Peer Review File

Manuscript Title: High atmospheric metal enrichment for a Saturn-mass planet

Reviewer Comments & Author Rebuttals

Reviewer Reports on the Initial Version:

Referees' comments:

Referee #1 (Remarks to the Author):

Dr. Sage,

I have finished my review of "High atmospheric metal enrichment for a metal-rich giant planet" by Bean et al. In this paper the authors obtain a thermal emission spectrum from 3-5 microns with JWST for a planet long-known to be anomalous in radius in the hot Jupiter population. It stuck out when it was found 2005, and it still sticks out now. It is a very interesting planet. Many previous studies have attempted to model the formation and structural evolution of the planet. It seems to be an end-member in the maximum bulk metallicity for a Saturn-to-Jupiter class giant planets. The paper is important and timely, and it may merit publication in Nature. It is an interesting counter-point to the recent Line al. (2021) Nature paper on the extremely low atmospheric metallicity of WASP-77Ab.

My main point is that I have some concerns regarding the robustness of the derived atmospheric properties. It is plausible that the error bar on the atmospheric metallicity could be larger than the authors have found so far. If this is the case, the statements made may not quite be as robust as the authors suggest, which may make Nature Astronomy, or another journal, a more appropriate publication.

The main issue is that the retrieval pressure-temperature (P-T) profile, which comes from reference (22) [although I note that the first reference for this parameterization is Guillot (2010), so is reference is made to (22) because it was the first use in a retrieval context? Please clarify] may be strongly impacting the retrieved abundances. I agree that the strong CO2 detection certainly implies a metal-rich atmosphere – that is obvious. But how metal-rich? It is well known that there is a degeneracy between atmospheric abundances and the P-T profile temperature gradient. The more isothermal a profile, the greater an abundance of an atom/molecule is needed to achieve a given absorption feature.

Figure 3 and Extended Data Figure 9, clearly show that there is significant "power" in the emission spectrum, from H2O and CO2, from pressures between 10^-3 to 10^-6 bar. However, in the chosen P-T profile parameterization, the entire region from 10^-4 to 10^-6 bar is isothermal. This is especially worrying for the CO2 abundance, as the right panel of Figure 3 shows emission spectrum contribution up to 10^-7 bar from 4.2-4.6 um. It is my concern that a more realistic P-T profile would yield smaller atmospheric abundances for H2O, CO2, perhaps CO, which would change the atmospheric metallicity and C/O ratio. The inadequacy of the chosen Guillot (2010) P-T profile, which comes isothermal fairly deep in the atmosphere, is discussed analytically in Parmentier et al. (2014, 2015), and it is plausible that it could be more problematic for a higher metallicity atmosphere. I suggest that the retrieval be re-run with a P-T profile parameterization that does not force this isothermal behavior, perhaps using the parameterization of Madhusudhan & Seager (2009), or another choice. Perhaps the results will be similar to what was already found. However, only a test will be able to show.

Some more minor comments:

On Figure 4 – the key figure of the paper -- the authors should also put on WASP-39b, with whatever error bar is deemed appropriate from a combination of the 5 JWST ERS papers. WASP-77Ab would also be appropriate.

It may be worth noting that another reason that the bulk metallicity could be even higher than the suggested 0.66 metal mass fraction is that I believe that the Thorngren et al. modeling methods assume a solar metallicity atmosphere grid for the planetary cooling. A ~100x solar atmosphere would allow the interior to cool off much more slowly, which would necessitate even more metals in the planetary interior to match the small radius.

Quoting the implied dayside planetary Teff in the paper, compared to an expected zero albedo dayside Teq for 2pi or 4pi reradiation, would be interesting as we build up a statistical sample of secondary eclipses with JWST. Such ratios are an important complement to phase curves in assessing energy redistribution.

Can the authors comment on the physical meaning of the 0.98 dilution factor for the planet. Is this...expected? Is it suggestive of nearly homogeneous dayside?

Referee #2 (Remarks to the Author):

I have read the manuscript entitled "High atmospheric metal enrichment for a metal rich giant planet" by Bean et al. I think the result is interesting--CO_2 absorption in a thermal emission spectrum--however there should be more discussion on the big picture implications as I am struggling to see why this result is significant given the current text. I also have numerous questions/suggestions on the modeling interpretation/analysis.

The primary result here, is a metallicity constraint in the atmosphere based upon the depth of the 4.2 um CO_2 absorption feature (and water). The high atmospheric metallicity is consistent with the high inferred bulk planetary metallicity given the planetary mass and radius. This result suggests that atmospheric metallicity is correlated with bulk planetary metallicity.

Perhaps I am missing a subtle discussion in the text, but is the correlation between bulk metallicity and atmospheric metallicity all that surprising? Is this not what is inferred for Neptune/sub-Neptune type planets, for example, GJ 436b (e.g., Morley et al. 2017)? I suppose what makes this target particularly interesting is that it is a "jovian" category planet with such a high metallicity; though this was already known from the discovery paper (e.g, Sato 2016)—the JWST observations effectively confirm that it is indeed elevated in metallicity.

My biggest concern is in drawing conclusions regarding atmospheric vs. bulk metallicities based upon the metallicity inferences of a single planet. It is not immediately obvious that this isn't simply a result of astrophysical scatter—there is no context in the manuscript beyond the solar system planets to assess the significance of this planet as a potential outlier. For instance, are there already correlations between bulk metallicity (e.g., those presented in say, Thorngren et al. 2016) and constraints from HST/Hi-res measurements? The manuscript seems to brush off previous atmospheric metallicity constraints, but there have been several uniform analyses looking at atmospheric metallicity (Changeat et al. 2022 being the most recent for emission). Some reported constraints in the literature combining STIS+WFC3 have resulted in metallicity constraints (e.g., Welbanks et al. 2019) comparable to those reported here. There are also several reliable high resolution ground based observations of giant planets that have led to very precise atmospheric metallicity constraints (e.g., WASP-18b- Brogi et al. 2022, WASP-77Ab-Line et al. 2021, KELT-20b-Kasper et al. 2022, KELT-9b-Kasper et al. 2021). The authors should consider at least looking into these results, perhaps performing the same bulk vs. atmospheric metallicity exercise adding to Figure 4, to provide further context for their own result showing that HD149.. is an outlier, as right now, there is very little.

While certainly broad wavelength coverage and high signal to noise has the potential to offer more stringent constraints on atmospheric abundances, such analyses are not immune to the modeling assumptions used to derive them. As such, I have a series of questions/comments/tests to explore the reliability of the inferred metallicity.

