review from referee 1

Referee report on manuscript zt5005 by Sanchez del Rio et al., entitled "X-ray lens aberrations retrieved by deep learning from several beam intensity images".

Sanchez del Rio et al. present a convolutional neural network (CNN) for retrieving the error profile of X-ray lenses using only the intensity of the propagated beam at several distances around the focal position. The methodology holds great promise for optical metrology and more broadly for improving the accuracy of simulations of real optical systems. While the subject matter will be of significant interest for readers of Journal of Synchrotron Radiation, the manuscript needs to be substantially revised to make it suitable for publication.

Below various issues are highlighted that should be addressed in the revised version:

Title:

1.	The title mentions lens aberrations, which are calculated as part of the methodology for retrieving the lens error profiles. However, the derived aberrations are never presented in the manuscript in any figure, instead comparisons are always made using the lens error profiles. The manuscript title should reflect that the article concentrates on the lens error profiles rather than the aberrations.

Abstract:

2.	The "phase problem" is mentioned in the Abstract but never referred to again. Potentially this work could be related to the phase problem, but this needs to be at least expanded upon in the introduction if it is to remain in the Abstract.
3.	In the last sentence of the abstract the authors claim that their work demonstrates that their method can be extended to other optical systems beyond an X-ray lens, but the manuscript does not expand upon this point at all.

Introduction:

4.	The authors say "they [the surface errors] are measured by the metrology laboratories that most synchrotron facilities have" (p2, 2nd paragraph). One should straightforwardly cite a couple of articles describing appropriate examples from various synchrotrons such as the ESRF.
5.	It is stated the "The shape of these profiles is compatible with the experience learned from direct in-situ experiments (Celestre et al., 2022)" (p3, 2nd paragraph). This sentence is somewhat vague. Are the authors saying that the profiles are similar to those in the cited manuscript? That experimental paper describes the measurement of 2D profiles of diamond and beryllium lenses, so it is difficult to compare to the 1D radial profiles presented in this manuscript. In addition, the geometrical aperture of the real lenses studied in the previous publication are much smaller than simulated here (~ 500 um compared to ~ 1500 um).

Methods:

6.	While the authors say that the studied optical configuration is "a single X-ray lens illuminated by a monochromatic X-ray beam" (p4, 1st paragraph), it is not clear whether this is truly monochromatic (zero bandwidth) or instead has finite bandwidth (~ 1 eV). Our reading of the text suggests the former, as the authors never discuss the effect of a finite energy bandwidth on their simulations. Of course, in practice a DCM would typically be used to monochromate the beam, but no DCM is shown in the screenshot presented in Figure 1. The authors should make comment regarding the implications of using a DCM for a real beamline, especially the potential impact of thermally-induced deformations on the 1st DCM crystal, which would surely affect the shape of the wavefront of the beam, and therefore impact upon the methodology described in the manuscript for measuring the lens error profiles.
7.	Figure 1 (p5) shows the OASYS workspace used for the modelling. If the two slits shown in the model are fully open (and therefore do not affect the simulation) surely it would be clearer for the reader to exclude them from the figure altogether. In addition the widget captions are rather small and some are

overlapping.

8.	The authors should clarify the type of aberrations that all of the included Zernike coefficients correspond to (p6, 2nd paragraph). At present this is only described for Noll numbers 22 and 37.

9.	It is not clear what the purpose is of Figure 2 (p8). One would expect that there would be more structure in panels (b) or (c) which would be encoding information about the error profiles shown in panel (a). From Figure 2 it is difficult to see how one would be able to retrieve such profoundly different error profiles (panel (a)) based on the information in panels (b) and (c). The authors should clarify this point by either providing more discussion of the simulations presented in Figure 2, or by recreating this figure with more illuminating data in panels (b) and (c).

10.	The text explaining the CNN used in the model (p9, 1st paragraph) is almost entirely lifted from Saha et al., 2020. The authors should reword this text, and should further highlight the very close correspondence of their CNN with PHASENET.

