nature portfolio

Peer Review File

Hidden mass in early galaxies revealed by bottom-heavy initial mass functions

Corresponding Author: Ms Chloe Cheng

This manuscript has been previously reviewed at another journal. This document only contains information relating to versions considered at Nature Astronomy.

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #2

(Comments for the Author)
I have read with great interest the authors' manuscript, which is well presented and of clear importance, describing a highly novel and interesting JWST observing programme. This paper clearly merits publication in Nature Astronomy, however I have a few concerns that I would recommend the authors address first, in order to strengthen their conclusions.

Major points:

Your conclusions re the masses of "impossibly early galaxies" are largely based on the assumption that the IMFs you measure in the central regions of your objects are representative of their whole stellar populations. However, it can be seen from the cutouts presented in Fig. 1 that the NIRSpec shutters only cover a small fraction of the cores of these galaxies. I think it would be useful to include a discussion of what fraction of the light from/mass of your galaxies falls into your slitlets (in particular the regions used for your spectral extractions) and how this might impact your conclusions, e.g., if the IMFs are more MW-like outside these central regions. A new column or columns in Table 2 containing this information would be useful. A contrary point to discuss is obviously the much-more-compact nature of the z>3 sources, which presumably grow via mergers to become your oldest z=0.7 objects, so maybe the central regions of your sources are the right places to be looking. What physical scales do your slitlets cover compared to r_e for typical z>3 quiescent galaxies? Clearly IFU follow-up would be of interest here to properly settle this point. Papers by Padraig Alton+17+18 might be interesting to discuss here?

The pipeline and crds versions used to reduce the data products used in this paper are quite out of date (crds v1322 is now more than a year old). This is of particular importance because there were major updates made to the NIRSpec calibration files in Sep 2025, e.g., new flat field and dark reference files. In this context, I would strongly advise the authors to re-reduce their data with the most recent pipeline and crds versions to make sure that their results are not sensitive to these updates.

Related to the above, given the extremely high SNR of the NIRSpec continuum data, have the authors considered the effects of a.) NIRSpec flicker noise introduced during detector readout? Do you use the clean_flicker_noise step in the level 1 pipeline - if so what mode and parameters, if not I think this would be worth testing to make sure it doesn't impact the results; b.) NIRSpec extraction "wiggles" due to undersampling of the spatial PSF (e.g., Perna+23, Dumont+25, Leung+26). Both of these feel like they would be particularly important in this case of trying to tease out extremely subtle continuum features.

Whilst I appreciate that ALF and the models it contains are currently unique in their capability to fit the parameters of interest for this paper, it always concerns me that this makes it extremely difficult to evaluate the likely level of systematic uncertainty in results such as these. For example, Byrne & Stanway (2023) and Jones, Byrne & Stanway (2025) show several tests demonstrating that different sets of stellar spectral model libraries predict quite different strengths for commonly studied rest-optical absorption features at the same age and metallicity. Can the authors comment in the paper on the likely/possible/plausible level of systematic uncertainty on their results due to modelling choices? If this isn't possible, I think it would at least be sensible to acknowledge the existence of this issue.

Minor points:

Line 64: emotive language "ambitious" not really the right style here in my opinion, people will draw their own conclusions from the data you present. Again on line 209.

Line 149: The sentence that starts here is a little confusing, I would clarify that your "mass offsets" are virial-to-stellar mass offsets, presuming this is what you mean?

Line 182: Not sure I quite follow the logic here? Surely only applies if stars have started dying within the mass-range that gives rise to the features you're measuring - is this the case for these objects? A bit more clarification would be useful on this point I think.

Line 200: clarify "above" 1 solar mass instead of saying beyond?

Reviewer #3

(Comments for the Author)
Key results:
- Presents ultra-deep JWST/NIRSpec MSA spectroscopy from IMFERNO combined with deep optical LEGA-C spectra for 9 massive quiescent galaxies at z = 0.7, enabling unusually broad rest-frame wavelength leverage for IMF work.
- Uses full-spectrum fitting with alf allowing simultaneous inference of age, multi-element abundances, and low-mass IMF slopes, and reports bottom-heavy IMF normalizations in several galaxies (alpha_IMF_low_mass up to 5).
- Connects inferred IMF-driven M/L shifts to revised stellar masses and compares these against virial-mass estimates as a plausibility check.

Validity:
- No single issue prohibits publication in principle, but several unquantified modeling/assumption sensitivities currently prevent the strongest claims from being considered fully supported at Nature's standard.
- The most consequential validity risks are: fixed IMF low-mass cutoff and fixed high-mass slope, incomplete quantification of IMF–abundance degeneracies, and limited accounting of SPS/systematic uncertainties relative to the reported extreme alpha_IMF values.
- The environmental/selection context (one pointing associated with the "COSMOS Wall") plus n=9 also limits generalizability; this is not fatal but must be made explicit and considered when generalizations are made.

Originality and significance:
- The observational combination (very high-S/N JWST near-IR absorption features + LEGA-C optical constraints for the same galaxies) is novel and of broad interest to galaxy evolution and stellar population communities.
- The topic is timely given JWST-era "early massive galaxy" mass-tension discussions; the work will be of immediate interest in astrophysics and may resonate across cosmology/galaxy-formation modeling.
- The manuscript should moderate language implying "first robust IMF measurements beyond the local Universe," because prior non-local IMF constraints exist. If the authors mean "first with JWST ultra-deep NIRSpec IMF-sensitive features at this S/N plus full-spectrum optical–NIR modeling," that novelty should be stated precisely.

Data & methodology:
- Data quality appears exceptional (ultra-deep JWST spectroscopy; deep LEGA-C), and the wavelength coverage is well chosen for separating age/abundance constraints (optical) from dwarf/giant-sensitive diagnostics (NIR).
- The methodology is broadly appropriate (full-spectrum fitting; multi-parameter inference; tests with alternative SFH; feature masking tests), and the paper acknowledges several known pitfalls (TiO region exclusion; NaD ISM sensitivity; Wing–Ford detector gap cases).
- Reproducibility is not yet at best practice: "data products available upon reasonable request" is insufficient for a result this sensitive to reduction choices and fitting configuration. For a robust revision, the authors should publicly release (or deposit in an archive):
- reduced IMFERNO 1D spectra used for fitting, masks, and per-galaxy resolution/LSF curves,
- exact fitting setup (wavelength windows, polynomial order, priors, alf version/commit),
- the flux-rescaling procedure to photometry in a fully reproducible form.
- The aperture/PSF/slit-orientation mismatch between LEGA-C and NIRSpec is a central methodological concern: joint fits assume that the combined spectra correspond to a single representative stellar population. The current correction focuses on velocity dispersion offsets; it does not fully address potential stellar population/IMF gradients combined with different spatial weighting. A more quantitative aperture-mismatch/gradient sensitivity assessment is needed.

Appropriate use of statistics and treatment of uncertainties:
- The inference relies on MCMC posteriors, which is appropriate, but the manuscript should be explicit that reported uncertainties are statistical-only unless systematics are included (currently, systematics appear largely unquantified).
- For extreme alpha_IMF values (and "mass x3–4" claims), systematic SPS uncertainties (stellar libraries, response functions, isochrones, continuum normalization, LSF modeling) can be comparable to or larger than the statistical error bars; the revision should include an error-budget style discussion and/or sensitivity tests that bound plausible systematic shifts in alpha_IMF.
- Please ensure all error bars are defined in every relevant figure legend and that probability language (e.g., ">2sigma") is

consistently tied to clearly stated uncertainty definitions (posterior standard deviation, credible intervals, etc.).
- Provide convergence diagnostics or a brief statement on how MCMC convergence was assessed (mixing, effective sample size, chain stability).

Conclusions:
- The conclusion that some z=0.7 massive quiescent galaxies show spectral evidence consistent with bottom-heavy low-mass IMFs is plausible and potentially robust, but it is not yet demonstrated to be robust against the most important systematic/model assumptions.
- The strongest interpretive claim—that the oldest systems imply z>6 "impossibly early" galaxies are likely 3–4x more massive—currently reads over-extended. It depends on (i) fixed IMF assumptions (low-mass cutoff; fixed high-mass slope/upper limit), (ii) extrapolation across epochs and assembly histories, and (iii) unquantified systematics. This should be rewritten in clearly conditional language and supported with quantitative sensitivity bounds.

