Reply to Dr. Z. Zhang

First of all, we would like to thank the reviewer for reading our paper carefully and providing constructive comments. In the revised manuscript, we have tried to accommodate all the suggested changes. The modifications from the originally submitted version are highlighted in the revised manuscript. Please see our specific responses below.

Comments on "Retrieval of optical thickness and droplet effective radius of inhomogeneous clouds using deep learning" by Okamura et al.
This paper documents a retrieval algorithm based on the deep learning neural network (DNN) for retrieving the cloud optical thickness and cloud effective radius from spectral cloud reflectance observations. The DNN algorithm is trained by synthetic cloud fields from LES and simulated cloud reflectances using 3D radiative transfer models. It is shown that a great advantage of the DNN algorithm is its apparent immunity to the socalled 3-D radiative transfer effects. The "traditional" Look-up-table method suffers from significant biases due to the illuminating and shallowing effects, while the retrievals from the DNN algorithm are less affected by these biases and agree better with the "ground truth" from LES.
Overall, I found this paper interesting and exciting, and certainly suitable for AMT. On the other hand, I do have a few questions/suggestions that are listed below and hope they can help the author further improve the paper. For full disclosure, I know almost nothing about neural network or machine learning. So, my comments will be mostly from the perspective of cloud remote sensing which is my research field.

comments

1) Robustness of the results: I'm very excited to see that the DNN-based algorithm is able to overcome the influence of 3-D effects (illuminating and shadowing) and yield retrievals in close agreement on LES. My biggest concern is that if this result is robust enough. I hope I am not mistaken, but it seems the DNN in this study is trained using only two LES cases in Figure 1 and moreover only applied to these two cases. If so, frankly, I not completely convinced if the algorithm will generate same successful retrievals if it is applied to other LES scenes or real satellite images. To convince me and the readers, the authors should consider a "blind test", in which they apply the algorithm to the LES cases other than the two training cases. For example, the authors can tweak the meteorological conditions in the LES (e.g., inversion strength, large-scale forcing etc.) to generate different cloud types/scenes, and then apply the DNN algorithm to assess and report its performance. r Overall, the authors need to demonstrate the robustness of their algorithm and results.

Response: We understand this issue. Indeed, we limited our tests to two cases of boundary layer clouds and do not expect the DNN trained in this study perform well for cirrus cloud (for example) that is geometrically thick and optically thin. In general, NN is known to perform well for data that are similar to those used in the training. It is not very confident that the NN work well for data that are very different. This is a common issue in every NN/DNN-based cloud retrieval methods (Faure et al., 2011; 2012; Cornet et al., 2004; 2015; Evans et al., 2008; Kox et al., 2014; Minnis et al., 2014; Strandgen et al., 2017).

This study is just a feasibility study, but it is encouraging to see that the 3D radiative effects (e.g., illuminating and shadowing) are reasonably corrected in the results presented in this paper. The convolution filters shown in Fig. 8 "suggest" that the DNN indeed learned meaningful patterns of 3D radiative transfer although it is difficult to interpret how the filters correct the 3D effects. At least, in this study, we tried to expand the variety of training data by scaling the cloud optical thickness artificially. As a result, we could show the performance tests for a wide range of optical thickness from 0.1 to 100.

Additional simulations using a LES model is technically difficult for now because it is too expensive with time and computational cost in mind. More tests including different types of cloud should be done in the future works. In addition, when more cloud data are available, DNN should be retrained for additional cloud data. This will be important for practical applications in the future. We have added a few words in the last sentence in the conclusion as "...will require training using realistic cloud fields for various types of cloud."

2) Complexity of the training: Note that 3-D radiative effects depend on many factors, not only just COT, CER and solar geometry, but also cloud top inhomogeneity, cloud geometrical thickness and surface reflectance among others [Várnai and Davies, 1999; Várnai and Marshak, 2001; 2002] as well as instrument characteristics. I'm wondering which ones of these factors have to be part of the training and which ones do not need to be. Take surface reflectance for example. Can we train the algorithm using only one surface reflectance and then it will work for all other types of surface? In addition to 3-D effects, the retrievals are also affected by many other factors, the presence of drizzle, atmospheric absorption, surface reflectance etc. It is not clear from the paper to what extent these factors are considered in the DNN algorithm training, and which ones are not. Overall, I'm trying to figure out how "smart" the algorithm is. If we have to worry about all the above-mentioned details in the training, then the practical usefulness of the algorithm becomes questionable.

Response: In this study, we assumed surface reflectance of zero for simplicity to study a feasibility of DNN-based retrieval methods. When analyzing actual satellite data, it would be possible to approximately correct the surface reflection component in observed radiance to get cloud-only radiance, if we know the surface reflectance. On the other hand, cloud-top roughness, geometrical thickness and vertical inhomogeneity within cloud are included in the training as they are from the LES cloud data. This is because we wanted to test whether the DNN can retrieve cloud column properties (or correct the 3D effects) even with the complexities that likely appear in real clouds. As the reviewer pointed out, the 3D radiative effects depend on many factors. Basically, all such important factors should be included in the training of DNN. This is a next step for practical applications to actual observation data. If we do not think deeply, we may just add solar geometry, surface reflectance and more parameters to the input vector of DNN. Although such a simple addition should be easy with the help by the current techniques, an appropriate DNN architecture for addition of input parameters should be investigated. We have added the last sentence in the conclusion as "An appropriate DNN architecture for addition of input parameters should be investigated in the future."

