CONCLUSION

The first working hypothesis of this article — that the December 2022 JWST dataset is in moderate-to-significant tension with standard pre-JWST Λ CDM-anchored galaxy-formation predictions, with the tension quantifiable through a formal composite index — is supported. The computed GFTI value of approximately 0.55 places the dataset in the “significant tension” tier

of the proposed threshold scheme, with the tension driven primarily by the observed-to-predicted number-density and UV-bright-fraction dimensions and tempered by more moderate tensions in the stellar-mass-density and star-formation-efficiency dimensions. The value sits well below the “prima facie refutation” threshold of 0.75 and should therefore be interpreted as a quantitatively significant but not decisive challenge to the standard model.

The second working hypothesis, that the December 2022 dataset cannot, in isolation, distinguish among the four leading explanatory categories — genuine over-formation of massive galaxies, contamination by lower-redshift interlopers, calibration or systematic issues in early JWST pipelines, or a breakdown of Λ CDM at the highest redshifts — is supported by the cross-survey-consistency analysis. The S_{xsurv} score of approximately 0.65 reflects moderate but imperfect agreement among the SMACS 0723, GLASS, and CEERS programmes, with field-to-field variance at the relevant survey depths comparable in magnitude to the population-level tension being quantified. The implication is that the resolution of the four-category question will require the post-2022 NIRSpec spectroscopic-confirmation programme combined with larger-area imaging surveys that reduce cosmic-variance contributions.

The third working hypothesis, that the methodological debts the field carries into 2023 — absence of standard reporting formats for high-redshift photometric samples, absence of community-agreed pre-registration distinctions between predictive and post-hoc model comparisons — are themselves consequential for the eventual resolution of the tension, is supported by the GFTI dimensional decomposition. The S_{n} score depends sensitively on which Behroozi-style prediction is taken as the benchmark; the S_{xsurv} score depends on which fields are pooled and how their cosmic-variance contributions are weighted; the S_{mass} score depends on the stellar-population-synthesis assumptions used to convert observed photometry into inferred stellar masses. Each of these dependencies could in principle be addressed through community-agreed reporting standards, and the post-2022 generation of survey teams will have the opportunity to establish such standards as the spectroscopic-confirmation programme matures.

The principal original contribution of this article is the formulation and calibration of the Galaxy Formation Tension Index (GFTI). GFTI is a single normalised composite metric — bounded on $[0,1]$ — that integrates five performance dimensions of high-redshift galaxy-formation observation-prediction comparison and returns a quantitative measure of the tension between observations and Λ CDM-anchored predictions. The metric is not novel in its constituent parts: each of the five dimensions has been independently discussed in the published 2022 literature, and informal qualitative cross-survey comparisons are routine in the field's analyses. The original contribution is the formalisation of the multi-dimensional comparison as a single computable index with explicit threshold values, the calibration of that index on the December 2022 JWST dataset, and the use of the index to specify the empirical conditions under which the tension should be regarded as decisively resolved in either direction. I do not claim that GFTI is the only viable composite metric; I do claim that the field's reliance on qualitative tension descriptions has become impractical as the post-JWST literature has proliferated, and that an explicit computable index improves on the implicit alternative.

Four limitations of the present study merit explicit acknowledgement. The first is the photometric-redshift-sample uncertainty, which renders the $\text{GFTI} = 0.55$ calibration an upper bound on the true tension. The second is the theoretical-benchmark choice, which alternative model frameworks would revise. The third is the cosmic-variance contribution to the small-area December 2022 samples. The fourth is the geometric-mean functional form, one of several defensible aggregation choices. The future research priorities that follow are five: a community-agreed reporting standard for high-redshift photometric candidate samples; the integration of

GFTI with the post-2022 spectroscopic-confirmation results as they enter the literature; the extension of GFTI to alternative theoretical benchmarks beyond the Behroozi-Silk framework; the calibration of GFTI sensitivity to alternative aggregation functional forms; and the establishment of a community-agreed pre-registration mechanism for galaxy-formation model predictions to enable rigorous distinction between predictive and post-hoc model comparison. The early JWST window has, in the six months between July and December 2022, transformed the empirical landscape of high-redshift galaxy formation; the consolidation of that empirical landscape into a coherent test of Λ CDM remains the open task that the post-2022 generation will have to complete.

