We thanks the referee for their time in reviewing our paper. In the following we reply in turn to the points raised, and provide further information for clarification when required:

1. *The authors study the projected sensitivity of the LHeC for constraining PDFs within the framework of the PDF4LHC15 set. The PDF4LHC15 set has been averaged from MMHT, CT14 and NNPDF3.0. The latter set has been updated to NNPDF3.1 with significant changes, thus the PDF4LHC15 set is slightly outdated by now. More importantly, the PDF4LHC15 set mixes PDFs derived using very different theoretical treatments. This concerns for instance different heavy-quark mass schemes (ACOT, FONLL and RT) or the gluon PDF in the low-x or the high-x region. This is problematic as the chosen statistical combination of PDF sets may lead to an overestimate of the uncertainties at these kinematics. Such issues have been discussed extensively in: A Critical Appraisal and Evaluation of Modern PDFs A. Accardi et al. Published in Eur.Phys.J. C76 (2016) no.8, 471 DOI: 10.1140/epjc/s10052-016-4285-4 e-Print: arXiv:1603.08906 [hep-ph]*

   The referee is right that this is not the only possible choice, and there are certain aspects (more recent data, flavour schemes etc) that are omitted by taking PDF4LHC15. On the other hand, no more recent combination is currently available, and taking e.g. only one individual set (CT, MMHT, NNPDF, ABM...) risks being under–conservative. Therefore we view PDF4LHC as the most appropriate baseline for the sort of study we wish to perform here. We have added some discussion on these points, and a reference to the paper above, to the beginning of section 2.2.

2. *It is unclear, why the authors use the HERAPDF parameterization as an example for a PDF parameterization with a restricted number of parameters. This parametrization, subsequently used in the ATLASepWZ16 analysis lacks sufficient flexibility in the light flavor PDFs and has been shown to fail in comparisons to data, most recently in: Measurement of associated production of a W boson and a charm quark in proton-proton collisions at s=13 TeV, CMS Collaboration (Albert M Sirunyan et al.) Published in Eur.Phys.J. C79 (2019) no.3, 269 DOI: 10.1140/epjc/s10052-019-6752-1 e-Print: arXiv:1811.10021 [hep-ex]*

   The referee is right to point out these potential issues with insufficient flexibility. Indeed, this is precisely the point we are investigating here. That is, if one takes an overly restrictive parameterisation in these type of projection studies, then one may over–estimate the impact on the corresponding PDF uncertainties. We take HERAPDF as a concrete example of this. Indeed, this set has been used in earlier previous LHeC PDF projections, in particular those presented in the LHeC CDR, arXiv:1206.2913 (while more recently some additional freedom has been added, but always being HERAPDF–like). We agree with the referee that there is evidence that this parametrisation is not sufficiently flexible to describe all available collider data, which is something that we discuss in our manuscript.

3. *The authors use pseudo-data for the heavy-quark neutral-current structure functions to explore additional constraints on the gluon PDFs. It is well-known that the gluon PDF in these reactions is correlated with the strong coupling constant and the quark masses mc and mb. See for example ABMP16: Parton distribution functions, alphaS, and heavy-quark masses for LHC Run II S. Alekhin, J. Blumlein, S. Moch, R. Placakyte Published in Phys.Rev. D96 (2017) no.1, 014011 DOI: 10.1103/PhysRevD.96.014011 e-Print: arXiv:1701.05838 [hep-ph] Fixing*

the strong coupling constant and the charm and bottom quark masses (in some renormalization scheme) introduces a bias in the projections. At the same time the LHeC's potential in constraining α_S and heavy-quark masses remains unexplored.

In terms of bias, this should not be an issues for our study, as we are only interested in a closure test, where the pseudodata by construction in agreement with the theory, i.e. with the same values of strong coupling and quark masses. In other words, we are only looking at the impact on the PDF uncertainties and not the central values. So taking a different choice of input strong coupling or quark masses would not have a significant effect on this.

On the other hand, the referee is right to point out the sensitivity of the LHeC measurements to the value of the strong coupling constant and to the heavy quark masses. A study of this is however beyond the scope of the current paper. Moreover, while in principle it is true that a complete set of LHeC PDF projections should include some estimate of the possible theoretical uncertainties associated to the uncertainties in the values of the $\alpha_s(m_Z)$ and m_c, m_b used in the fit, it has been demonstrated by related studies that the LHeC data itself can be used to constrain these fundamental SM parameters down to the point where their contribution to the total PDF uncertainty can be neglected.