3) Cloud mask: It is not clear from the paper how cloud masking is treated in the retrieval/training. If retrievals are done at the resolution coarser than the LES grid, then some pixels are inevitably partly cloudy. How are the partly cloudy pixels treated in the retrievals and training?

Response: Subpixel clouds are not considered in this study. Horizontal resolution is 280 m for both training and test. We have added a sentence in the first paragraph in Section 2.1: "The area averaging was done over a cloud region of 280 m for x- and y-axis; For simplicity, subpixel clouds are not considered in this study."

4) Definitions of CER: When cloud microphysics varies both vertically and horizontally, then the definition of CER can be very tricky. For example, Eq. (1) applies well to a single LES cell, no problem. (the root and meaning of the parameter need to be explained in detail though). The equation (2) for column-mean CER becomes tricky. First of all, does the vertical average takes into account any vertical weighting for example due to photon penetration depth [Platnick, 2000; Miller et al., 2016]? Some explanations are needed either way. Second, what the column-mean ? How to compute it? Third, what is the significance of the column-mean CER in Eq. (2)? Does it help understand the cloud radiative effects? Does it help the modelers validate their cloud microphysics simulations? Can it be used in combination of COT retrieval to estimate LWP? After defining the column-mean CER for a single column, the authors also need to explain how to aggregate/define the CER over multiple LES columns horizontally. For example, if the retrievals are done at 10x10 pixels, and each pixel has a slightly different column-mean CER, then what is the CER for the 10x10 pixel ensemble?

There are a few recent studies that discussed this topic. Maybe they are helpful [Miller et al., 2016] and [Alexandrov et al., 2012]

Response: We thank the reviewer for this comment and suggestions of references. There is no consensus on a representative CDER definition for cloud column, in the community. We needed to define some representative CDER to treat vertical inhomogeneity in the retrieval. First we made coarse resolution (280 m) data of LWC and N. Then Eq. (1) and (2) are applied to define local CDER and column mean CDER. As in Eq. (2), the column mean CDER (Re) is without vertical weighting. Although we think the definition is enough for our current purpose, this issue will be focused when sub-pixel cloud inhomogeneity is in mind. We have added a few sentences in Section 2.1:
"There is no community consensus on a single definition of CDER that is representative of the full column in the case of a vertically inhomogeneous cloud. Nevertheless, this study introduces the retrieval of such a representative CDER, …"
"It should be pointed out that there are other possibilities for column-average CDER (Miller et al., 2016). "

5) Plane-parallel albedo bias: This study focuses on the impacts caused by IPA, but there is another type of bias, plane-parallel-albedo bias (PPHB). It is not clear to me if the DNN described in this study could also take care of the PPHB. Note that recently, Zhang et al. [Zhang et al., 2016] described a novel method to correct the PPHB, which might be helpful for this study.

Response: In principle, the DNN should be able to handle PPHB as is shown by pioneering work using NN in Faure et al. (2001). We have added a citation to Zhang et al. (2016) in the Introduction, as follows:
"Zhang et al. (2016) recently described a novel method to correct the effect of in-pixel cloud inhomogeneity using subpixel reflectance variabilities."

6) Lack of technique details: I agree with the other reviewer that many important technique details are lacking from the current paper. Currently, the paper is rather short, so there is plenty of space to add in more detailed description and discussion, especially for Section 3 Method. Just to give an example, what are the meaning of Eq. (8) and (9)? Why do they provide the “relationships between inputs and outputs variables” of DDN, what kind of relationship?

Response: According to a comment from the referee 1, we have moved explanations on fundamental DNN techniques to Section 3.1, where fundamental deep learning techniques are summarized. We have added explanations in several parts. We hope revised manuscript is easier to read. However, we have not added very long explanations on the techniques that are really technical and not essential to the conclusions of this paper. We think it is better to leave the technical details of each optimization and deep learning technique for readers to consult the textbooks or references cited in the current manuscript. The deep learning techniques are rapidly growing in broad areas, upon many successes in engineering and applications for the artificial intelligence. Essential characteristics of each deep learning techniques are described in the current manuscript.

Although DNN can generally approximate nonlinear functions, it is expected that DNN may approximate the functions with a smaller number of DNN layers if the nonlinearity is less. The functions in Eq. (8) and (9) are less nonlinear to radiances. This kind of simple transform help better performance of DNN retrievals. We have modified the introductory sentences as follows:
"Although DNN can generally approximate nonlinear functions, it is expected that less nonlinear functions can be approximated by fewer DNN layers. Constructing efficient DNN thus makes it desirable to linearize the relationship between input and output variables to some degree. Because the radiances are highly nonlinear with respect to the COT and CDER, it is convenient to transform the COT and CDER by some simple functions."