My apologies for the delayed report, I must confess that in part this is due to an uncertainty over what to do with this paper.

The authors present a large grid of photoionisation models across a very large parameter space. The modelling is in principle useful and worthy of publication, but the current manuscript has so many weaknesses and inaccuracies that I do not feel I can recommend acceptance even with significant modifications.

That is not to say that the work in here might not become suitable for publication, and each one of my concerns in isolation could be dealt with some effort. But the sheer number of concerns and issues with the paper means I cannot foresee whether the final version would be worthy publication. Under these circumstances I am forced to recommend that the paper be rejected in its present form, but I have written a fairly

comprehensive report that I hope is useful.

Major concerns:

1. The paper is phrased as focusing on predictions for high-z galaxies but there is little appeal to high-z observations to inform the choice of parameter range to explore and as a consequence significant parts of the paper are of little interest for the galaxy population that it is setting out to explain.[Revise so that the paper is focused on low-*z* galaxies instead…perhaps “How much SF and how much AGN?” – Chris]

2. Parameter ranges. The parameter ranges explored are very much larger than any explored in the literature, and a significant part of the discussion focuses on the behaviour of the models far away from the parameters normally used for star forming galaxies. This concerns densities and the incident ionising flux and it seems the ranges of both have been chosen based on a misunderstanding of the literature. I go into details below. This needs to be fixed and the end-result should be a much reduced parameter space and I am concerned that the difference from other papers in the literature will be less clear.[Revise so that density is cut off at 10^6 and flux at the grain sublimation point. - Helen]

3. Equivalent widths given as reference. The equivalent widths of galaxies at wavelengths >~ 2000AA are sensitive to the star formation history of galaxies on time-scales >100 Myr or so, as well as the different attenuation of the continuum and emission lines

(e.g. Charlot & Fall 2000; Calzetti 1997), and the EWs of MIR/FIR lines are sensitive to the underlying dust emission model adopted. None of this is discussed in the paper and the results are therefore not as interesting/useful as they could be and the discussion and conclusion would almost certainly be modified if the grids had been discussed in terms of incident flux at Earth rather than EWs for instance. This needs substantial work and rewording of the discussion section.

4. Inappropriate modeling in the FIR. The modelling setup is acceptable for most UV and optical lines as far as I can see, but the treatment of some lines such as [S II], [O I], [C II] is not satisfactory. Cloudy has for all lines been run with either a column

density or temperature cut-off. Neither of these are well-chosen for the PDR regions where these ions emit - the typical temperatures here are ~100-1000K and column densities <~10^21 cm^2 (e.g. Hollenbach & Tielens 1997) - conditions your models would not reach (the temperature cut-off will happen before this). You could improve the [S II] and [O I] predictions by going to lower temperatures/lower ionisation fractions, but the treatment of [C II] requires more thought - or removal from the paper (probably the best option). For instance, it is normally agreed that C II originates in the warm ISM

with temperatures of ~100-200K (see e.g. Israel & Maloney 2011; Stacey et al 1991) and your stopping criteria preclude you from properly tracing this line.[Remove these lines from the plots (Helen) and analysis (Chris)]

5. Choice of ionising spectrum and metallicity treatment. The authors have chosen a reference SED with solar metallicity and with very significant WR contribution. This is sub-optimal but perhaps acceptable when the focus is star-forming galaxies in the low-z

Universe as in Richardson et al (2016), however it is completely unphysical and would warrant a lot of discussion to justify the choice. If the focus were only on relatively low energy lines - say E<20-30 eV, then ignoring the metallicity variation of the SED with metallicity is not too bad an approximation. However at higher energies the treatment of the Wolf-Rayet phase is essential and for single star evolution this depends primarily on mass-loss, the effect of binaries (e.g. the BPASS models by Eldridge et al) is also a

key uncertainty. We therefore \_know\_ there must significant trends in the hardness of the ionising spectrum with metallicity, particularly at high energies. Ignoring this effects and still using the models to interpret observations, as in section 5.1 & 5.2 is misleading and would be rather confusing for the high-z community.