1. My primary concern is a potential bias in the CO2 abundance (which drives the metallicity) due to the isothermal nature of the temperature-pressure profile at pressures lower than ~1 mbar. The manuscript indicates (Lines 175-176, and Figure 3) that the core of the CO2 band probes pressures as low as 1E-7 bar. This suggests that nearly 3 decades in pressure resides within the isothermal region of the atmosphere, and appears (Figure 3, right) to encompass nearly half of the CO2 contribution function. Realistic simulations of atmospheres (either through GCM's or 1D-radiative convective models) are not isothermal to these low pressures (e.g, Fortney et al. 2008, Molliere et al. 2016) for planets in these temperature regimes. My concern is that the CO2 abundance can keep increasing (and hence metallicity) without consequence as the band core is "saturated". I would recommend the following experiments to alleviate this concern:

a. Take the best fit parameters (those chosen for Figure 2) and artificially adjust the slope of the TP profile between ~1E-4 bar and 1E-7. Perhaps by extending the temperature gradient produced in the Guillot 2010 profile between 1E-3 - 1E-4 bar. This would illustrate the sensitivity of the CO2 band depth to the temperature gradient at these shallower pressures.

b. Implement a more flexible TP profile parameterization that does not, by construction, go isothermal at low pressures. Several choices exist and include the Madhusudhan & Seager 2009 parameterization and/or simple slabs of linear temperature gradients (e.g., as in Line et al. 2016). The point here is to show that the answer is robust regardless of the TP parameterization and this isothermal artifact.

c. Inspecting the corner plot in ED Figure 7, it appears that a number of parameters "run up" against their prior boundaries. For instance, C/O and logKth (also "dil", but an upper prior bound of 1 is justified). I am concerned that the [M/H] an C/O constraints may be influenced by some of these prior bounds. It's not immediately obvious why these upper prior bounds are chosen. While there is not an "obvious" tilt to the logKir vs. [M/H] 2-D histogram, it would be prudent to extend the prior range on both logKir and C/O to test that no correlation persists at higher values and that the abundance results are independent of the prior bound choices.

2. It would be beneficial to plot a 1D-radiative convective model TP on top of Figure 3 to illustrate the physical plausibility (or not) of the retrieved TP. If a new TP profile parameterization is used (as suggested above), it would also be informative to plot that on top as well.

3. The manuscript discusses a photochemical model was used to explore the potential existence of SO2. I encourage the authors to show the impact (or lack there of) that vertical mixing and photochemistry play in influencing (or not) the CO2 abundance profile. For instance, the retrieved equilibrium CO2 abundance profile in ED Figure 8 appears to vary by about an order of magnitude. Depending on the vertical mixing strength (and the presence of a homopause), it is not unreasonable to suspect that quenching could occur at the lower pressures where the temperature is cooler, which could perturb CO2 abundance by up to an order of magnitude away from equilibrium over the pressure levels indicated by the CO2 contribution functions.

4. I would also encourage the authors, as a robustness check, to run a free retrieval with the key gases to see if it results in comparable molecular abundances to those that result from the thermochemical consistent retrieval analysis.

Minor

-on Line 102 – 104 say that several orbital parameters are fixed when fitting the light curve. How do the previously reported uncertainties in those parameters propagate into either the white-light secondary eclipse depth and into the subsequent secondary eclipse spectrum? It is negligible, then it is good to say so.

-It would be useful to quantify the detection significances of each of the gases, following the standard practice described in the series of ERS WASP-39b papers (via Bayes factor analysis). This can be readily accomplished within the described retrieval framework by “removing” each given gas.

Author Rebuttals to Initial Comments:

Overall response:

We thank the referees for their review of the paper and helpful suggestions. We have implemented all of the proposed tests and we find that the result is unchanged. The assumed form of the temperature-pressure profile does not impact the results significantly, nor does vertical mixing and photochemistry.

We consider the result noteworthy because this one measurement of a single planet contradicts the idea of a simple "mass-metallicity" trend like is seen in the solar system at high confidence. The mass-metallicity trend has been a powerful driving force in the field of exoplanet atmospheres since it was introduced generally by Fortney et al. (2013) and as a specific power law relationship for giant planets by Kreidberg et al. (2014). The latter paper has 271 citations and has been the inspiration for major observing programs, space mission proposals, and theoretical work.

However, nearly all the previous work on the mass-metallicity relationship is a confused mess because of the lack of the reliability of the metallicity measurements. As described in the text, previous studies were stymied by insufficient data that often gave contradictory results. Only three previous measurements used spectroscopic detections of both oxygen- and carbon-bearing species to infer the metallicity (WASP-77Ab, WASP-39b, and WASP-18b). None of these measurements disagree with the mass-metallicity trend seen in the solar system. Two of them are also quite recent (WASP-39b, WASP-18b), being submitted for publication in the same timeframe as this paper. And one of them is for a planet that is 11 M_Jup (WASP-18b), which is well beyond the range of planets seen in the solar system. We have added two of these measurements to Figure 3 (was Figure 4) as added context. More details are given in the point-by-point responses below.

In contrast to all this previous work, our measurement for this Saturn-mass planet disagrees with the solar system mass-metallicity trend at >4 sigma confidence. This is the first definitive proof that If there is a mass-metallicity trend for close-in exoplanets then the cosmic variance around the trend must be nearly equal to the full range of values of the trend itself.

In addition, we gently suggest that Referee #2 is undervaluing the connection we have made between bulk and atmospheric metallicity. There are no credible and high-precision atmospheric metallicity measurements for exoplanets with high bulk metallicities. The referenced example of GJ436b does not have a single spectroscopic detection of a molecule in its atmosphere. The metallicity is only inferred from Spitzer photometry, which is known to yield degenerate solutions.

The editor requested that we come up with a better title, shorten the paper, and remove additional introductory material after the first, bold-faced paragraph. Therefore, we have moved Figure 1 to the Extended Data and we have removed most of the second and third paragraphs. At this point the main text of the paper is as long as it can be and there is a full complement of 10 Extended Data figures.

Point-by-point responses:

Referees' comments:

Referee #1 (Remarks to the Author):

Dr. Sage,

I have finished my review of “High atmospheric metal enrichment for a metal-rich giant planet” by Bean et al. In this paper the authors obtain a thermal emission spectrum from 3-5 microns with JWST for a planet long-known to be anomalous in radius in the hot Jupiter population. It stuck out when it was found 2005, and it still sticks out now. It is a very interesting planet. Many previous studies have attempted to model the formation and structural evolution of the planet. It seems to be an end-member in the maximum bulk metallicity for a Saturn-to-Jupiter class giant planets. The paper is important and timely, and it may merit publication in Nature. It is an interesting counter-point to the recent Line al. (2021) Nature paper on the extremely low atmospheric metallicity of WASP-77Ab.

Note that WASP-77Ab has been added to Figure 3 (previously Figure 4).