Suggested revisions:
1. IMF parameterization stress tests (critical)
- Quantify sensitivity of alpha_IMF and inferred stellar mass amplification to the fixed low-mass cutoff (e.g., m_c = 0.05, 0.08, 0.10 M☉).
- Explicitly state the IMF upper-mass limit used in alf and estimate its impact on M/L and alpha_IMF.
- Quantify the effect of fixing the high-mass slope above 1 Msun to Salpeter: provide "what-if" bounds for plausible alpha_3 values, or clearly bound the conclusions to the fixed-alpha_3 assumption.

2. IMF–abundance degeneracy demonstration
- Provide posterior covariance (corner) plots for representative galaxies (most bottom-heavy + near-MW) showing alpha_IMF (or alpha1/alpha2) vs [Na/Fe], [Ca/Fe], [Fe/H], and age.
- Add controlled-abundance tests (e.g., informative priors or fixed-abundance runs) to show how α_IMF shifts when Na/Ca constraints are tightened/loosened.
- Report which spectral regions most constrain Na/Ca independently of IMF, and show the impact of masking NaD for the key galaxies.

3. Systematics and robustness suite
- Vary continuum-normalization choices (polynomial order) and key wavelength windows (including/excluding TiO region; excluding individual IMF-sensitive features) and summarize alpha_IMF stability.
- More quantitative handling of aperture mismatch/gradients: demonstrate that plausible gradients cannot drive the reported alpha_IMF extremes, or present results as aperture-specific.

4. Virial-mass comparison moderation
- Reframe virial-mass agreement as a plausibility check rather than strong validation; clarify the limitations (rotation, anisotropy, dark matter, slit losses, aperture mismatch).
- Address the "unphysical" outlier transparently (evidence for face-on inclination; sensitivity of conclusions to excluding it).

5. Extrapolation language
- Rewrite z>6 implications conditionally ("if similarly bottom-heavy IMFs were already in place at formation…") and provide a bounded range rather than a single amplification factor.

References: appropriateness and completeness
- The manuscript generally cites relevant IMF spectroscopy and dynamical/lensing context.
- If claiming first/most robust beyond-local measurements, please more carefully position against prior non-local IMF constraints (and ensure the novelty statement is precise and defensible).
- Ensure that key controversies around Na/Ca-based IMF inferences and SPS limitations in the NIR are adequately referenced and discussed in the context of your degeneracy-control evidence.

Clarity and context
- The abstract is generally clear and motivating, but it currently reads more definitive than the evidence warrants regarding downstream implications (mass amplification; cosmological tension).
- Improve clarity by surfacing the most consequential assumptions (fixed low-mass cutoff; fixed high-mass slope; feature gaps/masks; aperture mismatch) earlier in the main text, not primarily in Methods/limitations.
- Separate "measurement" statements from "extrapolation" statements more explicitly.

Version 1:

Reviewer comments:

Reviewer #2

(Comments for the Author)
I thank the authors for their thorough treatment of my comments and recommendations, and can now recommend the paper for publication.

Reviewer #3

(Comments for the Author)
I thank the authors for their detailed response and for the substantial revisions made to the manuscript. The paper is significantly improved. In particular, the authors have addressed many of the main concerns raised in the first round: the IMF parameterization is now stated more explicitly, the lower- and upper-mass limits are clarified, several robustness tests have been added, the aperture-specific nature of the measurements is discussed, and the high-redshift interpretation is phrased more conditionally.

I believe the manuscript is now close to being suitable for publication in Nature Astronomy. However, I still recommend that the authors make a few additional clarifications, mostly concerning wording, scope, and literature framing.

Main remaining points

1. The manuscript should consistently refer to constraints on the low-mass IMF

The authors now state the IMF parameterization clearly in the Methods: the low-mass cutoff is fixed at 0.08\,M_\odot, the two low-mass slopes are varied, and the high-mass slope from 1\,M_\odot to 100\,M_\odot is fixed to the Salpeter value. They also explicitly note that the method is sensitive to the relative fraction of low- to high-mass stars, not to the full IMF shape at formation.

This is a very useful clarification. However, I still think the language throughout the manuscript should be made more consistently precise. In several places, the text refers broadly to “the IMF” being measured or constrained. Given the modelling setup, it would be more accurate to refer to the low-mass IMF, the low-mass end of the IMF, or the dwarf-to-giant sensitive IMF normalization, depending on context.

This distinction is important because the high-mass IMF slope at formation is fixed in the models and remains unconstrained by the spectra. This is not a minor semantic issue: the implications for stellar masses, remnant fractions, early star formation, chemical enrichment, and galaxy-formation tensions depend on the behaviour of the IMF over the full stellar mass range, not only the low-mass end.

I therefore recommend that the authors check the manuscript carefully and ensure that claims about “IMF measurements” are qualified where appropriate as constraints on the low-mass IMF.

2. The broader variable-IMF literature context should be treated more carefully

The revised manuscript now includes an important caveat that a bottom-heavy low-mass IMF does not exclude a shallower high-mass slope, and that such a possibility could reduce high-redshift stellar masses. This is a useful addition.

However, I find the current framing still somewhat too narrow. The idea that the IMF may vary with environment, star-formation conditions, metallicity, density, or cosmic epoch - and that different stellar-mass regimes may behave differently - has a long history in the literature. The current text risks over-emphasising one recent phenomenological framing while under-representing the broader body of work on variable IMFs.

I am not asking for an exhaustive review, and I do not think the paper should become a literature survey. However, if the manuscript makes broad statements about reconciling bottom-heavy low-mass IMFs with high-redshift galaxy-formation tensions, the authors should acknowledge that this belongs to a wider and older discussion of environmentally dependent IMF behaviour. A more neutral wording would be preferable.

For example, along the lines of:

More generally, a bottom-heavy low-mass IMF does not exclude the possibility that the high-mass IMF at formation differed from the Milky Way form. Variable IMF scenarios in which different stellar-mass regimes respond differently to the physical conditions of star formation have been explored in a broad literature, and such behaviour could affect inferred stellar masses and their implications for galaxy-formation models.

This would avoid implying that one recent named IMF prescription is the conceptual origin of this possibility. I would also suggest avoiding unnecessary emphasis on a specifically named “concordance IMF” formulation, since the relevant physical point is simply that different stellar-mass regimes of the IMF may vary differently with star-formation conditions, an idea that is not unique to one recent parameterization and to my opinion there is no need to introduce new names for every slope value of the IMF.

3. High-redshift implications are improved, but should remain explicitly conditional throughout

The authors have substantially improved the high-redshift interpretation. The revised text now states:

“If similarly bottom-heavy IMFs were already in place when these galaxies formed…”

and gives an amplification factor of $\sim 4 \pm 1$, rather than presenting the increase as an unconditional result.

This is a major improvement. Still, the abstract and summary remain quite strong. For example, the abstract states that the oldest galaxy is likely a descendant of JWST’s “impossibly early” galaxies and that the latter may have had similarly bottom-heavy IMFs increasing their masses by $\sim 4 \pm 1$.

This is acceptable if phrased carefully, but the authors should ensure that the distinction between direct measurement and extrapolation is clear everywhere. The direct measurement is at $z \sim 0.7$, in aperture-specific spectra. The implications for $z > 4$–6 systems remain conditional on assumptions about formation histories, stellar mixing, IMF gradients, and the persistence of the low-mass IMF signature.

I recommend a final pass through the abstract, main text, and summary to ensure that all statements about high-redshift galaxies remain conditional.

4. Aperture specificity and IMF gradients are now discussed well, but should remain visible

The new aperture discussion is very helpful. The authors now state that the spectra probe approximately the inner half-light radius, that local early-type galaxies can have bottom-heavy cores and more MW-like outskirts, and that if the IMF becomes fully MW-like beyond $1R_e$, the correction factors could be closer to $\alpha_{\rm IMF}/2$. They also explicitly state that the IMF constraints should be considered aperture-specific.

