We would like to thank the referee for the comments and suggestions to improve our work. We have attempted to address the points raised by the referee. Before discussing our detailed response, we would need to point out that the aim of the paper is the first of many steps to account for the colour gradient bias for Euclid. In particular, it explores the feasibility of the planned approach and what data may be needed to achieve this. In contrast, the actual calibration for the bias will require the exact shear measurement method, which is not available now. Thus we do not plan to study the details of bias dependence on the galaxy morphology/model, which might be different, but rather wish to provide a general picture of the colour gradient bias, and also estimate the necessary number of galaxies from HST. We admit that there lots of work remains to be done to implement this in the actual data analysis, which involves multiple groups within Euclid.

We enlarged the labels in some figures. All the changes in the draft are shown in bold font. The detail responses are given below.

Page 1, second column, I think Jarvis et al. 2016 is not the right reference - this presented the shear catalogs rather than the cosmological constraints https://arxiv.org/abs/1507.05552.

We corrected the reference.

Page 4, first column, first paragraph. The description of the galaxy properties is clear, but some justification for the choice of properties (sersic indices, half-light-radii, bulge-to-disc flux ratio), is required e.g. you could provide some references that demonstrate these choices are representative of the expected Euclid weak lensing sample.

We actually had some internal discussion about this issue, but we chose the galaxy models used by Semboloni et al. 2013, because these galaxy models have by construction a large colour gradient compared to general galaxies. We use two galaxy sizes, where the small one corresponds to the smallest galaxies in the Euclid weak lensing analysis (i.e. the possibly largest colour gradient bias we may have). At the current stage, we do not plan to cover the full Euclid weak lensing sample. Rather the aim is to show that in order to calibrate the colour gradient bias, two narrower band images from HST will be sufficient for most of the galaxies (for those having strong emission lines we may still need more information, maybe more bands). Moreover, we want to estimate how many HST images we need to perform the calibrations. Also we do not plan to perform bias calibration at this stage, since the colour gradient bias fully depends on the measurement methods, for which we have already made some tests; but to really perform the calibration, we need to apply the shear measurement method that is going to be use for Euclid, which is not possible at the moment.

-Page 5, second column, first paragraph. I found this a bit confusing, you say first that "Provided the average bias that is caused by colour gradients is known for a selection of sources, it is possible in principle to obtain unbiased estimates of the ellipticity correlation function", but then you also say it is important that the correction for the average bias accounts for variation in redshift and color. It is unclear whether for a selection of sources a single average correction could be applied, or e.g. a correction that depends on galaxy colour needs to applied e.g. on a per-galaxy basis.

The correction will be applied on per-galaxy basis, but based on its colour and redshift (and perhaps other parameters). We remove "average" in the sentence, which may cause confusion. Note that the correction for an individual galaxy may not be quite correct, but for an ensemble of (similar) galaxies it is. Hence the correction introduces some additional noise in the ellipticity measurement, but as the CG bias is so small, this is completely negligible compared to the intrinsic ellipticity.

Section 3.2 is clearly explained, but some details about the filters would be useful. e.g. do the HST filters used span the range of the Euclid filter?

We add this information in the new version (footnote in page 5). The filters do span the Euclid VIS range.

Page 5, 2nd column final paragraph: Is there color gradient bias in the simulated HST images, since they are also deconvolved with an effective PSF? I assume that this is accounted for (i.e. will contribute to the measured delta m in figure 5) by using a wavelength dependent PSF for the HST simulated images, but I think it is worth stating this explicitly.

Yes, this is a good point and is the reason for the delta m in Fig.5. We add a few explanations in the second paragraph of Section 3.2.

Page 6, first column, line 36: "The residual bias is within the tolerance for Euclid...". For galaxy S, this does not appear to be the case for z~0.5, please explain this. Does the Euclid requirement shown represent the whole systematic error budget for multiplicative shear biases? Please comment on whether the achieved accuracy is likely to be sufficient given that there will be other sources of multiplicative biases.

This requirement from Euclid is only for colour gradient bias. We add this in the caption of Fig.5. We explain this below Eq.15. This actually also relates to the last point. The emission line caused large colour gradient cannot be accounted for in the method using two bands.

Page 6, 2nd column, line 38: This seems a potentially important point. One of the stated improvements over S13 is that noise is included in the HST images, but given that real galaxies are not bulge+disk, could your deconvolution method potentially intoduce model biases to the CG correction? If so, please include some estimate of the size of this effect.

There is certainly some bias in our CG estimate. We cannot give a general estimate for this now, since it depends on several factors. We make additional tests, simulate the galaxy image by bulge + disk, but fit them with one Sersic profile. It does cause some extra error, which can be either an over- or an underestimation, and it depends on several factors of the galaxy images. It requires future work, but currently the estimate is sufficient for our purpose, i.e. we show that the bias itself is not expected to be very large and that noise in the HST data does not prevent us from estimating the bias. But the actual bias estimates will likely depend on morphology and require a more careful handling of the HST data. This work is currently ongoing.

Page 7, 2nd column, last line: "This results the effective PSF in..." Please correct.

Done.

Page 8, 2nd column, last paragraph: "which will not cause significant selection effects in our analysis" Please add some quantitative justification of this statement.

Schrabback et al. (2016) automatically mask and remove very extended low-$z$ galaxies, which are otherwise often falsely deblended into substructures. As such galaxies will likely be removed in a Euclid weak lensing analysis as well, this should not cause a relevant bias. For their KSB shape measurements Schrabback et al. (2016) additionally remove large (but not extremely large) galaxies based on the half-light radius due to postage stamps restrictions. As some of these galaxies may enter a Euclid analysis, we do not employ this additional selection here. We now clarify this in a new footnote.

\* Section 4.1: Any differences in methodology between section 3 and 4 are concerning, since section 3 is meant to validate the results of section 4. In particular, only a single sersic index is varied in section 4, whereas two are varied in section 3. Furthermore, you say there were a large fraction of failures when attempting to use two Sersic indices on the HST data. Given that this was not observed in the simulations in section 3, does this mean that those simulations were not representative of the HST data? If so, how are the results of section 3 applicable to section 4, and why should we believe the results of section 4? Apologies If I've missed the point here, if I haven't, then the treatment of this issue needs to be significantly improved to demonstrate the results in section 4 can be treated as having the accuracy estimated in section 3.

This is indeed a relevant limitation of the CG bias analysis. We admit that the current simulations cannot fully represent the HST data. The main reason of failure fitting two Sersic models is that some galaxies cannot be modelled by that. It requires some further collaboration in the future with other groups, including the image simulation and shear measurements. For the current problem, we perform two additional tests: 1) in the simulations, we also fit the galaxies with one Sersic component. The estimation has some extra scatter, but can reflect the main properties of the bias. 2) We also fit the HST data with two Sersic components, in the successful fitting (about 70%), the estimation shows a similar relation with the colour of the galaxy, although the scatter is larger. We add a small part to explain this point at the end of section 4.1.

- Figure 14: Caption says "(top)" and "(bottom)", should be "(left)" and "(right)"

Done.

- Page 11, first column, first paragraph: "biases for individual galaxies are generally well below 1%; much smaller than the intrinsic shapes of galaxies". I don't think a comparison of the biases to intrinsic (dispersion in) shapes of galaxies is useful here, unless "biases" here means just random noise on the corrections, rather than coherent biases. Please clarify.

We just remove the second part of the sentence.