Likewise the choice to cover a metallicity range of 0.2<Z/Z\_sun<5 is puzzling when the focus is on the high-z Universe where metallicities in the ionised gas are known to be <0.5 solar in general (e.g. Steidel et al 2014, 2016 or papers from the MOSDEF survey; Stark et al 2014, 2015, 2016).

To be brutally honest, with the assumptions for the ionising SED presented in the paper, the results will be mostly ignored by the high-z community, at least as long as there is no clear justification from a physically motivated point of view of the choices made. The

community is well aware of the metallicity dependence of the hard ionising continuum in stellar populations, and the challenges this pose and ignoring this fact is problematic and would as I said, probably lead to the work being ignored.

Resolving this leads to a few different options:

a) Rerun models with a physically motivated choice of ionizing spectra, tying gas-phase metallicity and stellar metallicity in some way (non-trivial and there is no agreement in the literature on how to do this - the most common is to fix them to be the same).

b) Argue that current binary models (e.g. Stanway et al 2014, 2016, BPASS2), prolong the WR phase at all metallicities so that adopting something like the Padova SED is an acceptable approximation. This is viable but would require a comparison of the Padova SED with BPASS2 SEDs to justify this assumption.

c) Remove the discussions of trends with metallicity from the paper since these are the ones that are most misleading - they indicate trends with gas-phase metallicity while keeping the ionising source at solar metallicity, a physically meaningless situation.

6. Comparison and predictions for JWST.

This section I am afraid is not useful at the moment. It ignores crucial details needed for predictions such as the IGM absorption to the galaxies and the detection limits are not properly discussed or at least not in a way that is useful to the reader. What star formation

rates, what magnitudes, what integration times, what observing plans are the authors considering, none are explained and it also seems there is a major confusion between the NIRCam and NIRSpec instruments.

I should however say that I think the inclusion of this discussion in the text \_is\_ a good idea. It just needs to be done more rigorously and perhaps be branched off as a separate section (or for that sake a different paper). Reference to the JWST ETC

(<http://jwstetc.stsci.edu/etc>/) is also advised. [I think this should be left in the paper, but I’m clueless about how to handle the observing info - Chris]

7. Dust transmission in the ISM. When you do a LOC modelling approach, the integral that is often written down, e.g. in Richardson et al (2016) is

L = \int L\_line(nH, \phi) Psi(...) dnH d\phi,

where Psi characterises the distribution of emitting clouds, it is in effect a weighting function. For applications for AGN there is a fairly clear physical model on how to choose this, but for star-forming galaxies this is a lot more complex. The essential part

here which is ignored in the present paper and in Richardson et al, is the distribution of dust in the interstellar medium.

The different clouds live in different parts of the ISM and those with highest density typically also sit behind the highest column of absorption. Note that this is not absorption \_in\_ the HII region (which in included in the modelling), but rather in the ISM

surrounding them. This must be included in LOC modelling of star forming galaxies - and it should also be kept in mind for the discussions here - otherwise you will draw the wrong conclusions.

In this case the lines and the continuum can also experience strongly different attenuation - something that leads to all EWs in the current work to be overestimates of the true observed EWs. Since this has been well known in the literature now for at least two decades (e.g. Calzetti 1997; Charlot & Fall 2000) it is necessary to discuss this in the work.

------------

Other concerns:

-----------------

Overall modelling strategy.

This is more of a subjective comment and is not in the category of major concern, but it is still a significant worry I have about the overall approach. The LOC modelling is very appealing in an AGN context where there are clear density variations and strong variations in the incident ionising flux. And I do think it is an interesting avenue to proceed down for star-forming galaxy models as well.

That said, I am not sure that the approach taken for AGNs really is the optimal to take for star-forming galaxies: We do not see significant variations in the density of HII regions across cosmic time - at most a factor of ~100 or so, much less than is seen in AGNs. This is why we often do not explore a wide range in densities in HII region models and is why I am not convinced that an integral over \_density\_ is that useful for LOC modelling of star-forming galaxies. Certainly not without some treatment of the ISM attenuation as discussed above. Incident ionising flux or ionisation parameter on the other hand \_is\_ an important variable.