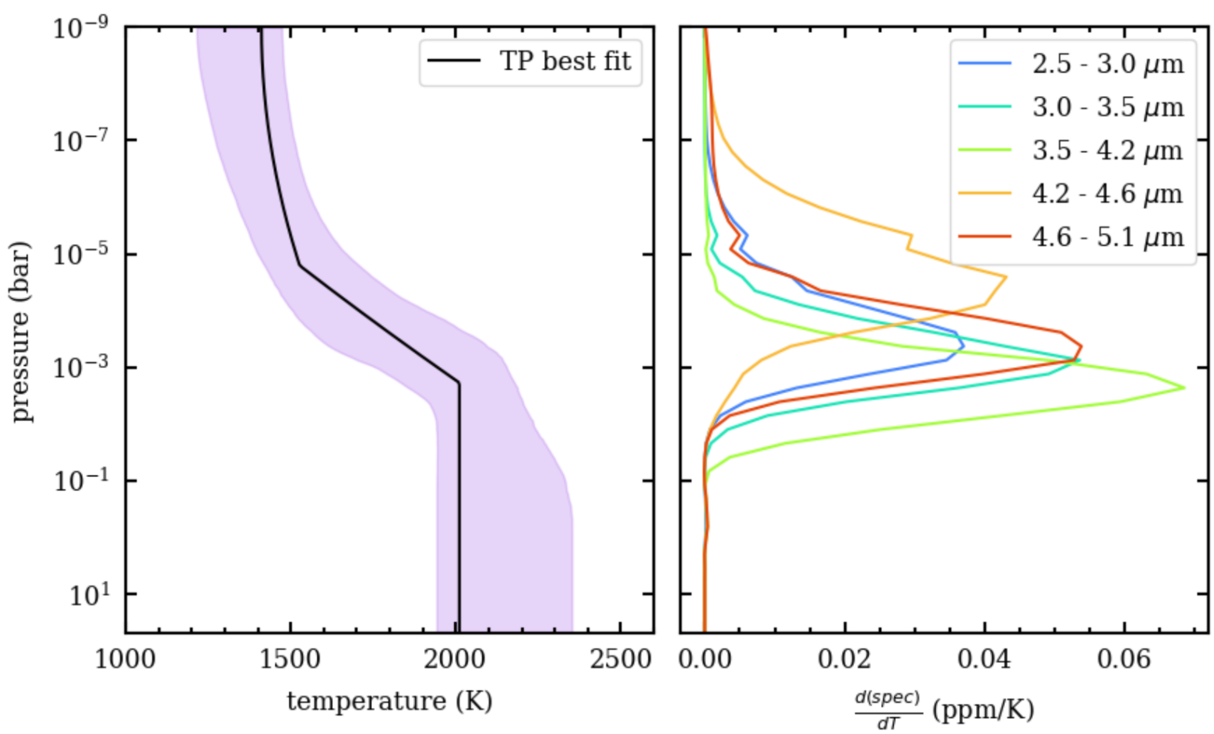
My main point is that I have some concerns regarding the robustness of the derived atmospheric properties. It is plausible that the error bar on the atmospheric metallicity could be larger than the authors have found so far. If this is the case, the statements made may not quite be as robust as the authors suggest, which may make Nature Astronomy, or another journal, a more appropriate publication.

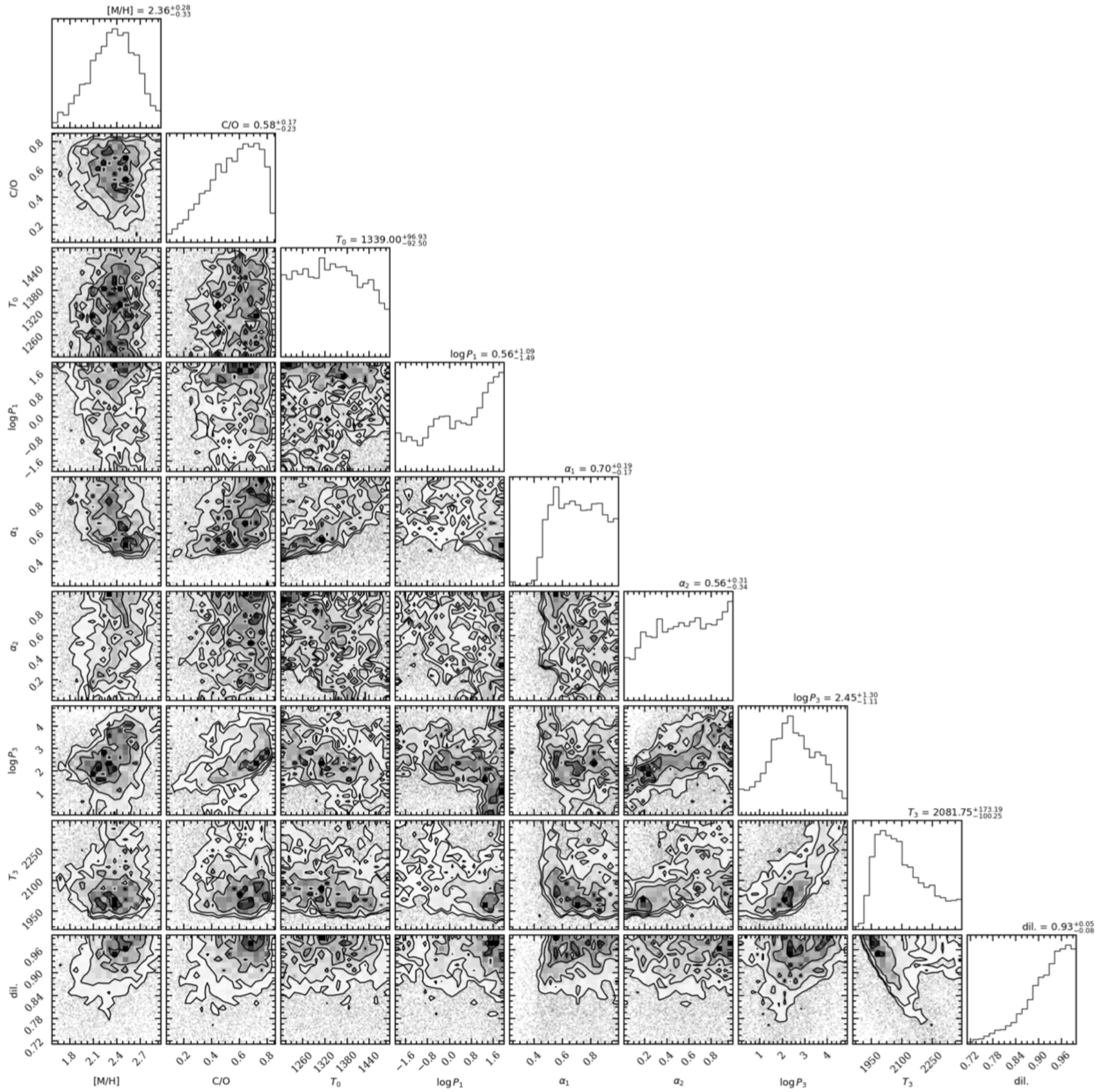
The main issue is that the retrieval pressure-temperature (P-T) profile, which comes from reference (22) [although I note that the first reference for this parameterization is Guillot (2010), so is reference is made to (22) because it was the first use in a retrieval context? Please clarify] may be strongly impacting the retrieved abundances. I agree that the strong CO2 detection certainly implies a metal-rich atmosphere – that is obvious. But how metal-rich? It is well known that there is a degeneracy between atmospheric abundances and the P-T profile temperature gradient. The more isothermal a profile, the greater an abundance of an atom/molecule is needed to achieve a given absorption feature.

