Responses to reviews

Reply to anonymous referee 1

We thank the reviewer for the time and efforts she/he spent reading our manuscript carefully and providing valuable comments and suggestions. In the revised manuscript, we have tried to accommodate all the suggested changes. Please find below our responses. The reviewer's comments are in italic. Changes/additions made to the text are given in quotes.

This article presents the possibility to use of deep neural network (DNN) to retrieve cloud properties (optical thickness and effective radius) accounting for the horizontal photon transport. This is so far not accounted for in the classical algorithm, that use the homogeneous cloud assumption. This is a move in the way to improve remote sensing algorithm and account for 3D radiative effects. However, the presentation and explanations do not put the paper in favor. Consequently, several precisions and corrections need to be added in the paper before publications. They are indicated below.

Major comments

1) The originality of the paper seems to be more related to the possibility to make a multi-pixel inversion of cloud properties than to use DNN. It should appear in the title. I suggest "Feasibility study of multi-pixels retrieval of cloud optical thickness and effective radius using deep neural network"

Response: We appreciate the suggestion. Indeed, this is a feasibility study of such a multi-pixel inversion. It is also important that the deep learning techniques are applied to retrieval of inhomogeneous clouds. Thus, by combining the reviewer's suggestion and our points, we have changed the title as
"Feasibility study of multipixel retrieval of optical thickness and droplet effective radius of inhomogeneous clouds using deep learning"

2) Abstract is too succinct and need to be completed.

Response: We have rewritten the abstract, adding explanations about the two DNNs

3) In the introduction, the authors described similar works, they did previously to retrieve COT and CEDR accounting for neighboring pixels (Iwabuchi et Hayasaka, 2003). Through the paper, the disadvantages of this previous method comparing to the new

one are not sufficiently explained. I did not understand "which was an obstacle to generalizing the algorithm (p2, li10)". Which obstacles? Does it not the same problem with the NN method ? The authors should add a discussion about the advantages/disadvantages and about the implementation of each method in the introduction or in the conclusion A comparison of previous method and DNN method in terms of results will also be valuable.

Response: We have rewritten explanations about limitations in the previous study by Iwabuchi and Hayasaka (2003), as follows:
"Since 3D radiative effects vary with COT, CDER, cloud geometrical thickness, cloud top roughness, and sun–cloud–satellite geometry, Iwabuchi and Hayasaka (2003) had to construct different sets of fitting coefficients, thus limiting the applicability of the technique in practice. In addition, their method was based on linear regression, which is not flexible to account for any nonlinear 3D radiative transfer effects."

A comparison with the previous method in Iwabuchi et Hayasaka (2003) is technically difficult. One reason is that the previous method was developed for a feasibility study assuming a particular type of cloud model (Fourier transform cloud), which is quite different from that in the present study. In addition, the assumed cloud inhomogeneity was in one horizontal dimension while the present study treats two horizontal dimensions.

We have added an explanation of one of advantage of DNN, ease of addition of input and output variables, as follows:
" In addition, input and output parameters can easily be added, and structures can be modified - another advantage of deep learning"

4) li 15-22: Some important references are incorrectly cited: Faure et al. (2001) is about the retrieval of mean cloud properties accounting for the sub-pixel heterogeneities while Faure et al. (2002) concerns the retrieval of cloud parameters from high-resolution data using adjacent pixels which is a different study. This second one is the closest to the current study. Correct it also in section 4.2. These two papers are major because they are the first papers in the fields but they are limited to fluxes and not applicable to real data. Following the paper of Faure et al., 2001, which is for medium spatial resolution, Cornet et al. (2004) present ways to apply to real data heterogeneous cloud retrieval using NN. It is finally tested on real data in Cornet et al. (2005) on MODIS data. The paragraph citing these studies about cloud neural network retrieval needs to be clarified.