Furthermore, there are other factors that arguably are at least as important: HII regions are \_not\_ coeval across a galaxy, and with life-times of at least 10 Myr one has to contend with an age-spread in the LOC modelling, HII regions are also not all the same metallicity in a galaxy - this also must be taken into account, and finally HII regions appear to be leaky (e.g. Oey & Kennicutt 1997, there are quite a few more recent papers too) so there is a significant contribution to the flux of some lines from ionising radiation that is outside the HII region (evolved stars, ionising radiation escaping HII regions). The latter point is perhaps not so important at high redshift (I do not know), but at low redshift it can be important for a LOC modelling approach.

Now this particular paper does not aim to solve all these aspects, but I wanted to bring it up because these are points that are important to keep in mind for the overall discussion.

Incidentally it is perhaps here worth consulting Kobulnicky, Kennicutt & Pizagno (1999) who looked at a related issue but focusing on the mixture of metallicities within the integrated spectrum, rather that variations in ionisation conditions.

Detailed comments (Chris handles everything below this point)

------------------

Parameter ranges:

Density: This is discsused in section 3.1.4 and it is argued based on observations of compact and ultra-compact HII region that these have densities exceeding 10^6 cm^(-3).

I am rather confused by these limits and their relevance. I do not dispute their values but they are as far as I know derived for the molecular gas density in these compact regions (this is certainly true for the Beuther et al paper) This is not directly related to the density of the actually ionised HII region that you are interested in for the simulation and such high densitities are indeed very far from any densities actually measured at high redshift (e.g. Lehnert et al 2009) which are almost always below 1000 cm^(-3).

Indeed pressure equilibrium makes this reasonably likely as well - molecular gas has T < 100K and an HII region typically ~10^4 K, so you would expect the density to be ~a factor 100-1000 lower, so going from 10^4-10^6 to 10-10^4 or so.

Thus having a density grid going to 10^10 is completely unnecessary for the application to high-z galaxies. A grid that extends to 10^4 would be more than enough. For a proper PDR simulation a higher density could be appropriate - but then Cloudy might not be the

optimal choice.

Incident ionising flux:

This is poorly explained in the text where limits from Stasinska & Leitherer (1996) and Levesque et al (2010) are discussed. It is said that (p 16): "the upper limit is set by assuming the theoretical maximum QH".

It is not clear what theoretical maximum QH is referred to, but from the numbers it seems it must be the one from Stasinska & Leitherer. However this seems to have been misunderstood: the value of QH depends on the amount of stars being formed - this is often normalised to 1 Msun/yr when given explicitly. The problem is that the Stasinska & Leitherer models are for different normalisations - 10^3 to 10^9 solar masses - now when normalising the incident ionising flux that is of course ok, but there is also a size to consider - 10^9 Solar masses within 10^16 cm gives an average density within the

region 8 orders of magnitude above your grid limit, that makes no physical sense at all. This needs serious re-assessment and will move the grid down towards what previous studies have used.

The use of the term Star-formation history

---------------------------------------------

On page 8, for instance, you start talking about the star formation history and the way this is used is confusing to me, and I believe to most people working the field.

The specific statement is that the SFH is the most important factor in the line emission with stellar metallicity having little, if any effect.

This is a rather surprising statement! The SFH should have almost no effect at high energies - all the effect should be due to the Fe/H of the stellar population.

I see that this is more or less a misunderstanding due to different ways of phrasing things so let me explain how I personally think about this (and I think this is fairly wide-spread within the field):

- For the far-UV part of the spectrum the dominant contribution to the light is in general the very most massive stars, although at E > 50 eV Wolf-Rayet stars come in as well. This means the spectrum depends on whether stars are actively forming, or whether they are not. The star formation \_history\_ - ie. the long term star formation activity is completely irrelevant on time-scales >10 Myr or so and it is not usual to refer to changes on time-scales <10 Myr as star formation history, but rather as the age of the starburst. It is indeed very true, and commonly shown, that the age of the starburst is an important factor for the far UV spectrum - in agreement with your figure 1. That is not normally called 'star formation history' and using this term in this context will confuse the reader greatly.