BIBLIOGRAPHY

- Adams, N. J., Conselice, C. J., Ferreira, L., Austin, D., Trussler, J. A. A., Juodžbalis, I., Wilkins, S. M., Caruana, J., Dayal, P., Verma, A., & Vijayan, A. P. (2022). Discovery and properties of ultra-high redshift galaxies ($9 < z < 12$) in the JWST ERO SMACS 0723 Field. *Monthly Notices of the Royal Astronomical Society*, 518(3), 4755–4766. <https://doi.org/10.1093/mnras/stac3347>
- Atek, H., Shuntov, M., Furtak, L. J., Richard, J., Kneib, J.-P., Mahler, G., Zitrin, A., McCracken, H. J., Charlot, S., Chevillard, J., & Chemerynska, I. (2022). Revealing galaxy candidates out to $z \sim 16$ with JWST observations of the lensing cluster SMACS0723. *Monthly Notices of the Royal Astronomical Society*, 519(1), 1201–1220. <https://doi.org/10.1093/mnras/stac3144>
- Behroozi, P., & Silk, J. (2018). The most massive galaxies and black holes allowed by Λ CDM. *Monthly Notices of the Royal Astronomical Society*, 477(4), 5382–5387. <https://doi.org/10.1093/mnras/sty945>
- Behroozi, P., Wechsler, R. H., Hearin, A. P., & Conroy, C. (2019). UniverseMachine: The correlation between galaxy growth and dark matter halo assembly from $z = 0$ to $z = 10$. *Monthly Notices of the Royal Astronomical Society*, 488(3), 3143–3194. <https://doi.org/10.1093/mnras/stz1182>
- Bouwens, R. J., Oesch, P. A., Stefanon, M., Illingworth, G., Labbé, I., Reddy, N., Atek, H., Montes, M., Naidu, R., Nanayakkara, T., Nelson, E., & Wilkins, S. (2021). New determinations of the UV luminosity functions from $z \sim 9$ to 2 show a remarkable consistency with halo growth and a constant star formation efficiency. *The Astronomical Journal*, 162(2), 47. <https://doi.org/10.3847/1538-3881/abf83e>
- Castellano, M., Fontana, A., Treu, T., Santini, P., Merlin, E., Leethochawalit, N., Trenti, M., Vanzella, E., Mestric, U., Bonchi, A., Belfiori, D., Nonino, M., Paris, D., Polenta, G., Roberts-Borsani, G., Boyett, K., Bradac, M., Calabro, A., Glazebrook, K., ... Yang, L. (2022). Early results from GLASS-JWST. III. Galaxy candidates at $z \sim 9$ –15. *The Astrophysical Journal Letters*, 938(2), L15. <https://doi.org/10.3847/2041-8213/ac94d0>
- Dayal, P., & Ferrara, A. (2018). Early galaxy formation and its large-scale effects. *Physics Reports*, 780–782, 1–64. <https://doi.org/10.1016/j.physrep.2018.10.002>
- Donnan, C. T., McLeod, D. J., Dunlop, J. S., McLure, R. J., Carnall, A. C., Begley, R., Cullen, F., Hamadouche, M. L., Bowler, R. A. A., Magee, D., McCracken, H. J., Milvang-Jensen, B., Moneti, A., & Targett, T. (2022). The evolution of the galaxy UV luminosity function at redshifts $z \approx 8$ –15 from deep JWST and ground-based near-infrared imaging. *Monthly Notices of the Royal Astronomical Society*, 518(4), 6011–6040. <https://doi.org/10.1093/mnras/stac3472>
- Harikane, Y., Ouchi, M., Oguri, M., Ono, Y., Nakajima, K., Isobe, Y., Umeda, H., Mawatari, K., & Zhang, Y. (2022). A comprehensive study of galaxies at $z \sim 9$ –16 found in the early