Response: As the reviewer requested, we revised Section 1 and clarified the description about

previous studies. The main discussion in Faure et al. (2001) was the application of NN to retrieval of mean cloud properties, taking sub-pixel inhomogeneity into account, but they also tested a NN that uses the radiances of target pixel (0.8km x 0.8km region) and eight adjacent pixels (0.8km x 0.8km regions) for retrievals, showing a promising result. Our DNN method is compared to the latter one in the Faure et al.'s (2001) study. We have added descriptions about this point in the Section 4.2, as follows:

" Originally, this NN had two hidden layers with 10 units each and it used the 0.8 km $\times$ 0.8 km area-averaged radiances at four wavelengths (0.64, 1.6, 2.2, 3.7 μm) at the target pixel and eight adjacent pixels (called as "with ancillary data"). It is described in the section 3.3 (2) in the original paper."

Faure et al. (2002) uses high resolution but one-dimensional cloud data, and we cannot directly compare our results with that of Faure et al. (2002). We have added the reference Cornet et al. (2005) in Section 1.

5) To my knowledge, this is the first time in atmospheric science that Deep Neural Network (DNNs) are used. More explanations are needed in a specific section explaining clearly how it works and allowing to understand some affirmations and vocabulary used in the text. For example, in the introduction, why "a DNN is more suitable for approximating complex non linear functions" than a classical NN? What is "automatic feature extraction", can the authors give an example? For the same reasons, Section 3.1 are confuse and consequently not very clear for a non-expert in deep learning. It needs to be separate with generality on the DNN in the specific section rewritten with more explanation and in a pedagogical way. Some schemas may also help to understand. Another section should specify to the choices made (see major comment 6 below). In the specific section about DNN should appears what is "shortcuts DNN"(p5, li 17) or what is convolutional layer ? How the filter weights are obtained? Can the authors also explain in few lines the paper of He et al. (2015) in order that the readers understand?

Response: We appreciate this comment. We have moved explanations on fundamental DNN techniques to Section 3.1. As the reviewer commented, some important explanations were missing in the previous manuscript. So we have added some more explanations (ex. "a DNN is more suitable for approximating complex problem", "automatic feature extraction", and "shortcuts"). We hope the revised manuscript is easier to read.

We have not added very long explanations on the techniques that are really technical and not essential to the conclusions of this paper. We did discuss essential characteristics of each deep learning technique in the current manuscript. 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. Indeed, the deep learning is applied to a remote

sensing problem of atmospheric target for the first time, to our best knowledge. However, it is a rapidly growing technique in broad areas of sciences, upon many successes in engineering and applications for the artificial intelligence. In the atmospheric sciences, too, we know at least a few research groups working on applications of deep learning techniques. Nowadays, it is easy to find a book for practical use of this technology even for undergraduate students.

6) p5 . There is also no enough explanation about the choice of the input vector and the architecture of the DNN. Li- 5-10: why these two input vectors? The paragraph should start with an explanation of the philosophy. The first input vector is built in order to correct IPA retrieval and the second to retrieve directly cloud properties. I'm wondering also why four wavelengths and not only two as for the bi-spectral method? Does the authors test this last configuration with two wavelengths? I'm wondering also how the architecture of the DNN was chosen (convolutional layer or not, activation function or not), does the authors made test to find the best architecture?

Response: As the reviewer commented, we tested numbers of different DNNs with different number of wavelengths, the use of convolutional layer, different activation function, and so on. The best two DNNs are shown in this paper. We have added the description about this in the manuscript, as follows:
"The above two DNN structures were obtained from various trial-and-error experiments. Different DNN structures were also tested. For example, we tested a DNN similar to DNN-2r but with four wavelengths, and one similar to DNN-4w but with only two wavelengths. However, DNN-2r and DNN-4w performed best. There is room for improvement in DNN structures, which should be investigated in the future."

7) P7- li 28-30: I am not completely agree with the assertion "the DNN retrieves COT values that are close to the true values assumed in the test, successfully corrected the phase lag." In Figure 6-a, near 11.7km, DNN-2r retrieval shows also large differences and near 17km clearly the DNN retrievals overestimate the COT and the phase lag is not completely cancelled. Can the authors be more precise in the description of the figures? In addition to cross-sections, could also the authors add the relative errors transects and the RMSE of the different retrievals to have more qualitative idea of the improvements. Same remarks concerning Re retrieval.