- The effect of metallicity on the far UV spectrum is significant, but I agree it is relatively small broadly speaking - but it has for instance an important influence on the Wolf-Rayet production which is crucial at high energies. This goes back to my point 5 above and

I will not further belabour this point here.

Thus for most people in the field, changes in star formation rate on time-scales less than, say, 50 Myr are not thought of as 'star formation history', but rather as effective starburst age. The effect of the effective starburst age is of course important (e.g. Mas-Hesse & Kunth 1999) but it is usually thought that trends on times less than 10 Myr are unlikely to be relevant for distant unresolved galaxies because there is no way you can synchronise star formation on galaxy-wide scales on time-scales shorter than 10 Myr. Thus the light from any galaxy is a mixture of different age HII regions.

Thus I would refrain from using the term star formation history about this because you will confuse your reader otherwise.

Page 2:

You write:

'Galaxy mergers commonly trigger the enhanced star formation rate (SFR) of starburst galaxies along the far left wing of the BPT diagram' - where do you get this from? I am not aware of a clear study saying this - it is true that Galaxy Zoo classifications show a slight increase in the galaxies with 'merger' classification in this part of the BPT diagram but that does not in any way mean that they truly are mergers, nor that they SF is truly driven by merger activity (e.g. Robaina et al 2009 for a slightly higher z perspective).

You write: 'in such galaxies' - What does 'such' refer to here - starburst galaxies or those on the far left wing?

Page 3:

'Similarly, at lower ionization' - Do you mean lower ionization parameter? Lower ionization in itself does not really make sense.

'could likely be the result of starlight, non-thermal sources, or a combination of the two' - This is surely true in all cases and has nothing to do with where the galaxies lie in the BPT diagram as such.

Paragraph starting: 'High ionization potential emission lines have historically signified AGN activity' - this is repeating the preceding paragraph, tidy up.

Sharazi -> Shirazi

Page 4:

'without any signs of AGN' - What does this in particular mean? AGNs may be present in many of the samples, but not detected with current facilities.

'This is largely due to the fact that at progressively higher redshifts one would expect a harder ionizing continuum due to the contribution of metal-poor (or metal-free) massive stars.' - This is true, but a given selection function will moderate this - z~2-3 star forming galaxies that are in current samples are not that much metal poor than local galaxies. See e.g. Steidel et al (2014, 2016), or any study of the z>2 mass-metallicity relation.

'LBG surveys have confirmed these features but have also shown positive equivalent widths for He II λ1640 and Ly-a (Cassata et al. 2013).' - Ly-a has been seen in emission for ages, that is certainly nothing new with Cassata et al! He II was also seen in emission in Shapley et al 2003, and quite a few papers before then as well - the challenge is whether a) it is nebular rather than a Wolf-Rayet feature and b) whether the ionising source is an AGN or star-formation. This should be updated and clarified.

', with densities on average an order of magnitude higher than those found in the local universe' - What 'density' are you referring to here? The electron density in ionised regions is one quantity, what you are using for the calculations. However the other density of relevance is the density of the ISM. These are not normally linked in Cloudy

calculations (like yours), but they are linked physically in any real system because the ambient pressure (ISM) constrains the expansion and evolution of a star forming region. Please clarify this.

'Such observations are consistent with Population III (PopIII) stars' - That is a very strong statement and I think it would be hard to fully justify this, precisely because of the strong metal lines. A Pop III generation of stars should be effectively zero metallicity and many of these features might not be seen. There are arguments that say we have seen Pop III signatures - but the more relevant references would then be e.g. Sobral et al (2015 - CR7) and Kehrig et al (2016, I Zw 18). Please expand and clarify.