This is a strong addition and addresses an important concern. I would only suggest keeping this point visible in the conclusions as well, because it is central to the interpretation of the stellar mass corrections. A short phrase such as “within the apertures probed by the spectra” would help avoid over-generalising the result to the entire galaxy.

5. Virial mass comparison should remain a plausibility check, not validation

The revised manuscript now frames the virial-mass comparison more carefully, as a plausibility diagnostic rather than a definitive validation. This is much improved.

However, the phrase “reinforcing the physical feasibility” is still fairly strong. I would suggest maintaining cautious language throughout this section, because the virial masses remain approximate and depend on velocity dispersion corrections, rotation, anisotropy, aperture effects, slit losses, and dark matter assumptions. The comparison is useful and reassuring, but it should not be read as an independent confirmation of the inferred IMF normalization.

A wording such as “is consistent with the physical plausibility of” would be slightly safer than “reinforcing”.

Minor comments

The statement “first robust constraints on the IMF beyond the local Universe from full-spectrum models” is much improved compared to the earlier version. I suggest retaining the qualifier “from full-spectrum models” wherever this novelty claim is made.

The Data and Code Availability sections are now much stronger. The plan to release the spectra as fitted with alf, resolution curves, wavelength windows, configuration files, the alf commit, and the photometric rescaling procedure is appropriate and important.

The authors now state that all quoted uncertainties are statistical. This should remain clear in figure captions and wherever extreme $\alpha_{\rm IMF}$ values or mass amplification factors are discussed.

The added corner plots and abundance-degeneracy tests described in the response are valuable. The manuscript should ensure that the main text sufficiently points readers to these tests, since IMF–abundance degeneracy is one of the central concerns for this type of analysis.

The discussion of systematic uncertainties in alf, stellar libraries, response functions, continuum normalisation, and LSF modelling is now substantially improved. I would keep the tone cautious, especially because the reported uncertainties themselves do not include all these systematics.

Reports/Response

We thank the referees very much for their careful reading of the manuscript and their useful feedback which has improved our analysis and manuscript. Please find our response to their comments below.

Reviewer #2 (Comments for the Author):

I have read with great interest the authors' manuscript, which is well presented and of clear importance, describing a highly novel and interesting JWST observing programme. This paper clearly merits publication in Nature Astronomy, however I have a few concerns that I would recommend the authors address first, in order to strengthen their conclusions.

Major points:

Your conclusions re the masses of "impossibly early galaxies" are largely based on the assumption that the IMFs you measure in the central regions of your objects are representative of their whole stellar populations. However, it can be seen from the cutouts presented in Fig. 1 that the NIRSpec shutters only cover a small fraction of the cores of these galaxies. I think it would be useful to include a discussion of what fraction of the light from/mass of your galaxies falls into your slitlets (in particular the regions used for your spectral extractions) and how this might impact your conclusions, e.g., if the IMFs are more MW-like outside these central regions. A new column or columns in Table 2 containing this information would be useful. A contrary point to discuss is obviously the much-more-compact nature of the z>3 sources, which presumably grow via mergers to become your oldest z=0.7 objects, so maybe the central regions of your sources are the right places to be looking. What physical scales do your slitlets cover compared to r_e for typical z>3 quiescent galaxies? Clearly IFU follow-up would be of interest here to properly settle this point. Papers by Padraig Alton+17+18 might be interesting to discuss here?
Thank you for raising this point. We agree that this is an important assumption that we have made in this work, which we did not explicitly state. To address this, we have added a paragraph (page 6, line 158) to discuss possible effects of IMF gradients on our application of alpha_IMF to the entire galaxy in Figure 3. See also our response to Referee #3's comments below. We have also added references to the Alton+2017 paper. We indeed plan to propose for follow-up IFU in the future.

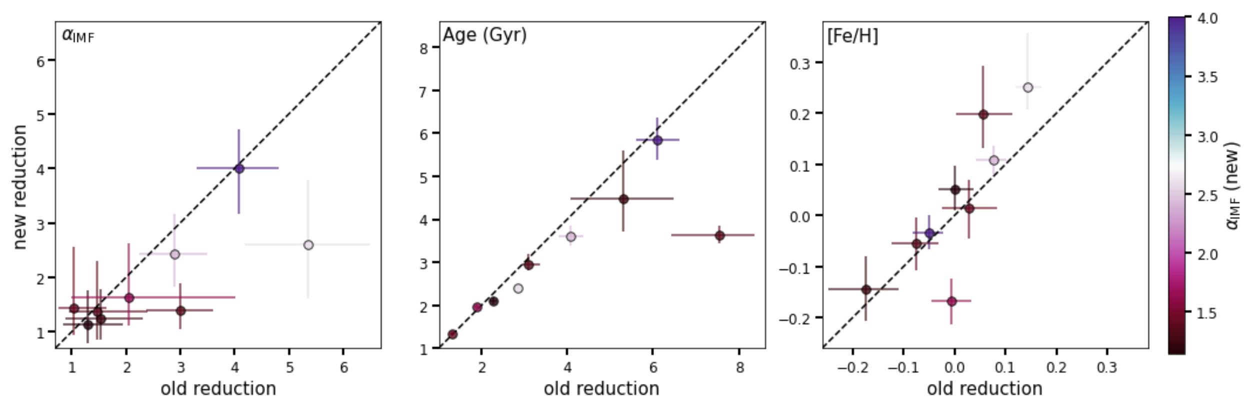
Our spectra cover approximately the inner Re of our galaxies, which we have now noted in the text. However, we have not added a column with the fraction of light in the slit, as we have custom shutter locations (i.e. often offset from the centre) and lengths. As a result of these factors, this fraction does not give us a lot of information about the region of the galaxy that has actually been observed. For example, for several galaxies, the slitlets in one direction go out to several Re, while for others, our slits are not even centred, and we do not capture the most central part of the galaxy.

The pipeline and crds versions used to reduce the data products used in this paper are quite out of date (crds v1322 is now more than a year old). This is of particular importance because there were major updates made to the NIRSpec calibration files in Sep 2025, e.g., new flat field and dark reference files. In this context, I would strongly advise the authors to re-reduce their data with the most recent pipeline and crds versions to make sure that their results are not sensitive to these updates.
We thank the referee for letting us know about this, our original reduction was done in the summer of 2025, before the reference files were updated. Following your suggestion, we have re-run the analysis with the re-reduced data. We have also updated our LSF corrections using improved light models. Finally, we have improved our velocity dispersion correction method (as we found an issue with our previous correction), see the Methods section.

In general, our results are consistent and our conclusions are mostly unchanged. Please see the attached one-to-one plot comparing some key recovered parameters from the original reduction and the new reduction. However, we note that these updates have resulted in two notable changes:

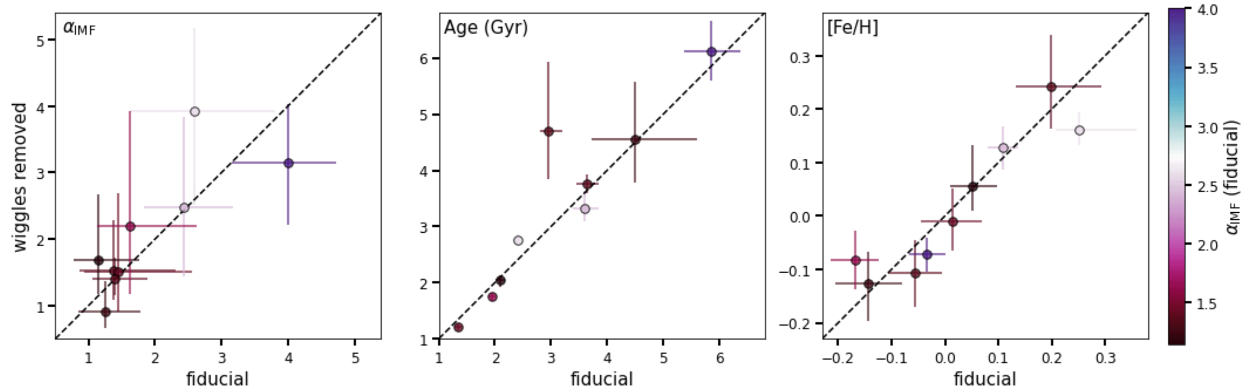
- *1158527 is no longer extremely bottom-heavy in Figure 2 and does not fall in the unphysical region in Figure 3. We have thus removed any discussion of 1158527 as an outlier.*
- *We now only have one galaxy that is very old and may be a descendant of JWST's impossibly early galaxies. We have updated our conclusions and the text to reflect this.*

We have updated the entire text and analysis to use the new reduction. We thank the referee for making this valuable suggestion.