Response: We have rewritten the sentence as
"Compared to the IPA retrieval, the DNNs retrieve COT values that are closer to the true values assumed in this test. The DNN-4w successfully corrects the phase lag as shown in Fig. 6c."
As the reviewer suggested, we have added figures of cross-sections of the relative errors and a table of RMSEs of the different retrievals in Figure 6.

8) p8, li 1-6: Concerning Re retrieval with the homogenous cloud assumption, it is not really surprising to obtain large differences between homogeneous assumption and true results. The overestimation with IPA is not only related with shadowing effects. Indeed, homogeneous cloud assumption involve homogeneous Re profile. From satellite remote sensing, the upper part of the cloud is retrieved (See for example Platnick et al. 2000). Therefore, if the effective radius is vertically increasing in the cloud, the retrieved Re is larger than the mean Re. For the DNN, training with heterogeneous clouds allows to learn the relation between vertically averaged Re and radiances. Discussion about this issue (p8, li 20-24) is too late in the paper and should be moved here. Try also to highlight better the shadowing by reporting for example the difference between true COT and homogeneous COT as in Cornet et al. (2015) or Marshak et al., (2006).

Response: Figures 5 and 6 are good for demonstrating spatial characteristics of retrieval errors, but the mean bias error for all test data are shown in Fig. 7. Differences between vertically homogeneous and inhomogeneous assumptions are well related to the mean bias in Re. We think it is better to see the discussion about the issue as in the current manuscript. We have modified the sentences as follows:
"IPA-retrieved CDER thus tends to be larger than column-mean CDER, whereas DNNs are by design trained to learn the relationship between the column-mean CDER and radiances by taking into account the vertical inhomogeneity. However, this vertical inhomogeneity effect on the IPA-retrieved CDER does not seem to be the main cause of the large positive bias."

The issue of difference between true COT and homogeneous COT as in Cornet et al. (2015) should be important, but the sub-pixel horizontal inhomogeneity is not considered in this paper.

9) P9, section 4.2: Authors made comparisons with previous works of Faure et al., 2002 but the settings are exactly the same. First, only pairs of wavelengths were used and not the four wavelengths mentioned. In addition, in the study of Faure et al., (2002), 15 neighboring pixels of each side of the target pixels were used (62 components in the input vector) and here only 3. This can change a lot the results. The comparisons have to be done again with the same parameters than the one used in Faure et al. (2002) or at least the same conditions that the DNN, that is 10 pixels for each side, otherwise, it is not possible to conclude the comparisons and to know really why retrieval is better (DNN or neighboring pixels?)

Response: As explained in the reply to the comment 4, we compared our results with that of Faure et al. (2001).

Minor comments

1) p. 1, li 17: why the bispectral method follows the IPA assumption? The authors should add reasons why in the text (time computation, simplicity, others?) and also insist about the independence of each cloudy columns which is considered infinite.

Response: We have rewritten the sentences as follows:
" The method is based on the independent pixel approximation (IPA) assuming plane- parallel, horizontally and vertically homogeneous cloud for each pixel of the satellite image because of high computational cost for simulation of three-dimensional (3D) radiative transfer. The observed cloud radiances result from three-dimensional (3D) radiative transfer in the cloud field, …"

2) p.2, li 5: Until which distance, have the neighboring pixels to be considered ? can the authors here or further in the text give some values and references ?

Response: We have added an explanation about the scales, as follows:
"The 3D radiative effects operate on horizontal scales that are determined mainly by cloud thickness and solar zenith angle. When the sun is oblique (i.e., with a solar zenith angle of 60◦ or larger), the maximum horizontal scale for 3D radiative effects is roughly 15–20 times larger than cloud thickness (Marshak and Davis, 2005).

3) p3, section 2.1: what are the resolution and dimension of the generated cloud fields?