In particular, a recent overview of LHeC studies in this respect can be found in the DIS2019 workshop talk by Claire Gwenlan

`https://indico.cern.ch/event/749003/contributions/3330763/attachments/1828248/2992940/DIS19-LHeC_Gwenlan.pdf`

showing that α_s can be measured with a precision of 0.15%, and similar considerations apply to the m_c and m_b measurements.

We have added a sentence to the end of section 2.2 discussing this point, and providing a reference to 1701.05838.

4. *The authors work with theory predictions in QCD accurate to next-to-leading order (NLO) while deep-inelastic scattering is known complete to next-to-next-to-leading order (NNLO) and some reactions even to N3LO. The low theoretical accuracy leads to a lower quality of the theoretical description (larger value of χ^2/dof.) of existing data and thus to PDFs, which have larger uncertainties.*

   While we certainly agree with the referee that for the description of actual experimental data it is mandatory that use theoretical predictions based on at least NNLO QCD calculations, here we are dealing with simulated pseudo-data for which by construction NLO QCD provides a satisfactory description, and therefore NNLO QCD corrections are not required (though the analysis could also be performed using NNLO calculations and no qualitative differences would be expected).

5. *The choice of a tolerance criterion $\Delta\chi^2 = T = 3$ is not well motivated in the study. It is currently used in some PDF fits to account for the lack of compatibility of data sets, although this is not necessary in a global fit. If a value of T different from unity is used, this mixes the projected PDF sensitivity of the LHeC with the question, whether compatible sets of data will be available then. In this context the authors' statement on page 10: "On the other hand, it has been known for some time, see e.g. [41], that fits to the HERA dataset only are found not*

to be consistent within the quoted uncertainties in comparison to those including collider data when using a textbook tolerance $T = 1$." should be revised as the ABMP16 paper cited above and reference therein provide counter examples.

The choice of a $\Delta\chi^2 = T = 3$ tolerance criterion is required by consistency given that this is the same approximate tolerance that defines the 68% CL intervals in PDF4LHC15. The use of a different tolerance, in particular $T = 1$, would not be consistent since then the LHeC pseudo-data would be treated in the global fit with a higher relative weight than all other experiments.

This said, precisely to investigate the point raised by the referee, in Fig. 4.2 we compare our baseline results, obtained with $T = 3$, with those corresponding to $T = 1$. As expected, with $T = 1$ there is a more marked error reduction, though the main qualitative findings of our paper are not changed.

As in any sensitivity/projection study, the choice of prior assumptions plays a crucial role here. If one assumes the HERAPDF parametrisation and a HERA-only dataset then one can consistently use $T = 1$ and then finds in general a more significant error reduction than in the case of using the PDF4LHC15 prior. But as we discuss in the paper such parametrisation is not able to describe all available precision collider measurements, as the PDF sets that enter PDF4LHC15 manage to instead.

It is beyond the scope of this paper to further discuss the motivations to use a tolerance criterion in the framework of a Hessian global PDF analysis, and here we just emphasize that our choice of $\Delta\chi^2 = T = 3$ is required by the internal consistency of the analysis.

6. *The authors' findings on the improvements of the light flavor PDFs through LHeC data are not entirely clear, since recent LHC data on W- and Z-boson production, differential in rapidity, has greatly increased the precision of the light-flavor PDFs. Such data are, however, not included in the PDF4LHC15 set. It is thus not clear, whether the LHeC can add substantial new knowledge.*

   In order to address this point, we have redone the LHeC profiling using NNPDF3.1 Hessian as baseline PDF set, which includes some of the most recent precision W and Z measurements, including the ATLAS 2011 W, Z rapidity distributions, the complete LHeC W, Z 8 TeV dataset, and the ATLAS and CMS Z p_T distributions at 8 TeV. As is shown in the new Fig. 4.3, even in this case one finds that the LHeC will have a marked impact of the light quark flavour PDFs, confirming the main results of our study obtained using PDF4LHC15 as the baseline PDF set.