Results:

11.	The initial part of the Results section (p9-10, until the beginning of Section 3.0.1) should be in the Methods, as it provides more details about how the CNN has been adapted from PHASENET (Saha et al., 2020).

12.	In Section 3.0.1 the authors state "It is remarked (Fig 3a) how the learning slope reduces at about 300 epochs." (p10, 2nd paragraph). Do the authors mean it is remarkable, or they just want to mention it? Is it possible for the authors to provide any insight about the CNN from the inflection points in Fig 3(a) and Fig 3(c)?

13.	There is significant overlap between the material covered in the Results and Discussion sections, and potentially the two could be combined.

Discussion:

14.	The authors explain that partial coherence of the X-ray beam was simulated using the coherent mode decomposition method introduced by them in a previous article (Sanchez del Rio 2020) (p15, 2nd paragraph). It appears the authors have used exactly the same description of the partially coherent X-ray beam that was presented in that manuscript (not just used the same method). If so, it should be clearly stated.

15.	It is not clear how helpful it is to include so many panels in Figures 4, 6 and 7. The reader also does not need to know that there were two different names for each sample (e.g. sample 101 is also sample # 3434). The lines in these figures also could be changed so they are not all solid for legibility. Different colours should be used for the different predictions in each column while keeping the colour of the 'original' the same in both. This will help the reader to quickly understand what is changing and what is not. The horizontal range of all the plots is the same: it would greatly help legibility if the figures were stacked together and only two horizontal axes were included at the bottom.

16.	Figure 5(b) could be included as an inset to Figure 5(a) so it is clearer that one is simply a zoomed-in view of the other.

17.	It was not clear to us how the predictions presented in the right column of Figure 7 were only defined over a window of 0.8 mm (according to the text) and yet are plotted over a range of about 1.5 mm. Perhaps something in the methodology is unclear.

18.	Throughout most the manuscript, the accuracies of different simulations are presented as a percentage, but on p21 (1st paragraph) they are presented as decimal numbers (e.g. 0.726 instead of 72.6%).

19.	Figure 8 claims to plot three curves, but only two are distinguishable due to the close agreement between the three. In any case this Figure does not provide significant additional information, and if the authors wish to highlight that the focused beam is similar in all three cases they can do so in the text.

Conclusions:

20.	The first sentence of the conclusion (p22) is poorly structured and difficult to follow.

21.	The Conclusions appear to be more concerned with plans for future work rather than summarising the results found in this work. In particular the authors should highlight how this work extends previous work - especially Saha et al., 2020 – in simulating the effect of partial coherence and also exploring the importance of using an orthonormal basis for the aberration coefficients.

Data availability:

22.	It is not clear whether the authors need to clarify when the GitHub repository was last accessed when providing the link.

There are many stylistic and typographic issues throughout the manuscript, a few are highlighted below:

23.		Parentheses are overused as a stylistic device throughout the manuscript
24.		The way the manuscript is divided into sections does not appear to have carefully chosen. For example, sections 4.0.X should surely be sections 4.X.
25.		On p1, no affiliation index is needed because all authors are from the same institution
26.		On p4, 1st paragraph: appex should be apex
27.		On p6, 2nd paragraph: "Error profile samples are by created" should be "Error profile samples are created by"
28.		On p6, 2nd paragraph: "tertian" should be "tertiary"
29.		On p12, 3rd paragraph: the sentence "However, a system using decomposition in non-orthogonal coefficients also well" does not make sense, it should surely be 'also WORKS well'.
30.		On p14, 2nd paragraph: "of a less good learning" could be "of poorer quality learning"
31.		On p21, 2nd paragraph: "guessed error profile for the sample 103 in 7" should be "estimated error profile for sample 103 in Fig. 7"
32.		The article which introduces the coherent mode decomposition method is cited twice (Sanchez del Rio, 2022a and Sanchez del Rio, 2022b)