Response: We added a description about the resolution of the generated cloud fields as "The resolution for the x- and y-axis of the cloud field is originally 35 m. For the z-axis, the resolution is 5 m at the bottom of the atmosphere, and it is coarse (less than 60 m) for the upper layers."
The dimension sizes are written in the manuscript.

4) p3, li 17: IPA (Independent Pixel Approximation) does not mean that the vertical profiles is homogeneous but only that each pixel is considered independently of his neighbors. Authors should speak about the homogeneous cloud assumption horizontally as well as vertically

Response: The IPA retrieval usually assumes vertically and horizontally homogeneous cloud in each pixel. We modified the introduction part as "the independent pixel approximation (IPA) assuming plane-parallel, horizontally and vertically homogeneous cloud for each pixel of the

satellite image".

5) p3, eq. 3: Why the authors used the square of the usual definition of the inhomogeneity parameter defined in the others study. For comparisons, it seems to be better to use the same definition.

Response: This is because the 3D radiative effects (e.g. 3D minus IPA radiances) are approximately linear to the square of the usual definition of the inhomogeneity. In this paper, we tried to show that the open cell cloud was much more inhomogeneous.

6) P5, li 9: Radiances at 3.75micron is used, I suppose that is only the solar part. It should be precise in the text that thermal correction need to be done before using this wavelength.

Response: We have added an explanation about this, as follows:
"It should be noted that only solar radiation is considered in the present study, which requires that the thermal radiation at 3.75 μm be corrected during pre-processing."

7) p5: explain why the number of pixels considered in the input vector (10x10) is larger than those considered in the output vector (8x8 or 6x6) and why it is not the same for the two DNN. 8) P6, li10: add the URL for the chainer framework

Response: We have added a sentence for the reason as follows:
"The reason for including margin pixels in the input field is to take into account the 3D radiative effects from the surroundings of the cloud field. The DNN-2r network consists of several fully connected layers."

9) P6, eq 8 and9: is there a justification for the choice of these functions?

Response: The functions in these equations are less nonlinear to radiances. This kind of simple transform helps better performance of DNN retrievals. We have added an introductory sentence as follows:
"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. "

10) P7, li 14: DNN and IPA are not really comparable: the first is an inversion tools as look-up tables and the second one is a direct model. It seems better to write multipixelsDNN inversion versus IPA-LUT inversion.

Response: We agree, but since we described our terms, we prefer to stick with the terms of "DNN and IPA retrievals". We modified the manuscript to use "IPA retrieval" in a consistent way.

11) Figure 5 and 6: Precise data corresponds to only the test data set or to a mix between the training and test dataset

Response: These are for test data.

12) p7, li 22 and Figure 5 and 6: precise the geometry of the observation: view zenithal and azimuthal angles?

Response: We added the geometry information to Figure 5 and 6, and in the manuscript.

13) p7: li 26: illumination and shadowed effects are well-known under IPA assumption: please add some references

Response: Várnai and Davies (1999), Várnai and Marshak (2002), and Marshak et al. (2006) are added as references about illuminating and shadowing effects.

14) p7: li 26: Large errors are due to the flattening of the relation-ship between radiances and COT due to saturation effects: a small difference in radiance lead to a quite large difference in COT

Response: We appreciate this comment and have added a sentence as
"This can be ascribed to weaker sensitivity of radiance to COT when COT is larger; a small difference in radiance leads to large difference in retrieved COT."

15) p8, li 9, figure 7 : I agree that the bias (mean error) is particularly large only for COT less than 1. For COT > 1, the difference in errors is not so important. For COT>10, the standard deviation is larger for IPA meaning that dispersion (roughening) is more important.

Response: Large IPA error dispersion for COT > 10 are shown for SZA of 60 degrees, which suggests that the roughening cause the error. This point is mentioned in the manuscript.

16) Figure 8: Could the authors indicate the COT and Re associated with the filters and be more precise in the description of the figure? Which filters patterns are "symmetrical around the center" and how is distributed the optical thickness? Also on which

figures does appear "the feature related to the solar direct beam"? Comments also the difference between wavelengths.