Note that the theoretical models at the moment do not allow for Pop III stars to exists at low (z<10 or so) redshift because of very quick enrichment. Whether they are right is a different matter of course!

'field is then capable of creating more highly ionized elements' - This does not follow logically from the preceding - a 'strong UV continuum' would mean a UV continuum that has a very high luminosity, but pumping up the luminosity will not allow more highly ionized elements to be seen. What I presume you mean, is a harder UV continuum. Even so the sentence is imprecise because the radiation field does not 'create' ionized elements. Please clarify and fix.

'Indeed, even the deepest ....' - I must admit I can not really make sense of this sentence - what 'line of reasoning' are you referring to, and what is the relevance of a N IV] emitter as a young starburst as compared to a massive evolved stellar population? 'Evolved' often means 'old-ish', so typically >1 Gyr. Please clarify.

Page 5:

'Follow-up work by Levesque et al (2010) ....' - This has been done many times before as well of course - for instance in Charlot & Longhetti (2001, Figures 1-3), Mas-Hesse & Kunth 1999, Mas-Hesse et al 1991, while the study in Levesque is clear and well done this phrasing makes it sound as if it was first done there which is misleading. Please be a bit more comprehensive in your referencing.

'... has proven useful in fitting galaxy spectra with large He II / Hb values observed in the local universe' - This is a statement with some qualifications I would say - their modelling was not very successful at matching the large He II/Hb ratios (see also Guseva et al 2000), but it does indeed work for other strong lines (as has been demonstrated for other BCD-like galaxies by other authors as well).

Page 7:

'Indeed, Ly-a at z> 6 becomes attenuated leaving...' - I am not sure 'attenuated' is the right word here as it seems to imply that Ly-a is not visible due to dust attenuation in the galaxy. This is unlikely to be the full story and the true reason for the decline in the number of Ly-a emitters at high redshift is unclear and depends on radiative transfer effects for which there is no current consensus. Please update and clarify - if you want to look at some Ly-a modelling papers, then e.g. Verhamme et al 2006, Dijkstra et al 2006; Dijkstra & Kramer 2012 are good places to start.

Page 9:

'as such, we have chosen the hardest SED possible for our baseline model.' - this goes back to my point among the major concerns above. I really do not see how your conclusion follows from the preceding argumentation. If you want to determine the conditions necessary, you surely need to choose the \_most correct\_ SED possible. Choosing an unrealistic SED would potentially lead you to make the wrong inferences about the conditions necessary.

Figure 1 - the discussion of this in the text does not match the labels on the sub-panels.

Page 10:

'overall hardness ... is fairly similar for non-rotating and rotating stars...' - I agree this is the case for the continuous SF models but I strongly disagree with this statement for the bursts - your panels have six orders of magnitude on the y-axis, any small shift is very

significant.

'.. to reach steady state' - please explain what steady state means here.

'... become much more apparent.' - Again, this statement is a bit confusing: for SSPs it is apparent also in Figure 1. Please tidy up.

'... stars begin to skip the WR phase' - 'Skip the WR phase' is a misleading phrasing - it makes it sound as if it goes on the same evolutionary track but just jump over the WR phase. I am sure this is not what you meant, but it is probably more correct to say: 'the less massive stars do not have sufficient mass-loss to enter the WR phase' or something like that.

'... the steady state Padova ... hardest ionizing spectrum' - firstly, do not use steady state to refer to the SED, and secondly this is surely correct, but this is because it is at solar metallicity and therefore has a vastly higher production of WR stars than at lower

metallicity. It is therefore not a suitable SED to use for lower metallicity simulations.

'... only included secular evolution ...' - what does secular evolution means here?

Page 12:

Where you refer to Shirazi & Brinchmann (2012) and Abel & Satypal you should really also refer also to Guseva et al (2000) who also studied He II/Hb in starbursting galaxies (following up from earlier work by Schaerer et al 1999), and also arguably Brinchmann, Kunth & Durret (2008) who also looked at He II/Hb and the relationship to WR stars

and also of relevance here is Thuan & Izotov (2005) who looked at Ne V, Fe V and He II emission in low metallicity galaxies.