Related to the above, given the extremely high SNR of the NIRSpec continuum data, have the authors considered the effects of a.) NIRSpec flicker noise introduced during detector readout? Do you use the clean_flicker_noise step in the level 1 pipeline - if so what mode and parameters, if not I think this would be worth testing to make sure it doesn't impact the results; b.) NIRSpec extraction "wiggles" due to undersampling of the spatial PSF (e.g., Perna+23, Dumont+25, Leung+26). Both of these feel like they would be particularly important in this case of trying to tease out extremely subtle continuum features.
We appreciate the referee's concern here. We test correcting the wiggles in our spectra (using the algorithm presented in Cheng+2025) and compare our refitted spectra. Please see the

attached one-to-one plot comparing our fiducial fits to our fits with the wiggles corrected. We note that in this figure we compare fits without the velocity dispersion correction. In general, we find that the wiggle correction does not significantly affect our results and does not change our conclusions.



We also note that we and others have shown in previous work that this effect is only significant for resolved spectra, as the wiggles average out in integrated spectra (see Cheng+2025 for detailed tests comparing wiggle-corrected spectra to uncorrected spectra). Moreover, we continuum-normalize the spectra, and as the wiggles are much broader than the features that we're examining, they do not have a significant effect. We have added a paragraph in the Methods to summarize the presence and effect of the wiggles, as well as our test (page 17, line 545).

For the flicker noise, we do not use the pipeline correction, but we perform our own 1/f correction, see the Methods section and Slob+24. We find no residual 1/f features. We additionally expect this effect to have more of an impact on the continuum, and as we continuum-normalize the spectra, we also do not expect this to have a significant effect on our results.

Whilst I appreciate that ALF and the models it contains are currently unique in their capability to fit the parameters of interest for this paper, it always concerns me that this makes it extremely difficult to evaluate the likely level of systematic uncertainty in results such as these. For example, Byrne & Stanway (2023) and Jones, Byrne & Stanway (2025) show several tests demonstrating that different sets of stellar spectral model libraries predict quite different strengths for commonly studied rest-optical absorption features at the same age and metallicity. Can the authors comment in the paper on the likely/possible/plausible level of systematic uncertainty on their results due to modelling choices? If this isn't possible, I think it would at least be sensible to acknowledge the existence of this issue.
We thank the referee for raising this, we agree it is an important concern. It is not possible to quantify the level of systematic uncertainty due to our choice of spectral library or specific modeling code in this paper, as there are currently no comparable codes available to constrain the IMF (as the referee also points out). However, we have added a caveat paragraph discussing the possibility of these uncertainties as a result of our reliance on alf (page 8, line

209). Additionally, we note that Lick index methods and full-spectrum modeling have been found to give similar results within uncertainties for local galaxies, see, e.g. Villaume+2017b, Lonoce+2021, Cheng+2023.

Minor points:

Line 64: emotive language "ambitious" not really the right style here in my opinion, people will draw their own conclusions from the data you present. Again on line 209.
We have removed this.

Line 149: The sentence that starts here is a little confusing, I would clarify that your "mass offsets" are virial-to-stellar mass offsets, presuming this is what you mean?
We have added this clarification.

Line 182: Not sure I quite follow the logic here? Surely only applies if stars have started dying within the mass-range that gives rise to the features you're measuring - is this the case for these objects? A bit more clarification would be useful on this point I think.
We have added some clarification to this point, please see the text. In particular, with this method we measure the relative fraction of low-mass stars, and thus excess low-mass stars have a slightly larger contribution if fewer stars with > 1 solar masses are still alive (as would be the case for older stellar populations). However, we have estimated that this effect would be < 10% for the ages of the galaxies in our sample.

Line 200: clarify "above" 1 solar mass instead of saying beyond?
We have added this clarification

Reviewer #3 (Comments for the Author):

Key results:
- Presents ultra-deep JWST/NIRSpec MSA spectroscopy from IMFERNO combined with deep optical LEGA-C spectra for 9 massive quiescent galaxies at z = 0.7, enabling unusually broad rest-frame wavelength leverage for IMF work.
- Uses full-spectrum fitting with alf allowing simultaneous inference of age, multi-element abundances, and low-mass IMF slopes, and reports bottom-heavy IMF normalizations in several galaxies (alpha_IMF_low_mass up to 5).
- Connects inferred IMF-driven M/L shifts to revised stellar masses and compares these against virial-mass estimates as a plausibility check.

Validity:
- No single issue prohibits publication in principle, but several unquantified modeling/assumption sensitivities currently prevent the strongest claims from being considered fully supported at Nature's standard.

- The most consequential validity risks are: fixed IMF low-mass cutoff and fixed high-mass slope, incomplete quantification of IMF–abundance degeneracies, and limited accounting of SPS/systematic uncertainties relative to the reported extreme alpha_IMF values.
- The environmental/selection context (one pointing associated with the “COSMOS Wall”) plus n=9 also limits generalizability; this is not fatal but must be made explicit and considered when generalizations are made.

We thank the referee for raising this point about the environmental/selection context. We have added a brief discussion to be explicit about this (page 7, line 204), where we discuss future directions.

Originality and significance:
- The observational combination (very high-S/N JWST near-IR absorption features + LEGA-C optical constraints for the same galaxies) is novel and of broad interest to galaxy evolution and stellar population communities.
- The topic is timely given JWST-era “early massive galaxy” mass-tension discussions; the work will be of immediate interest in astrophysics and may resonate across cosmology/galaxy-formation modeling.
- The manuscript should moderate language implying “first robust IMF measurements beyond the local Universe,” because prior non-local IMF constraints exist. If the authors mean “first with JWST ultra-deep NIRSpec IMF-sensitive features at this S/N plus full-spectrum optical–NIR modeling,” that novelty should be stated precisely.

We have moderated the language and have added the precise novelty. In particular, in the abstract, we note that these are the first robust IMF constraints in distant galaxies from full-spectrum models specifically. In the conclusions, we further elucidate this by stating “for the first time we robustly measured the IMF in a sample of massive, quiescent galaxies beyond the nearby Universe, using full-spectrum modeling of extremely deep \textit{JWST}/NIRSpec-MSA data.”

Data & methodology:
- Data quality appears exceptional (ultra-deep JWST spectroscopy; deep LEGA-C), and the wavelength coverage is well chosen for separating age/abundance constraints (optical) from dwarf/giant-sensitive diagnostics (NIR).
- The methodology is broadly appropriate (full-spectrum fitting; multi-parameter inference; tests with alternative SFH; feature masking tests), and the paper acknowledges several known pitfalls (TiO region exclusion; NaD ISM sensitivity; Wing–Ford detector gap cases).
- Reproducibility is not yet at best practice: “data products available upon reasonable request” is insufficient for a result this sensitive to reduction choices and fitting configuration. For a robust revision, the authors should publicly release (or deposit in an archive):
- reduced IMFERNO 1D spectra used for fitting, masks, and per-galaxy resolution/LSF curves,

We agree that this is important. We will publish all of the stated data products at the time of publication. For now, we have added sentences in the Data Availability and Code Availability statements indicating that we will do this.