Response: In general, this kind of analysis of NN coefficients is difficult to understand when the NN becomes deeper. To study the associations of these filters with COT and CDER, we should investigate the coefficients in the DNN-4w layers subsequent to the convolutional layers (as shown in Fig. 4). It would be hard to interpret the relationships between the two destination fully connected layers in the DNN-4w. In this paper, we just tried to check that the DNN-4w indeed learned meaningful patterns of 3D radiative transfer.

As suggested by the reviewer, we have added a few explanations about specific pattern shown in Fig. 8. We have also added an introductory note as follows:
"It is noted that the convolutional layer is designed to correct the 3D radiative effects appeared in radiances (Fig. 4)."

17) Section 4.2: are the same training set and generalization set used for all the NN trainings?

Response: Yes, the original dataset are the same as those used for DNNs. We have added a sentence as "Data used for training and test for the NN are from the same original datasets used for DNNs."

18) p9, li 30-33: add the issue concerning the vertical profiles for Re in the conclusion.

Response: As shown in Section 4.1, the vertical profile for Re seems to be not a main reason for the IPA retrieval bias in Re, at least, in the present cases. We think there is no strong need to mention this point in the conclusion.

19) P10: in the conclusion, can the authors insist on the limitations of using NN methods such as the one related to database used and extrapolation issues. In other words, how will work the DNN is the cloud is quite different to those used for the training dataset?

Response: It is a well-known issue of NN as the reviewer suggested. Training of NN should includes enough variety of realistic cloud. This study is just a feasibility study, but practical applications definitely require training for various types of cloud. We have added a few words as "...will require training using realistic cloud fields for various types of cloud."

20) P10: Following the previous points, can the authors speaks about the steps needed

in order to develop an operational multi-pixels algorithm?

Response: An expansion of cloud variety is one step. An appropriate DNN architecture for addition of input parameters should be investigated. For example, the sun-cloud-satellite geometry parameters are very different from the radiance image data that are used as input data in this study. There should be appropriate DNN structure to add the geometrical parameters. 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."

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."

Reply to anonymous referee 2

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.

The paper describes a new technique for satellite measurements of cloud optical thick- ness and cloud droplet effective radius. The key feature of the technique is that it takes into account 3D radiative effects and subpixel variability by considering not one pixel at a time, but by performing simultaneous retrievals over 10 by 10 pixel areas. The most important aspect of the technique is the use of a deep learning algorithm. This is a significant new development, and the study makes an important contribution on the path toward more accurate satellite retrievals of cloud properties. Overall, the method- ology is sound and the presentation is suitable. However, I believe that a few important improvements are needed in the analysis. My recommendation is therefore to make some major revisions. Please find below my detailed comments.

comments

Major issues:
1.
Page 7, Line 8 mentions that "The test dataset used for evaluation should be indepen- dent of the training dataset." My sense is that in this initial study the training and testing datasets are not fully independent, as they come from the very same cloud fields, and that this would be good to mention. (The two datasets include different randomly se- lected locations within the cloud fields, but the statistics of cloud properties are identical in the training and testing datasets.)
As noted in Page 10, Lines 7-8, it will be an important future step to examine the per- formance of the retrieval for a wider range of cloud parameters. It is reasonable to leave this (and the evaluation based on fully independent training and testing datasets) to a future paper, but even the current results could offer further insights into the ro- bustness of the proposed retrieval algorithms. Most importantly, one could examine not only the overall results, but also separately the results for open-cell and closed-cell convection cases. This would demonstrate that the same algorithm and training set improves retrieval accuracy for two very different types of cloud structures. I don't think the currently presented results show this: Overall error statistics may conceivably im- prove due to improvements for open-cell convection only, without any improvements for closed-cell convection. (Because retrieval uncertainties are likely larger for open-cell convection, it may be best to examine by what percentage DNN-2r and DNN-4w reduce the retrieval errors of IPA retrievals for open-cell and for closed-cell convection.) The paper did a good job in examining results as a function of optical thickness, but the new

analysis of already performed retrievals would help because open-cell and closed-cell convection cases differ in horizontal structure even at locations where vertical optical thicknesses are similar.