Reference to (22) is indeed because it was the first to use it in a retrieval context. See the longer response to the other comments below.

Figure 3 and Extended Data Figure 9, clearly show that there is significant “power” in the emission spectrum, from H2O and CO2, from pressures between 10^-3 to 10^-6 bar. However, in the chosen P-T profile parameterization, the entire region from 10^-4 to 10^-6 bar is isothermal. This is especially worrying for the CO2 abundance, as the right panel of Figure 3 shows emission spectrum contribution up to 10^-7 bar from 4.2-4.6 um. It is my concern that a more realistic P-T profile would yield smaller atmospheric abundances for H2O, CO2, perhaps CO, which would change the atmospheric metallicity and C/O ratio. The inadequacy of the chosen Guillot (2010) P-T profile, which comes isothermal fairly deep in the atmosphere, is discussed analytically in Parmentier et al. (2014, 2015), and it is plausible that it could be more problematic for a higher metallicity atmosphere. I suggest that the retrieval be re-run with a P-T profile parameterization that does not force this isothermal behavior, perhaps using the parameterization of Madhusudhan & Seager (2009), or another choice. Perhaps the results will be similar to what was already found. However, only a test will be able to show.

We have implemented the Madhusudhan temperature-pressure profile parameterization. Some figures summarizing the results are shown below. We find that the results agree with our original retrieval: the metallicity and C/O are both consistent within 1 sigma. The more flexible parameterization gives larger errors on the C/O. The fact that the profile goes nearly isothermal where it does is actually mainly due to the data. Indeed, in preliminary versions of the data we got a mild inversion in the retrieval, so the parameterization is not overly constraining. We elect to keep the original retrieval results because we think that the Madhusudhan profile is too flexible given the quality of the data. The conclusions of the paper would actually be stronger if we used that result instead because the metallicity is even higher.





Some more minor comments:

On Figure 4 – the key figure of the paper -- the authors should also put on WASP-39b, with whatever error bar is deemed appropriate from a combination of the 5 JWST ERS papers. WASP-77Ab would also be appropriate.

We added WASP-77Ab and WASP-39b to the figure (now Figure 3, was Figure 4). The WASP-77Ab result was already published and we should have included it before. The WASP-39b result is still somewhat preliminary at this point, but probably robust enough for this purpose with a conservative error bar. We cite the Feinstein et al. (in press) results.

It may be worth noting that another reason that the bulk metallicity could be even higher than the suggested 0.66 metal mass fraction is that I believe that the Thorngren et al. modeling methods assume a solar metallicity atmosphere grid for the planetary cooling. A ~100x solar atmosphere would allow the interior to cool off much more slowly, which would necessitate even more metals in the planetary interior to match the small radius.

This comment is relevant for cool giant planets, but in this case we have the hot Jupiter heating regulating the interior adiabat. Our heating was calibrated to match the radius, so as long as we're using the same atmosphere model (we are), the radius and entropy will be correct. What does end up being different compared to an atmosphere model with a self-consistent metallicity is the intrinsic temperature and heating power, which would both be somewhat high in the model we're using. Fortunately that's not important for this work. Also, the T_int scales so strongly with internal entropy that it's not likely to be that far off in our estimate.

Quoting the implied dayside planetary Teff in the paper, compared to an expected zero albedo dayside Teq for 2pi or 4pi reradiation, would be interesting as we build up a statistical sample of secondary eclipses with JWST. Such ratios are an important complement to phase curves in assessing energy redistribution.

We have added the bandpass-averaged brightness temperature and the planet's zero-albedo equilibrium temperature to paragraph 6.

Can the authors comment on the physical meaning of the 0.98 dilution factor for the planet. Is this...expected? Is it suggestive of nearly homogeneous dayside?

Given the length restrictions we are glossing over this. It isn't clear if the parameter has a straightforward physical meaning. For now we treat it as a nuisance parameter to get at the composition.

Referee #2 (Remarks to the Author):

I have read the manuscript entitled "High atmospheric metal enrichment for a metal rich giant planet" by Bean et al. I think the result is interesting--CO2 absorption in a thermal emission spectrum--however there should be more discussion on the big picture implications as I am struggling to see why this result is significant given the current text. I also have numerous questions/suggestions on the modeling interpretation/analysis.

The primary result here, is a metallicity constraint in the atmosphere based upon the depth of the 4.2 um CO2 absorption feature (and water). The high atmospheric metallicity is consistent with the high inferred bulk planetary metallicity given the planetary mass and radius. This result suggests that atmospheric metallicity is correlated with bulk planetary metallicity.

Perhaps I am missing a subtle discussion in the text, but is the correlation between bulk metallicity and atmospheric metallicity all that surprising? Is this not what is inferred for Neptune/sub-Neptune type planets, for example, GJ 436b (e.g., Morley et al. 2017)? I suppose what makes this target particularly interesting is that it is a “jovian” category planet with such a high metallicity; though this was already known from the discovery paper (e.g, Sato 2016)—the JWST observations effectively confirm that it is indeed elevated in metallicity.

The bulk metallicity of the planet was known but how atmospheric composition traces the bulk metallicity is totally unknown for exoplanets. It isn’t necessarily surprising that they are correlated, but until this work there was no precise empirical data for exoplanets to test the idea. The JWST data don’t “confirm” the high metallicity. The JWST data presented here measure the atmospheric metallicity for the first time.

We rebut the specific discussion of GJ436b in the general response above.

My biggest concern is in drawing conclusions regarding atmospheric vs. bulk metallicities based upon the metallicity inferences of a single planet. It is not immediately obvious that this isn’t simply a result of astrophysical scatter—there is no context in the manuscript beyond the solar system planets to assess the significance of this planet as a potential outlier.

We don’t feel it is an accurate representation to say there was no context. There was an entire paragraph dedicated to this (beginning “Motivated by the trend that is observed in the solar system…”). And the solar system context is the main comparison sample given the lack of reliable data for exoplanets. This is a JWST spectrum with a detection of water and carbon dioxide. We have both oxygen- and carbon-bearing species, we have a recognized metallicity indicator, and we have the precision and spectral resolution to break the degeneracy with the temperature-pressure profile. The “astrophysical scatter” is the actual point of this paper. This planet is not a “potential outlier”. It is >4 sigma deviant from the solar system trend.