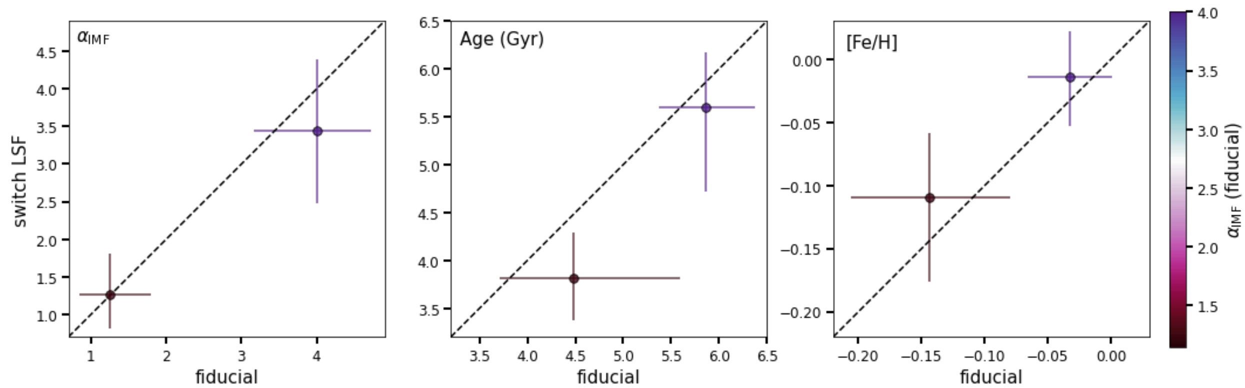
- exact fitting setup (wavelength windows, polynomial order, priors, alf version/commit),

We have added the alf commit version in the Code Availability Statement. We will also publish our alf configuration files for each object. The wavelength windows will be included in the files that contain the spectra for fitting. Alf determines the polynomial order based on the wavelength windows during fitting, so we have not included this here.
- the flux-rescaling procedure to photometry in a fully reproducible form.
Same as above, we will post this upon publication.

- The aperture/PSF/slit-orientation mismatch between LEGA-C and NIRSpec is a central methodological concern: joint fits assume that the combined spectra correspond to a single representative stellar population. The current correction focuses on velocity dispersion offsets; it does not fully address potential stellar population/IMF gradients combined with different spatial weighting. A more quantitative aperture-mismatch/gradient sensitivity assessment is needed.
We thank the referee for pointing this out, this is indeed an important concern. We have added a discussion about IMF gradients, also in response to Referee #2's comments, please see the response above. Regarding stellar population gradients (i.e. age and elemental abundances), Cheng+2024 found the gradients in the LEGA-C galaxies (some of which are included in our sample here) to be relatively weak. We have also tested fitting the LEGA-C and JWST spectra separately, and find that the ages and [Fe/H] abundances generally agree, demonstrating that the aperture mismatches are not likely to have a significant effect. However, to bring some more focus to this issue, we have also added a sentence about this in the main body. Additionally, we have stated that our IMF measurements are aperture-specific (based on the referee's comment below). A more quantitative assessment of the effect of gradients beyond the scope of this paper. However, spatially resolved IMF measurements in the IMFERNO galaxies will be the subject of future work.

Appropriate use of statistics and treatment of uncertainties:
- The inference relies on MCMC posteriors, which is appropriate, but the manuscript should be explicit that reported uncertainties are statistical-only unless systematics are included (currently, systematics appear largely unquantified).
We have added this statement to the Methods.

- For extreme alpha_IMF values (and "mass x3–4" claims), systematic SPS uncertainties (stellar libraries, response functions, isochrones, continuum normalization, LSF modeling) can be comparable to or larger than the statistical error bars; the revision should include an error-budget style discussion and/or sensitivity tests that bound plausible systematic shifts in alpha_IMF.
We agree with the referee, and we have added a caveat paragraph discussing the possible systematic uncertainties as a result of our modeling choices. In summary, we have performed several fitting tests as suggested by the referees. In particular, we find that our updated data reduction confirms the factor of 4 increase in mass. Furthermore, as discussed throughout this response, our constraints are robust to the tests suggested below (i.e. changing the order of the polynomial used for continuum normalization, fixing the low-mass cutoff to a different value, switching the LSF curves for two galaxies, see attached plot). Additionally, our virial mass comparison provides an excellent plausibility check.



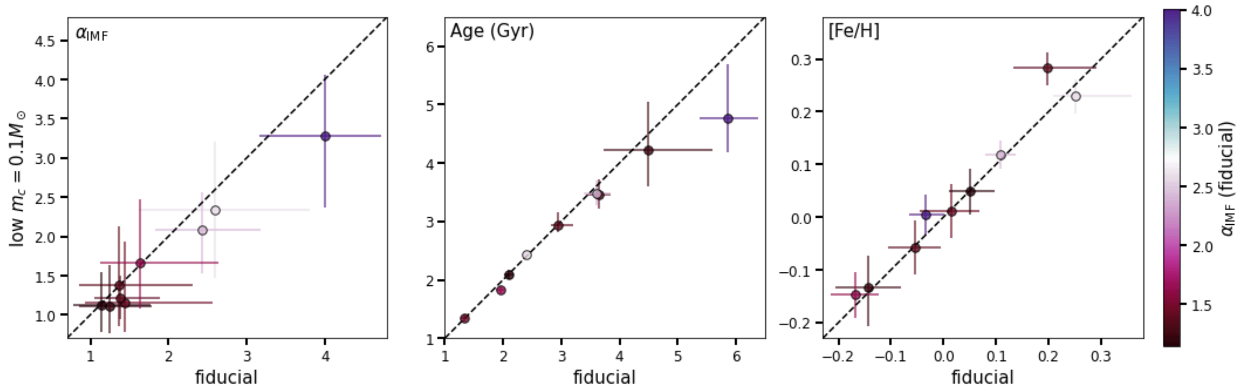
- Please ensure all error bars are defined in every relevant figure legend and that probability language (e.g., ">2sigma") is consistently tied to clearly stated uncertainty definitions (posterior standard deviation, credible intervals, etc.).
We have updated this to refer to standard deviations instead of sigmas, and have ensured that we state the error bar definitions in every figure caption.

- Provide convergence diagnostics or a brief statement on how MCMC convergence was assessed (mixing, effective sample size, chain stability).
We visually examine the chains to ensure that they are well mixed and converged. We have added a short statement about this in the methods.

Conclusions:
- The conclusion that some z=0.7 massive quiescent galaxies show spectral evidence consistent with bottom-heavy low-mass IMFs is plausible and potentially robust, but it is not yet demonstrated to be robust against the most important systematic/model assumptions.
- The strongest interpretive claim—that the oldest systems imply z>6 "impossibly early" galaxies are likely 3–4x more massive—currently reads over-extended. It depends on (i) fixed IMF assumptions (low-mass cutoff; fixed high-mass slope/upper limit), (ii) extrapolation across epochs and assembly histories, and (iii) unquantified systematics. This should be rewritten in clearly conditional language and supported with quantitative sensitivity bounds.
We have softened the language around our conclusions to be conditional.

Suggested revisions:
1. IMF parameterization stress tests (critical)
- Quantify sensitivity of alpha_IMF and inferred stellar mass amplification to the fixed low-mass cutoff (e.g., m_c = 0.05, 0.08, 0.10 M☉).
We agree that this is a useful test. Unfortunately, the alf code does not allow the user to fix the low-mass cutoff to a specific value, however there is an option to allow for variation in the low-mass cutoff in the fit. Thus, we have performed a preliminary test, where we have fit our sample with the variable low-mass cutoff option, but "fix" the low-mass cutoff by constraining it to vary only in a very small range around 0.1 Msun (one of the limits suggested by the referee). See the attached plot. Our constraints are presumably robust to the assumed low-mass cutoff. We note we have not explicitly allowed for variation in the low-mass cutoff, which would require

even higher quality data than we have here (S/N > 300/AA, see Conroy+2017). We have added a brief discussion about the possible effect of a variable low-mass cutoff in the Methods (page 17, line 533).