Response: We have added Table 2 to show RMS errors (in %) for open and closed cell cases, separately. The results show that DNNs generally retrieve more accurate COT and CDER than IPA method. For SZA of 60 deg., DNNs are better than IPA in both open and closed cloud cases, while there is an exception; IPA is better for COT for SZA of 20 deg. in the closed cell case. We have added explanations about this table, as follows:
"Table 2 shows the relative RMSE of estimated COT and CDER by the IPA and DNN retrievals for open-cell and closed-cell cases. In both cases, the retrieval accuracies for DNNs are obviously improved compared to the IPA retrieval. An exception appears for COT in closed-cell case when SZA is 20◦; COT RMSE of 24% for DNN is larger than 16% for IPA. The DNN-4w is better than DNN-2r in both cases."

As the reviewer suggested, we have revised the explanation about test datasets as "the training and test dataset include different randomly selected locations within the cloud fields, but the statistics of cloud properties are identical in the training and test datasets."

2.

Page 5, Lines 13 and 22: I wonder why scene parameters are estimated for 8 X 8 pixel arrays when using the DNN-2r method, but only for the central 6 X 6 pixel arrays when using the DNN-4w method. This could make sense if 3D effects acted over larger distances at 3.75 microns than at the 0.86 and 2.15 microns used by the DNN- 2r method, but neither my own physical reasoning nor the filter weights in Figure 8 suggest this. In fact, Figure 8 shows that DNN-4w retrievals at a pixel are strongly affected by 0.86 micron radiances 2 pixels away. This suggests that (at least for pixels at the edges of 8 X 8 pixel areas) the DNN-2r method cannot capture the portion of 3D effects caused by areas more than a pixel away. This probably contributes to DNN- 2r giving less accurate results than DNN-4w (a tendency mentioned in Page 9, Lines 31-32) and should be mentioned in the discussion of the differences between the two methods at the top of Page 10. (The discussion should also include the impact of additional wavelengths in DNN-4w and algorithmic differences.) Also, it could help to clarify explicitly in the paper whether DNN-2r retrievals inside (not along the edges of) 8 X 8 pixel areas are affected by radiances 2 pixels or more away. If they were, it could even make sense to analyze retrieval accuracy only for pixels in the central 6 X 6 pixels of 10 X 10 pixel areas (similarly to DNN-4w).

Response: We appreciate this comment. Honestly speaking, we had no strong reason to limit the output pixels in DNN-2r and set it as 8x8 pixels. On the other hand, because of convolution filters of 5x5 pixel size, the number of pixels were limited as 6x6 pixels. As the reviewer

pointed out, the difference of the number of pixels in the DNN output layer may have an effect on the retrieval performances. However, we think that it is not a main reason because 1) influences from neighboring pixels generally tend to weaken with increasing distance from the target pixel, 2) there are only 28 edge pixels in the 64 (8x8) pixels area, and 3) most of these edge pixels just miss only one side of neighboring pixels 2 pixels away. The 3D radiative effects actually operate on larger horizontal scales than 2 pixels (560 m) as captured by DNN-4w, while the effects tend to weaken for larger scales. As shown in Fig. 8, the filter coefficients do not vanish at the edges of the window. To improve retrieval accuracy, it is better to increase the number of adjacent pixels used for both DNN-2r and DNN-4w, as this point is mentioned in the conclusion.

In the initial stage of this study, we tested various kinds of DNN architectures with a different number of wavelengths, the use of convolutional layer, different activation function, and so on. For example, we tested one similar to DNN-2r but using four wavelengths or one similar to DNN-4w but using only two wavelengths. However, the DNN-2r and DNN-4w were the best performed ones. Therefore, at least we can say that a main reason for the better retrieval performance for DNN-4w than DNN-2r would be a combined effect of use of additional wavelengths and a different DNN architecture (e.g., use of convolutional layer and release from the reliance on the IPA retrieval). However, unfortunately, we could not make the reason very clear. We have added the description regarding the future improvements in the manuscript, as follows:
"The above two DNN structures were obtained from various trial-and-error experiments. Different DNN structures were also tested. For example, we tested a DNN similar to DNN-2r but with four wavelengths, and one similar to DNN-4w but with only two wavelengths. However, DNN-2r and DNN-4w performed best. There is room for improvement in DNN structures, which should be investigated in the future."