For instance, are there already correlations between bulk metallicity (e.g., those presented in say, Thorngren et al. 2016) and constraints from HST/Hi-res measurements?

There are no such results in the literature with high precision atmospheric metallicity measurements.

The manuscript seems to brush off previous atmospheric metallicity constraints, but there have been several uniform analyses looking at atmospheric metallicity (Changeat et al. 2022 being the most recent for emission). Some reported constraints in the literature combining

STIS+WFC3 have resulted in metallicity constraints (e.g., Welbanks et al. 2019) comparable to those reported here.

We already address this in the paper. While some prior studies may have yielded reliable results, it is not possible to identify which ones those might be given the wide range of contradictory results in the literature for the same planet using the same data. As explained in the paper, the main limitation was the previous data. The data weren't constraining enough, and some data sets were likely just plain wrong. So a wide range of assumptions were possible in retrievals and they gave different results. Only a handful of previous data sets were constraining enough - see response to the next point.

There are also several reliable high resolution ground based observations of giant planets that have led to very precise atmospheric metallicity constraints (e.g., WASP-18b- Brogi et al. 2022, WASP-77Ab-Line et al. 2021, KELT-20b-Kasper et al. 2022, KELT-9b-Kasper et al. 2021). The authors should consider at least looking into these results, perhaps performing the same bulk vs. atmospheric metallicity exercise adding to Figure 4, to provide further context for their own result showing that HD149.. is an outlier, as right now, there is very little.

This is an important suggestion and we have implemented it. We have chosen to include previous results for transiting (i.e., close-in) exoplanets that were based on spectroscopic identification of oxygen- and carbon-bearing species and that applied a retrieval framework to give quantitative constraints on the abundances. That gives WASP-77Ab (Line et al. 2021), WASP-39b (JWST ERS results, we choose to reference the Feinstein et al. results that are in press), and WASP-18b (Brogi et al. in press). The results for KELT-9b and KELT-20b don't include both carbon- and oxygen-bearing species. In the context of giant planet formation these are thought to be much more important than the refractories that those studies concentrated on.

The editor should note the results from WASP-39b and WASP-18b were both submitted for publication in the same timeframe as this paper, and thus they weren't included originally. The emerging results for WASP-39b were already referenced in the text of the paper.

We have added the results for WASP-77Ab and WASP-39b to Figure 3 (was Figure 4). These planets all agree fairly well with the mass-metallicity trend. We did not add the result for WASP-18b because the planet has a mass of 11 M_Jup and thus is far outside the range of planets seen in the solar system. The target of this study is a Saturn-mass planet, and thus is directly comparable to the solar system.

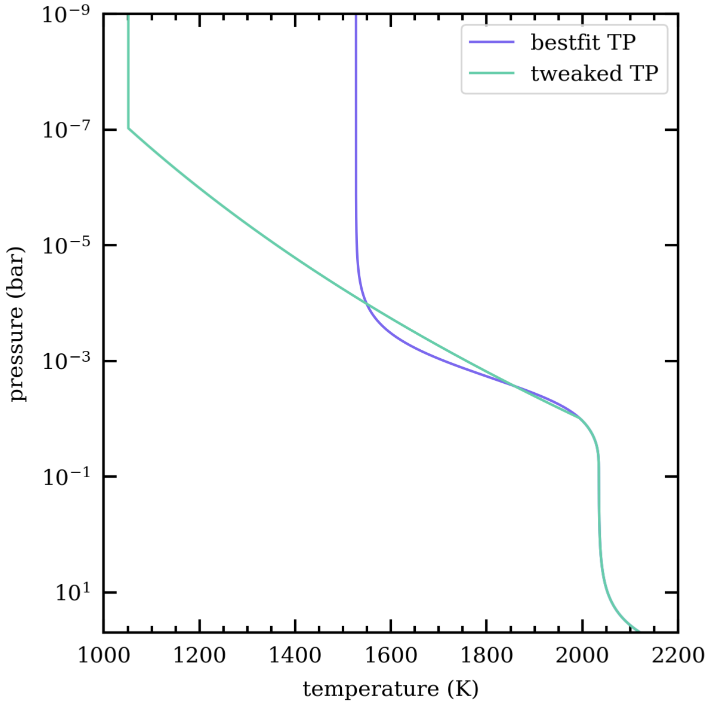
As can be seen in the new figure, HD 149026b is indeed a large outlier from the mass-metallicity trend and no other comparable exoplanet measurement is similarly discrepant.

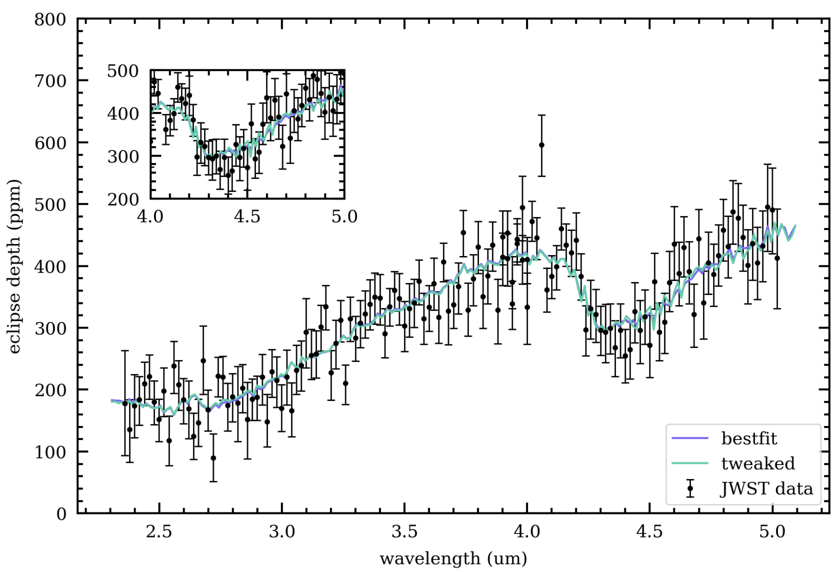
While certainly broad wavelength coverage and high signal to noise has the potential to offer more stringent constraints on atmospheric abundances, such analyses are not immune to the modeling assumptions used to derive them. As such, I have a series of questions/comments/tests to explore the reliability of the inferred metallicity.