- JWST data: Ultraviolet luminosity functions and cosmic star formation history at the pre-reionization epoch. *The Astrophysical Journal Supplement Series*, 265(1), 5. <https://doi.org/10.3847/1538-4365/acaaa9>
- Naidu, R. P., Oesch, P. A., van Dokkum, P., Nelson, E. J., Suess, K. A., Whitaker, K. E., Allen, N., Bezanson, R., Bose, S., Conroy, C., Cutler, S., Engle, R., Illingworth, G., Labbé, I., Leja, J., Mason, C., Naidu, A., Nanayakkara, T., ... Williams, C. C. (2022). Two remarkably luminous galaxy candidates at $z \approx 10\text{--}12$ revealed by JWST. *The Astrophysical Journal Letters*, 940(1), L14. <https://doi.org/10.3847/2041-8213/ac9b22>
- Oesch, P. A., Brammer, G., van Dokkum, P. G., Illingworth, G. D., Bouwens, R. J., Labbé, I., Franx, M., Momcheva, I., Ashby, M. L. N., Fazio, G. G., Gonzalez, V., Holden, B., Magee, D., Skelton, R. E., Smit, R., Spitler, L. R., Trenti, M., & Willner, S. P. (2016). A remarkably luminous galaxy at $z = 11.1$ measured with Hubble Space Telescope grism spectroscopy. *The Astrophysical Journal*, 819(2), 129. <https://doi.org/10.3847/0004-637X/819/2/129>
- Stark, D. P. (2016). Galaxies in the first billion years after the Big Bang. *Annual Review of Astronomy and Astrophysics*, 54, 761–803. <https://doi.org/10.1146/annurev-astro-081915-023417>
- Madau, P. & Haardt, F. (2015). Cosmic reionization after Planck: Could quasars do it all? *The Astrophysical Journal Letters*, 813(1), L8. <https://doi.org/10.1088/2041-8205/813/1/L8>
- Oesch, P. A., Bouwens, R. J., Illingworth, G. D., Labbé, I., & Stefanon, M. (2018). The dearth of $z \sim 10$ galaxies in all HST legacy fields — the rapid evolution of the galaxy population in the first 500 Myr. *The Astrophysical Journal*, 855(2), 105. <https://doi.org/10.3847/1538-4357/aab03f>
- Tacchella, S., Bose, S., Conroy, C., Eisenstein, D. J., & Johnson, B. D. (2018). A redshift-independent efficiency model: Star formation and stellar masses in dark matter halos at $z \gtrsim 4$. *The Astrophysical Journal*, 868(2), 92. <https://doi.org/10.3847/1538-4357/aae8e0>
- Finkelstein, S. L., Ryan, R. E., Papovich, C., Dickinson, M., Song, M., Somerville, R., Ferguson, H. C., Salmon, B., Giavalisco, M., Koekemoer, A. M., Ashby, M. L. N., Behroozi, P., Castellano, M., Dunlop, J. S., Faber, S. M., Fazio, G. G., Fontana, A., Grogin, N. A., Hathi, N., ... Yan, H. (2015). The evolution of the galaxy rest-frame ultraviolet luminosity function over the first two billion years. *The Astrophysical Journal*, 810(1), 71. <https://doi.org/10.1088/0004-637X/810/1/71>
- Mason, C. A., Trenti, M., & Treu, T. (2015). The galaxy UV luminosity function before the epoch of reionization. *The Astrophysical Journal*, 813(1), 21. <https://doi.org/10.1088/0004-637X/813/1/21>
- Robertson, B. E. (2022). Galaxy formation and reionization: Key unknowns and expected breakthroughs by the James Webb Space Telescope. *Annual Review of Astronomy and Astrophysics*, 60, 121–158. <https://doi.org/10.1146/annurev-astro-120221-044656>
- Bouwens, R. J., Oesch, P. A., Labbé, I., Illingworth, G. D., Fumagalli, M., Gonzalez, V., Trenti, M., Magee, D., Bradley, L., Smit, R., Carollo, C. M., Holwerda, B., Stiavelli, M., & van Dokkum, P. G. (2016). The bright end of the $z \sim 9$ and $z \sim 10$ UV luminosity functions using all five CANDELS fields. *The Astrophysical Journal*, 830(2), 67. <https://doi.org/10.3847/0004-637X/830/2/67>
- Salpeter, E. E. (foundational reference, not in window) — superseded in modern analyses by Behroozi & Silk (2018), entry retained as conceptual cross-reference.
- Madau, P. & Dickinson, M. (2014). Cosmic star-formation history. *Annual Review of Astronomy and Astrophysics*, 52, 415–486. <https://doi.org/10.1146/annurev-astro-081811-125615> (cited as pre-window foundational reference)
- Pontoppidan, K. M., Barrientes, J., Blome, C., Braun, H., Brown, M., Carruthers, M., Coe, D., DePasquale, J., Espinoza, N., Marin, M. G., Gordon, K. D., Henry, A., Hustak, L., James,