Minor issues:

Page 1, Line 23: What is meant by "cloud state"?

Response: In this context, "cloud state" means cloud vertical and horizontal inhomogeneity.
We replaced these words in the new manuscript as "cloud horizontal and vertical inhomogeneity".

Page 2, Line 23: The study by Evans et al. ("The Potential for Improved Boundary Layer Cloud Optical Depth Retrievals from the Multiple Directions of MISR", J. Atmos. Sci., 2008) should also be mentioned, as it also used a neural net for cloud retrievals.

Response: We have added Evans et al. (2008) as a reference.

Page 2, Lines 26-27: What is meant by "feature" and "feature extraction"?

Response: In this context, "feature" means a feature in the training datasets. This word is usually used in the classification problem (e.g., object recognition from input image) using DNN. We have modified the manuscript as follows:
"features in training datasets are learned hierarchically in the DNNs, although it is not very easy know how the features are described in the DNNs"

Figure 2: It would help to indicate the time elapsed during the 60 time steps along the horizontal axis, or to mention the time step in the figure caption.

Response: We have added an explanation in the Fig. 2 caption as
"A time step corresponds to one minute."

Figure 3: It would help to clarify why there is a fully connected layer near the top of the left column that operates only on radiances and not on the IPA-estimated scene parameters.

Response: We thought the layer should be in, at least, an either side (left or right side) of the figure. A layer can be inserted in both sides, but we did not have very strong reason. We believe this is a minor issue, but we have added a sentence as follows:
"In the first layers, radiances and IPA-estimated cloud properties are merged to obtain 8x8x2 elements (two elements per pixel for 8x8 pixels)."

Page 6, Lines 27-29: It would help to clarify whether all pixels within an LES scene are multiplied by the same randomly chosen number, or all individual pixels are multiplied by a different number. (My guess is the first option.)

Response: As the reviewer guessed, all pixels within a single scene are multiplied by the same number. We have added a sentence:
"The cloud extinction coefficients of all pixels within a single cloud scene are multiplied by the same number."

Page 6, Line 25 and Page 7, Lines 8-10: What does the word "samples" refer to? My guess is that each sample is a 10 X 10 pixel area. If my guess is right, can samples overlap? Also, it would help to mention the total number of pixels in the LES dataset, as this could show whether the training set includes almost all LES pixels or just a small fraction of them.

Response: As the reviewer guessed, a "sample" here means a 10 x 10 pixel area, and sample areas can be overlapped by other samples. The horizontal size of a LES scene is 28km x 28km. The total number of pixels in one LES scene is 100 x 100 = 10000 pixels. We have added some descriptions as mentioned above.

Figure 7: It could help to include into one of the panels a PDF of true optical thickness values.

Response: Figure 1 shows means and standard deviations of COT for the two cloud cases, from which rough shapes of COT PDF may be imagined. Although the distribution functions of true COT are not presented in this paper, the following figures show the joint histograms of the DNN-4w retrieval error and COT.