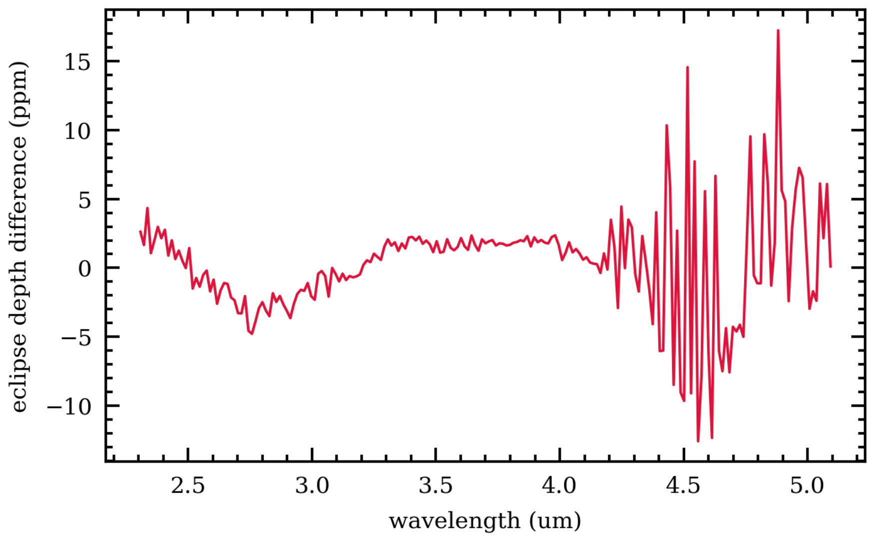
1. My primary concern is a potential bias in the CO2 abundance (which drives the metallicity) due to the isothermal nature of the temperature-pressure profile at pressures lower than ~1 mbar. The manuscript indicates (Lines 175-176, and Figure 3) that the core of the CO2 band probes pressures as low as 1E-7 bar. This suggests that nearly 3 decades in pressure resides within the isothermal region of the atmosphere, and appears (Figure 3, right) to encompass nearly half of the CO2 contribution function. Realistic simulations of atmospheres (either through GCM's or 1D-radiative convective models) are not isothermal to these low pressures (e.g, Fortney et al. 2008, Molliere et al. 2016) for planets in these temperature regimes. My concern is that the CO2 abundance can keep increasing (and hence metallicity) without consequence as the band core is "saturated". I would recommend the following experiments to alleviate this concern:

a. Take the best fit parameters (those chosen for Figure 2) and artificially adjust the slope of the TP profile between ~1E-4 bar and 1E-7. Perhaps by extending the temperature gradient produced in the Guillot 2010 profile between 1E-3 - 1E-4 bar. This would illustrate the sensitivity of the CO2 band depth to the temperature gradient at these shallower pressures.

We have performed this test. Some figures summarizing the results are given below. Basically, the spectrum is only weakly sensitive to the TP profile at low pressures. This is essentially what is already shown by the contribution functions. The data are sensitive to low pressures and thus constrain the TP profile in this region. But the bulk of the CO_2 signal comes from deeper in the atmosphere.


bestfit TP
tweaked TP
pressure (bar)
temperature (K)
10^{-9}
10^{-7}
10^{-5}
10^{-3}
10^{-1}
10^{1}
1000
1200
1400
1600
1800
2000
2200


eclipse depth (ppm)
wavelength (um)
bestfit
tweaked
JWST data
800
700
600
500
400
300
200
100
0
2.5
3.0
3.5
4.0
4.5
5.0



b. Implement a more flexible TP profile parameterization that does not, by construction, go isothermal at low pressures. Several choices exist and include the Madhusudhan & Seager 2009 parameterization and/or simple slabs of linear temperature gradients (e.g., as in Line et al. 2016). The point here is to show that the answer is robust regardless of the TP parameterization and this isothermal artifact.

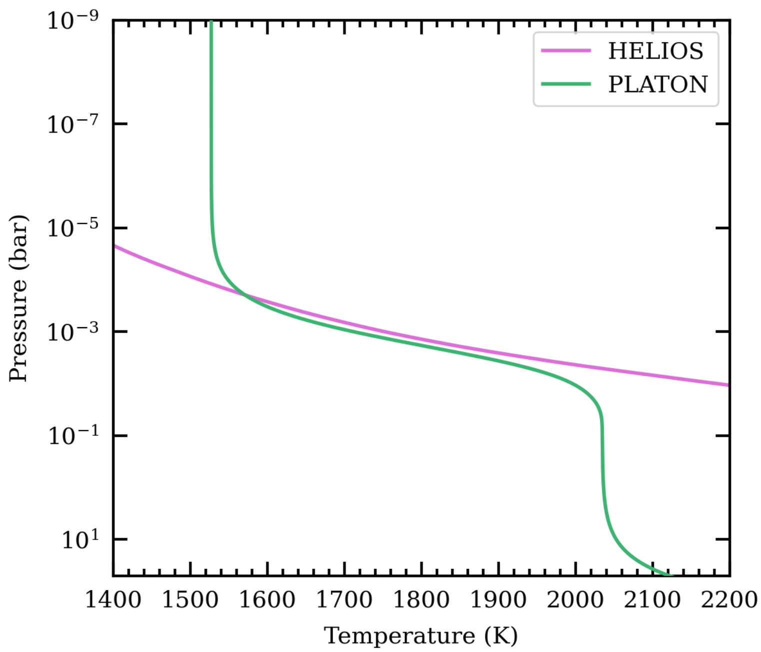
Done. See the response to Referee #1 above.

c. Inspecting the corner plot in ED Figure 7, it appears that a number of parameters "run up" against their prior boundaries. For instance, C/O and logKth (also "dil", but an upper prior bound of 1 is justified). I am concerned that the [M/H] an C/O constraints may be influenced by some of these prior bounds. It's not immediately obvious why these upper prior bounds are chosen. While there is not an "obvious" tilt to the logKir vs. [M/H] 2-D histogram, it would be prudent to extend the prior range on both logKir and C/O to test that no correlation persists at higher values and that the abundance results are independent of the prior bound choices.

The important kappa parameter does not actually run up against the prior (lower limit of zero). It just looks like it does given the range in the corner plot. This has been fixed. As described in the paper, there are no correlations (and thus degeneracies) between the composition and the temperature-profile parameters.

2. It would be beneficial to plot a 1D-radiative convective model TP on top of Figure 3 to illustrate the physical plausibility (or not) of the retrieved TP. If a new TP profile parameterization is used (as suggested above), it would also be informative to plot that on top as well.

We calculated a 1D-radiative convective model using HELIOS with the best-fit parameters from the chemical equilibrium retrieval (see below). It agrees well in the range of pressures that we are most sensitive to, although it disagrees at low and high pressures. We consider this sufficient support for the retrieval results. Further exploration of this is beyond the scope of the paper given the space limitations.