'... peak Wlambda of ... is 5 times greater for Padova continuous evolution track than the Padova instantaneous evolution track.' – This is rather puzzling and is not really the normally expected case, certainly in the UV, if you compare the stellar populations at a

comparable time. The WR population in each case should be similar at say 5 Myr but the continuous SF model should have more bright stars contributing to the continuum than a SSP model - in which case the EW should always be higher in the SSP models.

Basically this is a statement about the relative life-time of the stars that dominate the energy production for the emission line and for those that dominate the continuum. It would be useful to understand better this statement as it is so surprising - for instance, at what time you do the comparison, over what time-range it holds etc.

Page 13:

This is a page where my concerns on the stopping criteria come to the fore. What you have chosen is probably ok for the main lines you consider here but it is a concern for other crucial lines like [N II], [S II] and [O I]

'Full dust abundances are based on Orion (Baldwin et al. 1991)' – why is this relevant for high-z conditions? And are you specifying dust abundances or depletion patters or something else? Please give more details on this.

Page 15-16:

These are the pages where the ranges are given - I discuss my concerns

about this above.

Page 17:

Figure 3 - This really highlights the big discrepancy between the grid size and the relevant physical conditions. And why only those studies? Why not e.g. Dopita et al (2013); Charlot & Longhetti (2001) etc.?

There are also many measurements of electron temperatures in HII regions both at low and high redshift, using the [O III] 4363 line. These all agree that T\_e < few x 10^4 so the majority of the grid is not important for the class of galaxies that the paper focuses

on.

Page 18: '... this does not pose too much concern as most of the emission lines we track do not optimally emit ... .' – But these are close to the densities that we generally find at low and high redshift. It is exceedingly rare to find HII regions with densities above 10^3 cm^-3. Firstly I do not think this is a problem - it is what happens, but the sentence should be rewritten.

'Note that the three studies ... parameter space that represent HII regions ... we also explore the more extreme conditions...' - I would recommend looking at some of the literature on low redshift analogs (e.g. Borthakur et al 2014; Brinchmann, Pettini & Charlot 2008; Liu et al 2008; Bian et al 2016). It is indeed true that something like Orion

is completely unsuitable for high-z work (or even comparing to extreme star-bursts in the local Universe), but there is a large body of work indicating that there are star-forming regions at low-z that are showing comparable conditions to those at high redshift. Stars do not change \_that\_ much with redshift. It is true that you cover a much large range of parameter space, but most of it is irrelevant for the application you have in mind.

Page 19: 'While some optical and infrared emission lines emit in these extreme regions, the efficiency of reprocessing is generally very low making emission lines weak.' - This is true, but be aware that for applications to low redshift star-forming galaxies this emission might be very relevant (not important for your specific goal though I admit).

Page 24:

'do not predict ..., but create enough emission to be detectable by current optical instruments' - what does this mean? that is a completely non-quantitative statement and it must be made clearer. What are 'current instruments' and what are their limits? Furthermore He II 4686 (nebular) is often mixed with He II 4686 (WR-wind) and disentangling the two might also be challenging. Please make clear what you mean here.

Page 25:

You comment on the collisional deexcitation of [O III] which renders 4959+5007/4363 less useful as a thermometer - this is correct, but not entirely relevant for star-forming galaxies as densities almost never reach these levels.

Page 26:

[N III]3869 -> [Ne III]3869

Section 3.3.3: '... for selected optical ...' -> '... for selected IR

...'

This section is also one where you need to discuss dust emission mechanisms and their effect on the equivalent widths as I comment on above.

Page 27:

'Taking these ratios on our grids indicate that our simulations have temperatures around 104 K with log(nH) ~ 3.0, which is consistent with Figure 3.' - What does this sentence mean? I understood your 'simulations' to be the Cloudy runs, but here that is clearly not the case - what simulations are you referring to? Please clarify!