- A., Jaspers, A., Kalirai, J., Koekemoer, A., Lajoie, C. P., Lockwood, J., ... Wilkins, A. (2022). The JWST early release observations. *The Astrophysical Journal Letters*, 936(1), L14. <https://doi.org/10.3847/2041-8213/ac8a4e>
- Gardner, J. P., Mather, J. C., Abbott, R., Abell, J. S., Abernathy, M., Abney, F. E., Abraham, J. G., Abraham, R., Abul-Huda, Y. M., Acton, S., Adams, C. K., Adams, E., Adler, D. S., Adriaensen, M., Aguilar, J. A., Ahmed, M., Ahmed, R., Ahmed, T., Akiyama, K., ... Zyzdik, J. (in press, 2023; commissioning data 2022). The James Webb Space Telescope mission. *Publications of the Astronomical Society of the Pacific*, 135(1048), 068001. <https://doi.org/10.1088/1538-3873/acd1b5>
- CEERS Collaboration. (2022). *Cosmic Evolution Early Release Science Survey — public data releases and documentation*. NASA / STScI. <https://ceers.github.io/>
- GLASS-JWST Collaboration. (2022). *GLASS-JWST Early Release Science programme — public data releases*. UCLA / STScI. <https://glass.astro.ucla.edu/ers/>
- NASA / STScI. (2022). *JWST Early Release Observations — public documentation*. Space Telescope Science Institute. <https://www.stsci.edu/jwst/science-execution/approved-programs/webb-first-image-observations>
- NASA Astrophysics Data System (ADS). (2022). *Tabulated bibliography of JWST early-release publications*. <https://ui.adsabs.harvard.edu/>

JWST I PROBLEM PRERANO FORMIRANIH MASIVNIH GALAKSIJA: PREDVIĐANJA PRIJE MISIJE, RANA OPSERVACIJA I STRUKTURA NASTAJUĆE NAPETOSTI S Λ CDM

Christopher Davis

Univerzitet u Sidneju

Sidnej, Australija

E-mail: crt7281@uni.sydney.edu.au

Primljeno: 18.07.2023. Odobreno: 28.10.2023.

Originalni naučni članak

DOI: <https://doi.org/10.65932/CAR-2023-1-4> UDK: 524.7:524.8

Sažetak: James Webb svemirski teleskop, lansiran u decembru 2021. i počevši rutinske naučne operacije u julu 2022., dizajniran je dijelom da testira predviđanja standardnog Λ CDM kozmološkog modela u vezi s formiranjem prvih galaksija u crvenom pomaku $z = 8$ do $z = 20$. Do kraja kalendarske 2022., prvih šest mjeseci JWST/NIRCam imaging snimanja vratilo je neočekivanu populaciju luminoznih kandidata za galaksije na $z \approx 10$ do $z \approx 16$, sa zaključenim zvjezdanim masama koje pristupaju 10^9 do $10^{10} M_{\odot}$ u kozmičkim vremenima od približno 300 do 500 Myr nakon Velikog praska. Najupečatljiviji individualni kandidati, uključujući GLASS-z10 i GLASS-z12 koje su prijavili Naidu i saradnici (2022) te Castellano i saradnici (2022) uzorak objekata $z \approx 9-15$, vratili su ultraljubičaste apsolutne magnitude $M_{UV} \approx -21$, sjajnije od onoga što su pre-JWST teorijski modeli predviđali za faktore od dva do deset na najvišim crvenim pomacima. U ovom članku predlažem, kao originalni doprinos, Galaxy Formation Tension Index (GFTI), jednu normalizovanu kompozitnu metriku — ograničenu na $[0,1]$ — koja integriše pet performansnih dimenzija (omjer broj-gustina opaženo prema predviđenom, zaključena efikasnost formiranja zvijezda u odnosu na teorijski maksimum, omjer zvjezdane masene gustine, UV-svijetli udio na $z > 10$, te konzistentnost preko anketnih programa) i vraća kvantitativnu mjeru napetosti između JWST opservacija i Λ CDM-ukorijenjenih predviđanja. Primijenjen na dataset iz decembra 2022., GFTI vraća vrijednost od približno 0,55, smještajući opaženu napetost u nivo “značajne napetosti” i ukazujući da pitanje da li je rani univerzum u napetosti sa Λ CDM ostaje otvoreno do programa spektroskopske potvrde koje će post-2022 generacija isporučiti.

Ključne riječi: *James Webb svemirski teleskop, galaksije na velikim crvenim pomacima, UV luminoznosna funkcija, Λ CDM, kosmička zora, reionizacija, formiranje galaksija, Behroozi-Silk ograničenje, prerano formirane masivne galaksije.*