We have only tested fixing the low-mass cutoff to 0.1 Msun as the lower bound on the prior for the low-mass cutoff in alf is 0.08Msun. This is due to the fact that the isochrones only extend to this mass (see Conroy & van Dokkum 2012a), so fixing the low-mass cutoff to anything below this (i.e. the suggestion of 0.05Msun) would be extrapolating.

- Explicitly state the IMF upper-mass limit used in alf and estimate its impact on M/L and alpha_IMF.

We thank the referee for raising this point. We have added the upper mass limit before Eq. 1 in the Methods. We have added a short discussion of the impact of the upper-mass limit in the Methods section (page 17, line 527). In summary, the assumed upper mass limit should not have a strong effect on M/L or alpha_IMF.

- Quantify the effect of fixing the high-mass slope above 1 Msun to Salpeter: provide “what-if” bounds for plausible alpha_3 values, or clearly bound the conclusions to the fixed-alpha_3 assumption.

We agree with the referee that this is an important test, but unfortunately it is not possible to change the slope of the IMF beyond 1 Msun in alf.

2. IMF–abundance degeneracy demonstration

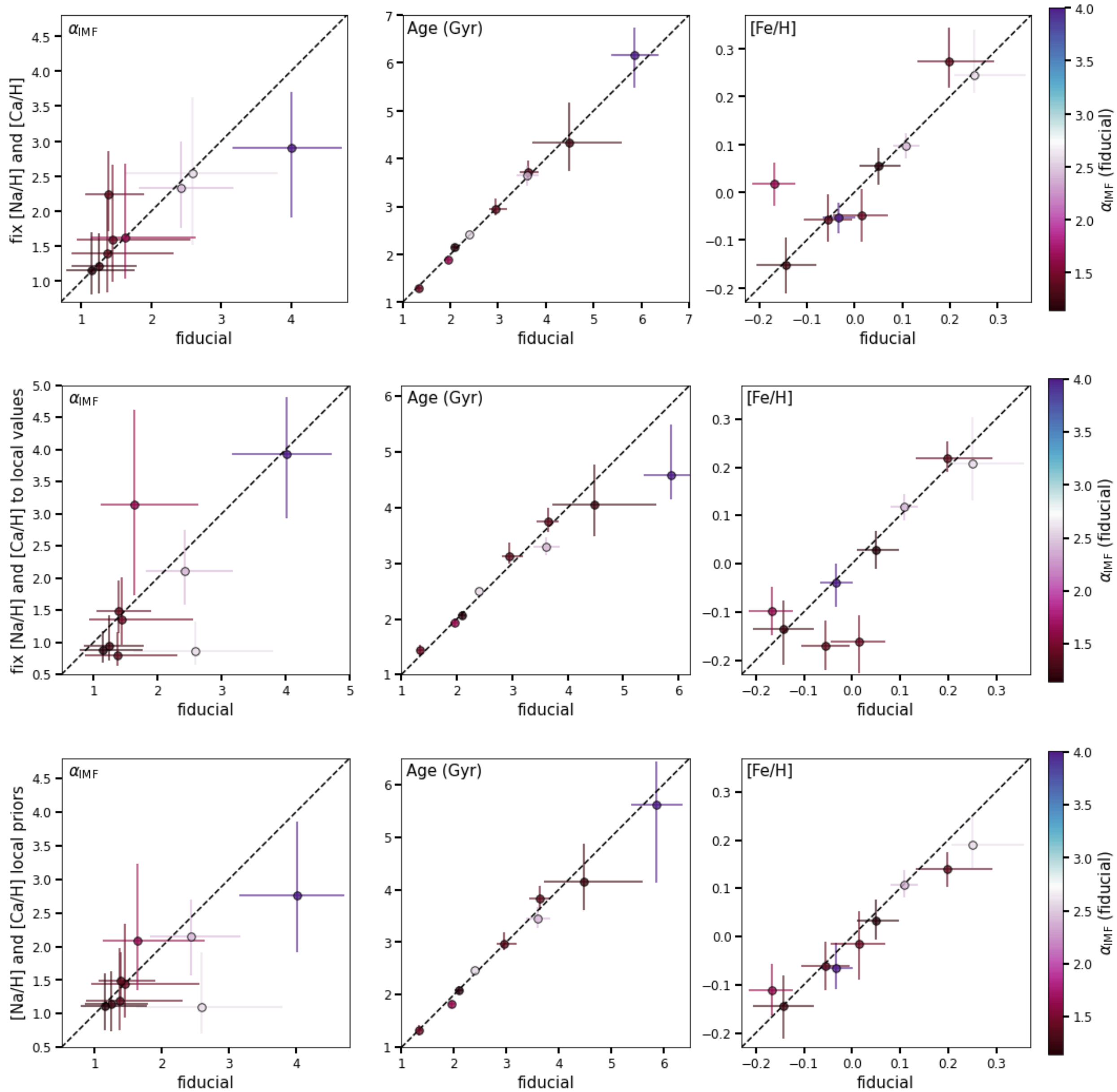
- Provide posterior covariance (corner) plots for representative galaxies (most bottom-heavy + near-MW) showing alpha_IMF (or alpha1/alpha2) vs [Na/Fe], [Ca/Fe], [Fe/H], and age.

We thank the referee for this suggestion. We have added corner plots for a bottom-heavy galaxy (1134272) and a MW-like galaxy (1146989) in the Extended Data.

- Add controlled-abundance tests (e.g., informative priors or fixed-abundance runs) to show how α_IMF shifts when Na/Ca constraints are tightened/loosened.

We have tested fixing the Na and Ca abundances to our best-fitting values, fixing them to the average values of massive early type galaxies at z = 0, and bounding the priors to the minimum and maximum values of massive early type galaxies at z = 0 (using the SDSS results from

Conroy & van Dokkum 2012 for the latter two tests). See attached plots. Our results are robust to the Na and Ca constraints.



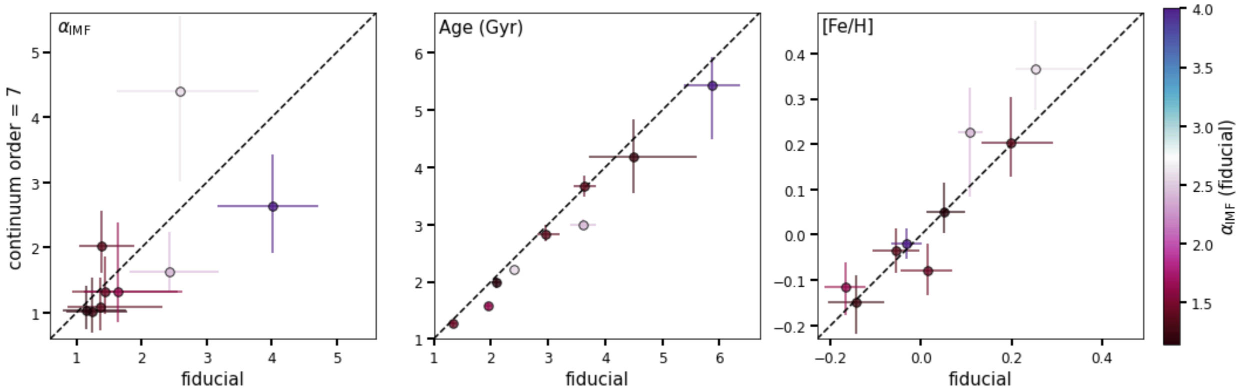
- Report which spectral regions most constrain Na/Ca independently of IMF, and show the impact of masking NaD for the key galaxies.

We have included a one-to-one comparison for when we include vs. mask NaD for our whole sample in the Extended Data section. We have also added a discussion of which regions most constrain Na and Ca in the Methods, see also Conroy & van Dokkum 2012b.