Page 28:

'high [Ne V] emission is likely due to AGN activity, however the simple presence of [Ne V] emission should not attributed to non-thermal excitation.' - this is not a sufficiently clear sentence. What does 'high' mean here? When (and how! see e.g. Thuan & Izotov 2005) should I conlcude it is AGN?

The discussion of [C II] here suffers from the concerns regarding the modeling I detailed above.

Page 29:

The metallicity range chosen (0.2<Z/Zsun<5) is a strange range for the purported goals: high-z galaxies are unlikely to have Z > 0.5 Zsun (lots of references, see e.g. the MOSDEF papers) and in many cases even lower metallicities are found to be needed (e.g. ~0.1 Zsun Stark et al 2015;2016). Z much larger than solar are essentially only found in the central regions of low redshift galaxies. This really needs better justification - as it stands it is not very useful for high-z, where a range 0.001 < Z/Z\_sun < 1 would be more relevant perhaps.

'we chose to only study the effects of varying metallicity of the cloud.' - I commented on this above, I find this problematic. The SED, stellar evolution and the gas-phase metallicity are all linked closely.

Page 30:

Equation 4: What is xi?

'This pocket of no emission was neither present in our solar simulations nor in our subsolar simulations.' - So when is it present? This sentence makes little sense as written. Please clarify.

'(whose ionization potentials is similar to that of [O III] λ5007: 47.9 eV and 54.9 eV respectively)' - Why do you bring up the ionisation potential for O++ to O+++? [O III]5007 is a forbidden line so the ionisation to O+++ is not the relevant step - rather it is the ionisation from O+ to O++. You should also be able to determine what process dominates for C III 977 emission in this region by looking at the Cloudy output. Please fix & improve.

Page 35:

'as Ferland et al. (1996) discuss, He II λ1640 decreases with increasing Z due to the increased ....' - These days this is more usually phrased with respect to the incrased opacity in the stellar atmospheres and the change in WR star content which you do not discuss here at all. Ferland's discussion is of course correct, but only relevant for situations where the shape of the ionising continuum does not significantly change. Since there are many more WR stars at higher metallicities (due to the increased winds), the He II emission is also boosted significantly relative to low-Z, in contrast (but not

disagreement!) with Ferland's argument.

'In regards to ....' - why do you bring this up? Your grid does not extend to <0.01 Z/Zsun. The whole paragraph is irrelevant where it is placed (the point is fair enough though).

Page 37:

'It should also be noted that ([O II] λ3727 + [O III] λλ4959, 5007)/Hb acts as a metallicity indicator.' - This is true, and references should be given - e.g. Pagel et al 1979; Mc Gaugh 1991; Kewley & Ellison 2008 for an overview, many more exists.

'(Nagao, Maiolino and Marconi, 2006, Raiter et al. 2010).' – A surprising choice of references - this has been well known at least since McGaugh 1991 and while these two papers certainly are good, they are not particularly known for the R23 analysis.

Page 38:

'We find that nearly all the gas is ionized in the regions in which [C II] emits' - This is a consequence of how you set up your models and is not a reflection of the physical truth. In fact C II mostly emits in the PDR where a lot of elements are not significantly ionised. (I presume your sentence referred to H and He but you should make this clear).

'Our predictions regarding the emission of fine structure lines are based on Hα emission, which is relatively flat across our grids' – I have absolutely no idea what you mean with this sentence, please clarify!

'Thus, we can only compare differences in the peak equivalent widths of [O I] 63 μm and [O III] 88 μm at constant phiH and nH values as indicators of differences in SFR. Otherwise, these differences should be interpreted as differences in emission based on the adopted Hand nH values. Observers should thus be cautious when using these fine-structure lines as indicators of SFR.' - Again, I have no idea what you are trying to say here, I can't even start to guess. Please clarify!