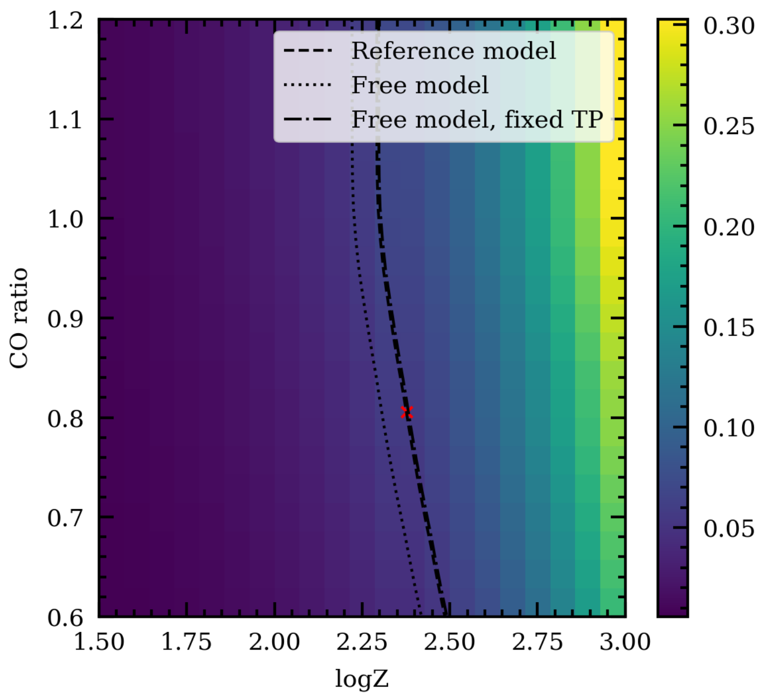


3. The manuscript discusses a photochemical model was used to explore the potential existence of SO2. I encourage the authors to show the impact (or lack there of) that vertical mixing and photochemistry play in influencing (or not) the CO2 abundance profile. For instance, the retrieved equilibrium CO2 abundance profile in ED Figure 8 appears to vary by about an order of magnitude. Depending on the vertical mixing strength (and the presence of a homopause), it is not unreasonable to suspect that quenching could occur at the lower pressures where the temperature is cooler, which could perturb CO2 abundance by up to an order of magnitude away from equilibrium over the pressure levels indicated by the CO2 contribution functions.

The non-equilibrium chemistry calculations indicate that the CO_2 abundance is not strongly impacted by vertical mixing. We implemented a quench pressure for CO_2 in the retrieval and it didn't change the results.

4. I would also encourage the authors, as a robustness check, to run a free retrieval with the key gases to see if it results in comparable molecular abundances to those that result from the thermochemical consistent retrieval analysis.

We ran a free retrieval and the results suggested lower CO_2 and H_2O abundances than the chemical equilibrium retrieval. However, we don't consider this result informative because it would require abundances very far from equilibrium. We find that in the free retrieval the abundances of the individual gasses are highly correlated with the kappa parameter. The spectrum has essentially no continuum windows outside the CO_2 and H_2O features. Therefore, a free retrieval can trade the temperature-pressure profile against the abundances. In this case the retrieval reduces the gas abundances and shifts the temperature-pressure profile to the left. Thus it produces a similar model fit with lower abundances. However, the retrieved ratio of the CO_2 to H_2O abundances is very similar as for the chemical equilibrium retrieval (~0.05). It is not possible to have that ratio simultaneous with the low CO_2 and H_2O abundances implied by the free retrieval for any reasonable combination of M/H and C/O values. Therefore, we consider the free retrieval degenerate with the current data set. The chemical equilibrium retrieval breaks the degeneracy between the composition and temperature-pressure profile parameters because it essentially uses the ratio of CO_2 and H_2O abundances required to fit the data as an additional constraint. The figure below provides some additional information ("CO ratio" is C/O, red X is the best-fit chemical equilibrium retrieval point, color represents the CO_2/H_2O abundance ratio).



Minor

-on Line 102 – 104 say that several orbital parameters are fixed when fitting the light curve. How do the previously reported uncertainties in those parameters propagate into either the white-light secondary eclipse depth and into the subsequent secondary eclipse spectrum? It is negligible, then it is good to say so.

The uncertainties in the parameters have no impact on the conclusions of the paper. This is a standard assumption in the field. A statement to this effect has been added to the paper.

-It would be useful to quantify the detection significances of each of the gases, following the standard practice described in the series of ERS WASP-39b papers (via Bayes factor analysis). This can be readily accomplished within the described retrieval framework by "removing" each given gas.

The spectrum deviates from the blackbody at 9.2σ confidence. By our calculations CO_2 is unambiguously identified at 3.2σ confidence. To arrive at this number we removed the CO_2 opacity, re-ran the retrieval, and calculated the significance based on the chi^2 value for the fit at the position of the CO_2 feature. H_2O is not uniquely identified at high confidence because it behaves almost like a continuum opacity source over this bandpass and the temperature-pressure profile can adjust to nearly completely compensate if it were absent.

This sounds like a modest detection of CO_2, but it is the significance of CO_2 specifically rather than just a generic spectral feature. And although the evidence for H_2O specifically is weak, not having it or CO_2 in the model results in strange abundances. We get a good fit (reduced chi^2 ~ 1.0) assuming a scaled solar abundance pattern and chemical equilibrium, and such a model indicates CO_2 and H_2O as the dominant opacity sources.

The text at the end of the 6th paragraph has been revised to include this information.

Reviewer Reports on the First Revision:

Referees' comments:

Referee #1 (Remarks to the Author):

Dr. Sage,

I have finished my second review of "High atmospheric metal enrichment for a Saturn-mass planet" by Bean et al. I apologize for the long delay, as I mentioned the original e-mail went to my spam folder, and I was only able to make time to read the revised paper over the weekend.

I am personally a bit surprised, but pleased, by the robustness of the derived atmospheric metallicity to the change in P-T profile. I find this work important and compelling. I do think that this planet is an important counter-point to the atmosphere mass-metallicity relation which has been a significant theoretical and observational driver for the field over the past decade. While it is true that the planet, in bulk, was known to be metal-rich, I think it is important for the field, and for re-orientating people's thinking, that the atmosphere is also extremely metal-rich. So I believe that this work merits publication in Nature.

A small additional comment: An aspect of the derived atmospheric C/O ratio that is well-appreciate by some, but ignored by others, is that the condensation of silicate clouds like $MgSiO_3$ / Mg_2SiO_4 is expected to remove something like 20-25% of the oxygen away from what would be available to be found in gaseous H_2O, CO, CO_2, etc. One might therefore expect a change in population level C/O ratio for planets hotter and colder than the Teff/log g range for the onset of MgSi-bearing clouds. On the other hand, the cold night sides of hot Jupiters (where these clouds may be stable for essentially all hot Jupiters) significantly complicates this situation and might mean that all hot Jupiters have already "lost" this oxygen from the gas phase. I realize that space is at a premium, but it would be great if the authors could mention that their derived profile is warmer than the MgSi-bearing silicates condensation curve (at least to my eye, i.e. Visscher et al. 2010, eq. 18 and 20).

Referee #2 (Remarks to the Author):

I thank the authors for their thorough responses to my original round of questions. Based upon these responses and the revised manuscript, I have (mostly minor, except perhaps the first point) additional comments/questions. Given the content, I feel that this article is more appropriate for Nature Astronomy.