3. Systematics and robustness suite

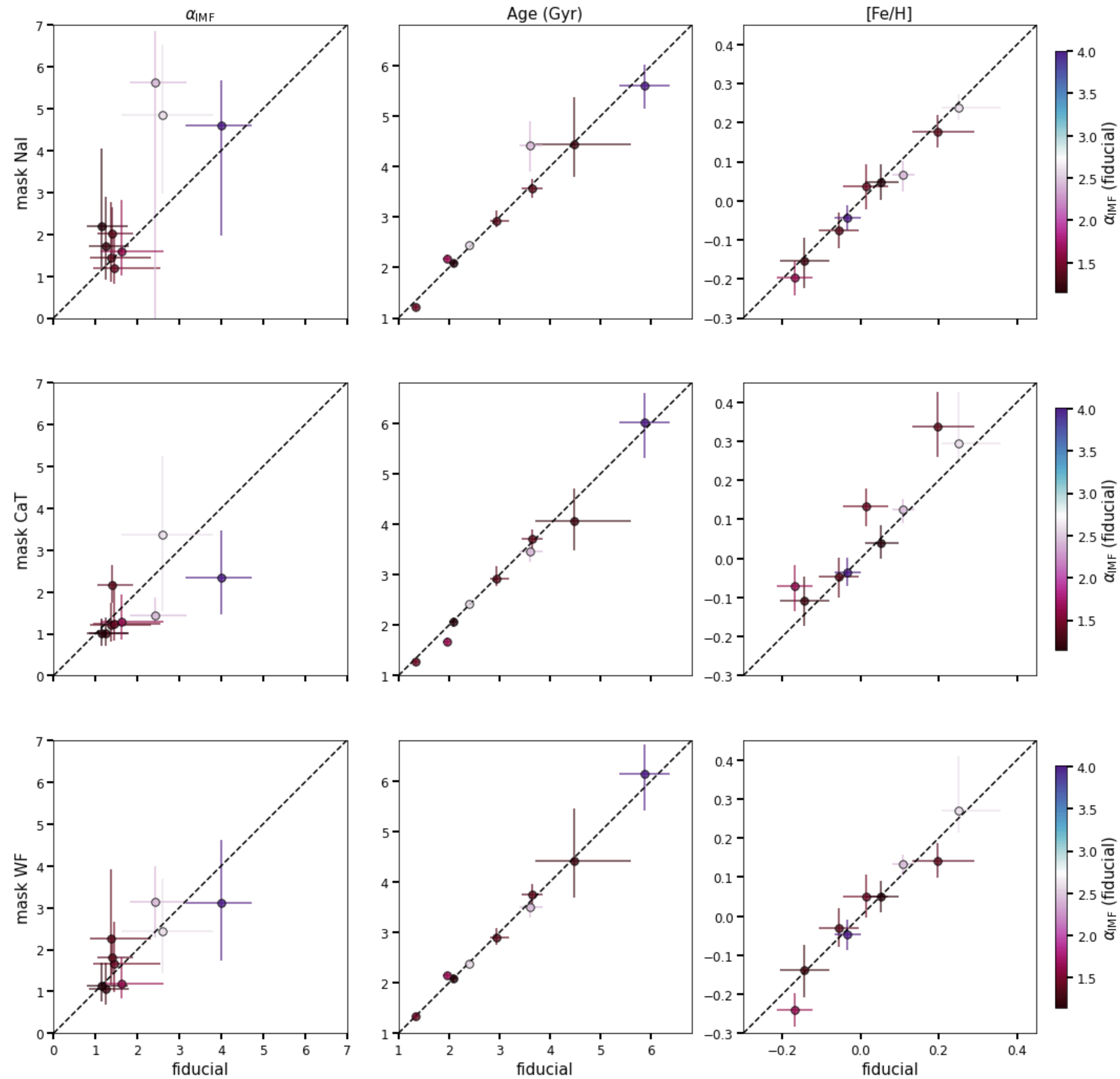
- Vary continuum-normalization choices (polynomial order) and key wavelength windows (including/excluding TiO region; excluding individual IMF-sensitive features) and summarize alpha_IMF stability.

Continuum normalization: we agree that this is an important test. We have thus tested fixing the order of the polynomial to n = 7 across the entire spectrum for each of the galaxies in our sample. See the attached plot. Our constraints are generally robust to the choice of polynomial order, although there are some small differences. We have added a short discussion about this in the Methods section (page 17, line 540). We also note that this has been tested in the original alf papers. In particular, in Conroy & van Dokkum 2012b, they demonstrated that variation in the polynomial order by +/- 2 induces changes to the spectrum at the < 0.1% level. Additionally, see Appendix B of Conroy+2018, where they tested changing the polynomial order and the method in which the polynomial was applied. Their results were also generally insensitive to these choices.



Including/excluding TiO: we have tested this, as stated in the Methods (page 18, line 588). Our constraints are robust to including/excluding the TiO bands.

Excluding individual IMF-sensitive features: we thank the referee for suggesting this test. We have performed this, see the attached plot. Our results are largely robust to these tests, although we note that there are some small differences. Crucially, this test has also been performed for alf in Conroy & van Dokkum 2012b and Cheng+2023. We have added some discussion of these tests to the text (page 18, line 612).



- More quantitative handling of aperture mismatch/gradients: demonstrate that plausible gradients cannot drive the reported alpha_IMF extremes, or present results as aperture-specific.
Please see the response to the comments above.

4. Virial-mass comparison moderation
- Reframe virial-mass agreement as a plausibility check rather than strong validation; clarify the limitations (rotation, anisotropy, dark matter, slit losses, aperture mismatch).
We thank the referee for this suggestion. We have stated that the virial mass comparison is a plausibility check. We have discussed rotation, anisotropy, dark matter, and slit losses in the text. In particular, we note that these galaxies may be rotating, which could influence their virial masses. Detailed anisotropic Jeans modeling will be performed in an upcoming study (Slob et

al. in prep.). In response to the previous comments from both referees, we have also added discussion of the apertures of our measurements.

- Address the “unphysical” outlier transparently (evidence for face-on inclination; sensitivity of conclusions to excluding it).

We thank the referee for suggesting this. We note that with the updated reduction, 1158527 is no longer an outlier, see the response to Referee #2’s comment above.

5. Extrapolation language
- Rewrite z>6 implications conditionally (“if similarly bottom-heavy IMFs were already in place at formation…”) and provide a bounded range rather than a single amplification factor.

We have rewritten our conclusions to be more conditional and have added an error bar to the amplification factor.

References: appropriateness and completeness
- The manuscript generally cites relevant IMF spectroscopy and dynamical/lensing context.
- If claiming first/most robust beyond-local measurements, please more carefully position against prior non-local IMF constraints (and ensure the novelty statement is precise and defensible).

We have added the precise novelty, see response to the comment above.

- Ensure that key controversies around Na/Ca-based IMF inferences and SPS limitations in the NIR are adequately referenced and discussed in the context of your degeneracy-control evidence.

We have added these references and discussions, please see responses to other comments.

Clarity and context
- The abstract is generally clear and motivating, but it currently reads more definitive than the evidence warrants regarding downstream implications (mass amplification; cosmological tension).

We have softened the language here.

- Improve clarity by surfacing the most consequential assumptions (fixed low-mass cutoff; fixed high-mass slope; feature gaps/masks; aperture mismatch) earlier in the main text, not primarily in Methods/limitations.

We have added discussion about these assumptions throughout the text, please see responses to other comments.

- Separate “measurement” statements from “extrapolation” statements more explicitly.

We have ensured that we clarify between “measurement” and “extrapolation” by implementing the conditional language discussed above.

Finally, we note that several concerns raised by the referees will be addressed by our group in future work. In particular, the detailed dynamical modeling of the IMFERNO galaxies is currently *underway and is showing promising results that confirm our findings here (Slob et al. in prep.). Additionally, our group has a recently approved Cycle 5 program to obtain redder spectra of the current IMFERNO sample and tackle this problem. We have added brief references to these upcoming works in the text.*

Referee response

We thank the referees for taking the time to carefully read our resubmission and for their additional feedback. Our responses to the comments are written below:

Reviewer #2 (Comments for the Author):

I thank the authors for their thorough treatment of my comments and recommendations, and can now recommend the paper for publication.