Page 40:

Figure 6a (and 6b, 6c): The labels of the lines are Cloudy specific and not easily

understandable to a general reader. They should be changed to actually include what lines are considered. Also the shape of the plots with 4.5 orders of magnitude on the y-axis and very compressed vertically makes them very hard to read. A different aspect is needed to be able to read the figures properly.

Page 42:

This discussion conflates the starburst age with the age of the galaxy. This is misleading and unlikely to be true physically.

'O-type stars tend to dominate the luminosity of starburst galaxies.' - It is more correct to talk about H II region - there is little reason to believe that a galaxy will form on time-scales < few Myr.

'The equivalent widths of many of the strong hydrogen recombination lines like Hα, Hβ, or Brγ can be used as age indicators because they measure the ratio of the young, ionizing over the old, non-ionizing stellar population.' - This is to some extent correct - this is for

instance discussed in Mas-Hesse & Kunth (1999) - it is an old result and some references would not go amiss here.

Page 43:

'Geneva and Padova continuous tracks, however, continue to emit constantly across the 6-8 Myr range.' - Surely the tracks themselves do not emit! Please rephrase.

Page 46:

'The physical reason for this apparent contradiction is that dust makes a substantial contribution to the overall opacity in our dusty simulations, which decreases the availability of high-energy photons to ionize and excite the gas.' - This explanation makes some sense for Ne V and Ar IV, but not for Si II and Mg II which are both low

ionisation energy lines.

A key factor here is likely dust depletion. Indeed the main effect of dust in HII regions is depletion of metals onto dust grains (e.g. Shields & Kennicutt 1995) and a discussion of depletion and its effects should enter here as well.

'Many of the equivalent widths of UV emission lines increase with the removal of dust since dust absorption peaks in the UV' - But in your model surely the dust attenuation of the continuum and the line is the same, so why should you see an effect on the EW?

'... while the region it emits across the LOC plane essentially disappears...' - It seems to expand significantly to my eyes?

'dust free models either get incorporated into the larger emission region in the plane or disappear, best seen with [S II] λ4078 and [S II] λ6720 ' - why?

Page 48:

'Thus we agree with the conclusion that AGN contribution is needed for local [Ne V] 14 mum and 24 mum emission' - This is insufficient – you have not considered shocks, nor X-ray binaries, nor the effect of normal binaries or uncertainties in the SED-shape, so how can you make this conclusion?

'We find that peak He II λ4686 emission does not change significantly as we vary from 0.2 to 5.0 Zsun' - Well it does if you change the ionising spectrum with metallicity which you should do. Your conclusion is caused by your assumptions and does not reflect

reality. Please fix.

'Thus, we suggest that perhaps there are low metallicity pockets within these local galaxies, which contribute to their overall He II λ4686 emission.' - I really do not understand how this statement follows from any of what you have written above. Please rewrite.

Page 50:

Here you focus again on equivalent widths but you should be aware that these can be hard to measure at high-z!

'With our dust-free simulation, we find a much higher peak log(WC III]) = 3.0 with typical emission around log(WC III]) = 2.0 (Figure 7a, column b). In our dusty low metallicity (0.2 Zsun) simulation, we find peak log(WC III]) = 2.7. ' - yes, but for completely unrealistic conditions. Revisit with application of densities actually observed in high-z galaxies.

'We find O III]1665...' - what line is this?

p 53:

The JWST comparison is confused and not sufficiently quantitative. While NIRCam has a grism, it operates at long-ish wavelengths (2.4µm-5µm) and will not be JWST main spectroscopic instrument. That will be NIRSpec and maybe you have mixed NIRSpec and NIRCam?

'With peak log(Wλ) around 2-3, these emission lines should be bright enough as well.' - for what? What exposure time, what assumptions about host galaxies, what SFR? There are no details and this part is not useful as it stands.

Page 54:

'We predict that C III λ977 and C IV λ1549....' - C IV might be, but why this CIII 977? It would be completely killed by the intervening IGM. Anything below 1216Å in the rest-frame is likely to be useless at high redshift due to IGM absorption.