The main thesis of the manuscript is that atmospheric metallicity is more correlated with bulk metallicity than with mass. I missed this when reviewing the initial manuscript, but there does not appear to be any quantification of "more correlated with". Specially, while Figure 3 provides the nominal fit parameters to the 7 planets/data points, no correlation coefficient or p-values are given. These would be useful metrics to provide so that a reader, especially a non-expert, would be able to determine exactly how correlated these quantities are (perhaps solar system planets by themselves, then including the exoplanet points).

I appreciate the deeper look into the T-P profile vs. CO_2 abundance degeneracy. While I am largely convinced that the choice of T-P profile parameterization (G10 vs. MS09) is inconsequential, I do have some minor clarification questions (I am assuming it is just plotting inconsistency). I am confused by what is shown in the corner plot given in the response under the MadhuSeager 2009 profile compared to the reconstructed T-P profile shown above the corner plot. The parameter, logP3, which I take to control the location at which the deep isothermal layer begins (top of layer), seems to cover ranges from 0 -4, which is 1 – 10,000 bars. However, the reconstructed TP profile suggests a more plausible range between 1E-3 and ~ 1E-1 bars. These ranges seem rather inconsistent. What is going on here? Is what is being reconstructed reflecting what is shown in the corner plot?

I also appreciate that the authors ran a free retrieval on the data. It is unfortunate that the free retrieval is unable to provide consistent absolute abundance constraints with those derived from the chemical equilibrium retrieval. I think it would be informative, and transparent, to provide the (C+O)/H based free retrieval metallicity (as was done for the WASP-77 data point, Line et al. 2021) for comparison. Why am I suggesting this? The manuscript argues for showing only exoplanets with C+O based metallicity constraints (as opposed to various HST constraints what are solely based on H2O). However, as the free retrieval is unable to detect evidence for water via the Bayes factor (as indicated in the response), it's unclear what exactly is constraining the metallicity in the atmosphere of HD149. Are the authors able to extract a C/H (and an O/H) based metallicity to compare directly to the solar system?

Author Rebuttals to First Revision:

Overall response:

We thank the referees and editor for their further helpful comments on the paper. We have addressed the detailed comments below and we have tried our best to get the paper in the correct format. There are no major changes, just two small additions (one to the main text, one to the Methods) as requested by the referees.

The instructions requested that we comment on our willingness to make the referee comments and responses public: We wish to participate in transparent peer review.

Point-by-point responses:

A small additional comment: An aspect of the derived atmospheric C/O ratio that is well-appreciate by some, but ignored by others, is that the condensation of silicate clouds like $MgSiO_3$ / Mg_2SiO_4 is expected to remove something like 20-25% of the oxygen away from what would be available to be found in gaseous H_2O, CO, CO_2, etc. One might therefore expect a change in population level C/O ratio for planets hotter and colder than the Teff/log g range for the onset of MgSi-bearing clouds. On the other hand, the cold night sides of hot Jupiters (where these clouds may be stable for essentially all hot Jupiters) significantly complicates this situation and might mean that all hot Jupiters have already “lost” this oxygen from the gas phase. I realize that space is at a premium, but it would be great if the authors could mention that their derived profile is warmer than the MgSi-bearing silicates condensation curve (at least to my eye, i.e. Visscher et al. 2010, eq. 18 and 20).

The uncertainty envelope for the retrieved T-P profile does cross the condensation curves for MgSi-bearing clouds. This is unlikely to impact the interpretation of the metallicity given the large errors. The impact on the C/O could be larger but it is likely to be difficult to discern at the individual planet level, so we leave the study of this for future work. We have added a short discussion to the end of the Methods.

The main thesis of the manuscript is that atmospheric metallicity is more correlated with bulk metallicity than with mass. I missed this when reviewing the initial manuscript, but there does not appear to be any quantification of “more correlated with". Specially, while Figure 3 provides the nominal fit parameters to the 7 planets/data points, no correlation coefficient or p-values are given. These would be useful metrics to provide so that a reader, especially a non-expert, would be able to determine exactly how correlated these quantities are (perhaps solar system planets by themselves, then including the exoplanet points).

The fits in Figure 3 are actually just to the solar system planets, as is described in the caption. We added a word to make this more clear. We also added the p value of the trend for the bulk metallicity panel to the caption (0.003). We don't give the fit parameters and p value for the trend with planet mass because this has been done many times in the literature. We also don't fit the trends including the exoplanets because the sample is still small. The focus of the paper is how the result for HD149026b compares with the well-defined solar system trends. The text already quantifies how well the new result for this planet compares to the trend with mass (>4 sigma discrepant, see first paragraph) and the new trend with bulk metallicity (agreement at 2.1 sigma, see paragraph 11).

I appreciate the deeper look into the T-P profile vs. CO_2 abundance degeneracy. While I am largely convinced that the choice of T-P profile parameterization (G10 vs. MS09) is inconsequential, I do have some minor clarification questions (I am assuming it is just plotting inconsistency). I am confused by what is shown in the corner plot given in the response under the MadhuSeager 2009 profile compared to the reconstructed T-P profile shown above the corner plot. The parameter, logP3, which I take to control the location at which the deep isothermal layer begins (top of layer), seems to cover ranges from 0 -4, which is 1 – 10,000 bars. However, the reconstructed TP profile suggests a more plausible range between 1E-3 and ~ 1E-1 bars. These ranges seem rather inconsistent. What is going on here? Is what is being reconstructed reflecting what is shown in the corner plot?

Sorry for the confusion on this, but the corner plot and the reconstructed TP profile are in different units (pascals and bars, respectively). All the figures in the actual paper are consistent.

I also appreciate that the authors ran a free retrieval on the data. It is unfortunate that the free retrieval is unable to provide consistent absolute abundance constraints with those derived from the chemical equilibrium retrieval. I think it would be informative, and transparent, to provide the (C+O)/H based free retrieval metallicity (as was done for the WASP-77 data point, Line et al. 2021) for comparison. Why am I suggesting this? The manuscript argues for showing only exoplanets with C+O based metallicity constraints (as opposed to various HST constraints what are solely based on H_2O). However, as the free retrieval is unable to detect evidence for water via the Bayes factor (as indicated in the response), it's unclear what exactly is constraining the metallicity in the atmosphere of HD149. Are the authors able to extract a C/H (and an O/H) based metallicity to compare directly to the solar system?

As discussed in the previous response, while H_2O is not uniquely identified given all the possible free parameters, if we make the mild assumptions of a scaled solar abundance pattern and chemical equilibrium then it clearly shows up. So the M/H and C/O are set mainly by the CO_2 and H_2O features that are seen (mentioned in the first paragraph), along with the weaker/non-existent features of things like CO and CH_4.

We elect not to quote a M/H based on the (C+O)/H from the free retrieval because the results are out of equilibrium at an implausible level and because we don't have good constraints on the CO (which carries a lot of the C and O).