Reviewer #3 (Comments for the Author):

I thank the authors for their detailed response and for the substantial revisions made to the manuscript. The paper is significantly improved. In particular, the authors have addressed many of the main concerns raised in the first round: the IMF parameterization is now stated more explicitly, the lower- and upper-mass limits are clarified, several robustness tests have been added, the aperture-specific nature of the measurements is discussed, and the high-redshift interpretation is phrased more conditionally.

I believe the manuscript is now close to being suitable for publication in Nature Astronomy. However, I still recommend that the authors make a few additional clarifications, mostly concerning wording, scope, and literature framing.

Main remaining points

1. The manuscript should consistently refer to constraints on the low-mass IMF

The authors now state the IMF parameterization clearly in the Methods: the low-mass cutoff is fixed at 0.08\,M_\odot, the two low-mass slopes are varied, and the high-mass slope from 1\,M_\odot to 100\,M_\odot is fixed to the Salpeter value. They also explicitly note that the method is sensitive to the relative fraction of low- to high-mass stars, not to the full IMF shape at formation.

This is a very useful clarification. However, I still think the language throughout the manuscript should be made more consistently precise. In several places, the text refers broadly to “the IMF” being measured or constrained. Given the modelling setup, it would be more accurate to refer to the low-mass IMF, the low-mass end of the IMF, or the dwarf-to-giant sensitive IMF normalization, depending on context.

This distinction is important because the high-mass IMF slope at formation is fixed in the models and remains unconstrained by the spectra. This is not a minor semantic issue: the implications for stellar masses, remnant fractions, early star formation, chemical enrichment, and galaxy-formation tensions depend on the behaviour of the IMF over the full stellar mass range, not only the low-mass end.

I therefore recommend that the authors check the manuscript carefully and ensure that claims about “IMF measurements” are qualified where appropriate as constraints on the low-mass IMF.

We thank the referee and agree that this is an important distinction. Where appropriate, we have qualified that we are constraining the “low-mass” or “low-mass end of the IMF” throughout the text.

2. The broader variable-IMF literature context should be treated more carefully

The revised manuscript now includes an important caveat that a bottom-heavy low-mass IMF does not exclude a shallower high-mass slope, and that such a possibility could reduce high-redshift stellar masses. This is a useful addition.

However, I find the current framing still somewhat too narrow. The idea that the IMF may vary with environment, star-formation conditions, metallicity, density, or cosmic epoch - and that different stellar-mass regimes may behave differently - has a long history in the literature. The current text risks over-emphasising one recent phenomenological framing while under-representing the broader body of work on variable IMFs.

I am not asking for an exhaustive review, and I do not think the paper should become a literature survey. However, if the manuscript makes broad statements about reconciling bottom-heavy low-mass IMFs with high-redshift galaxy-formation tensions, the authors should acknowledge that this belongs to a wider and older discussion of environmentally dependent IMF behaviour. A more neutral wording would be preferable.

For example, along the lines of:

More generally, a bottom-heavy low-mass IMF does not exclude the possibility that the high-mass IMF at formation differed from the Milky Way form. Variable IMF scenarios in which different stellar-mass regimes respond differently to the physical conditions of star formation have been explored in a broad literature, and such behaviour could affect inferred stellar masses and their implications for galaxy-formation models.

This would avoid implying that one recent named IMF prescription is the conceptual origin of this possibility. I would also suggest avoiding unnecessary emphasis on a specifically named “concordance IMF” formulation, since the relevant physical point is simply that different stellar-mass regimes of the IMF may vary differently with star-formation conditions, an idea that is not unique to one recent parameterization and to my opinion there is no need to introduce new names for every slope value of the IMF.

We understand the referee’s point here, and thank them for clarifying. We agree that it is important to discuss the idea that the IMF at different masses may vary in different ways. We *have thus modified the paragraph to name a few different examples (i.e. simultaneously top- and bottom-heavy, time-varying, and metallicity-dependent IMFs).*

3. High-redshift implications are improved, but should remain explicitly conditional throughout

The authors have substantially improved the high-redshift interpretation. The revised text now states:

“If similarly bottom-heavy IMFs were already in place when these galaxies formed…”

and gives an amplification factor of $\sim 4 \pm 1$, rather than presenting the increase as an unconditional result.

This is a major improvement. Still, the abstract and summary remain quite strong. For example, the abstract states that the oldest galaxy is likely a descendant of JWST’s “impossibly early” galaxies and that the latter may have had similarly bottom-heavy IMFs increasing their masses by $\sim 4 \pm 1$.

This is acceptable if phrased carefully, but the authors should ensure that the distinction between direct measurement and extrapolation is clear everywhere. The direct measurement is at $z \sim 0.7$, in aperture-specific spectra. The implications for $z > 4$–6 systems remain conditional on assumptions about formation histories, stellar mixing, IMF gradients, and the persistence of the low-mass IMF signature.

I recommend a final pass through the abstract, main text, and summary to ensure that all statements about high-redshift galaxies remain conditional.

We appreciate the referee’s comment. We have modified the abstract to say that this galaxy “may be” a descendant of the impossibly early galaxies. Additionally, we have made the phrasing addition about apertures suggested in the referee’s comment below. We have checked the rest of the text to make sure that all statements about high-z galaxies are conditional.

4. Aperture specificity and IMF gradients are now discussed well, but should remain visible

The new aperture discussion is very helpful. The authors now state that the spectra probe approximately the inner half-light radius, that local early-type galaxies can have bottom-heavy cores and more MW-like outskirts, and that if the IMF becomes fully MW-like beyond $1R_e$, the correction factors could be closer to $\alpha_{\rm IMF}/2$. They also explicitly state that the IMF constraints should be considered aperture-specific.

This is a strong addition and addresses an important concern. I would only suggest keeping this point visible in the conclusions as well, because it is central to the interpretation of the stellar mass corrections. A short phrase such as “within the apertures probed by the spectra” would help avoid over-generalising the result to the entire galaxy.

We thank the referee and appreciate the concern here. We have added this phrase to the summary paragraph of the paper.

5. Virial mass comparison should remain a plausibility check, not validation

The revised manuscript now frames the virial-mass comparison more carefully, as a plausibility diagnostic rather than a definitive validation. This is much improved.

However, the phrase “reinforcing the physical feasibility” is still fairly strong. I would suggest maintaining cautious language throughout this section, because the virial masses remain approximate and depend on velocity dispersion corrections, rotation, anisotropy, aperture effects, slit losses, and dark matter assumptions. The comparison is useful and reassuring, but it should not be read as an independent confirmation of the inferred IMF normalization.

A wording such as “is consistent with the physical plausibility of” would be slightly safer than “reinforcing”.

We thank the referee for suggesting this rewording, and we have implemented it as stated.

Minor comments

The statement “first robust constraints on the IMF beyond the local Universe from full-spectrum models” is much improved compared to the earlier version. I suggest retaining the qualifier “from full-spectrum models” wherever this novelty claim is made.

We have ensured that we have retained this qualifier throughout the text.

The Data and Code Availability sections are now much stronger. The plan to release the spectra as fitted with alf, resolution curves, wavelength windows, configuration files, the alf commit, and the photometric rescaling procedure is appropriate and important.

The authors now state that all quoted uncertainties are statistical. This should remain clear in figure captions and wherever extreme \alpha_{\rm IMF} values or mass amplification factors are discussed.

We have added that the error bars represent statistical uncertainties in all figure captions, and have also clarified this wherever we discuss specific alpha_IMF values/mass amplification factors.

The added corner plots and abundance-degeneracy tests described in the response are valuable. The manuscript should ensure that the main text sufficiently points readers to these tests, since IMF–abundance degeneracy is one of the central concerns for this type of analysis.

We have added a reference to the corner plots in the main text. We have ensured that we point towards the abundance degeneracy tests in the main text as well (page 3, line 113).

The discussion of systematic uncertainties in alf, stellar libraries, response functions, continuum normalisation, and LSF modelling is now substantially improved. I would keep the tone cautious, especially because the reported uncertainties themselves do not include all these systematics.

We have ensured the tone is cautious in this discussion.