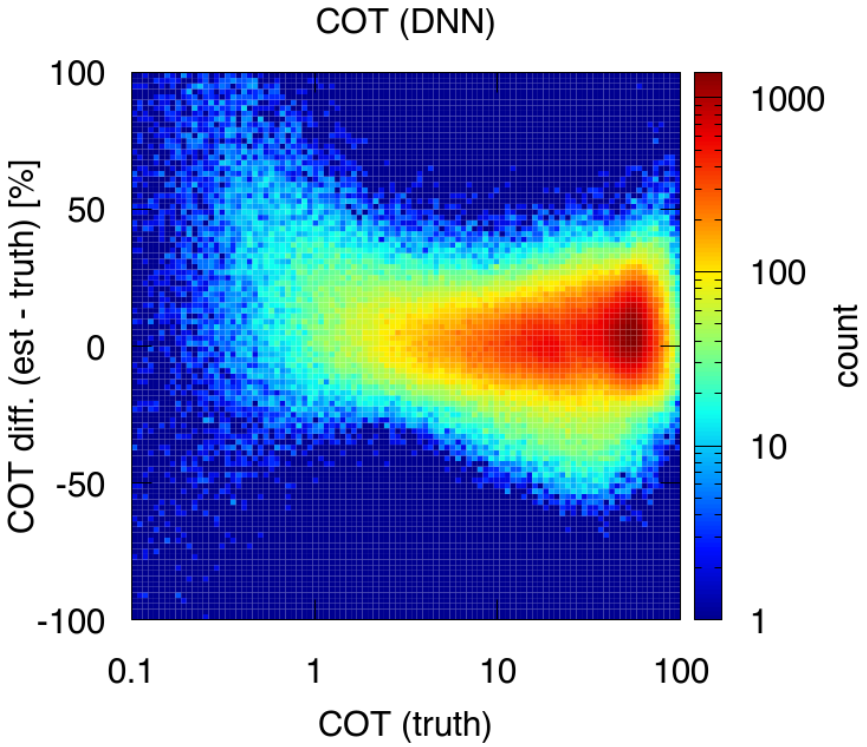

Figure 9: The legend should indicate which color shading corresponds to which line/method.

Response: We have revised Figure 9 as suggested.

Page 10, Lines 5-6: I am not sure the sentence "In the DNN-4w that we tested, we excluded 3D radiative transfer effects that occurred at horizontal scales greater than approximately 1.5 km (5 pixels)" is correct. Based on Figure 8, I thought that DNN-4w retrievals exclude 3D effects that occur at horizontal scales greater than 2 pixels (560 m). This is because I thought the pixel

whose properties we are retrieving is at the center of the filters in Figure 8, which means that only radiances two pixels away are considered. A correction of this sentence or a clarification of the meaning of filters in Figure 8 would help.

Response: The reviewer is right. We have fixed the sentence as
"...scales greater than 560 m (2 pixels)"

Somewhere in the text it would help to comment on whether the speed of calculations would be a concern for using DNN in operational retrievals in the near future. (For example, how does the speed of DNN compare to the speed of IPA and NN retrievals?)

Response: We did not exactly compare the computational costs of DNN, IPA, and NN retrieval. We have added an explanation about rather general facts about the DNN computational cost, in the last part of Section 3.3, as follows:
"As for computational cost, the training requires significant computation time, for which even one GPU helps considerably. Once the DNNs are trained, the retrievals using the present DNNs are generally very quick because they entail only very few simple manipulations of numerical data."

~~Retrieval~~ Feasibility study of multipixel retrieval of optical thickness and droplet effective radius of inhomogeneous clouds using deep learning

Rintaro Okamura[1], Hironobu Iwabuchi[1], and K. Sebastian Schmidt[2]

[1]Center for Atmospheric and Oceanic Studies, Graduate School of Science, Tohoku University, 6-3 Aoba Aramaki-aza, Sendai, Miyagi 980-8578, Japan
[2]Department of Atmospheric and Oceanic Sciences, University of Colorado, Boulder, CO, USA

Correspondence to: Hironobu Iwabuchi (hiroiwa@m.tohoku.ac.jp)

Abstract. Three-dimensional (3D) radiative transfer effects are a major source of retrieval errors in satellite-based optical remote sensing of clouds. The challenge is that 3D effects manifest themselves across multiple satellite pixels, which traditional single-pixel approaches cannot capture. In this study, we present two ~~retrieval methods~~ multi-pixel retrieval approaches based on deep learning~~. We~~ , a technique that is becoming increasingly successful for complex problems in engineering and other areas. Specifically, we use deep neural networks (DNNs) to ~~retrieve~~ obtain multipixel estimates of cloud optical thickness and column-mean cloud droplet effective radius ~~simultaneously~~ from multispectral, ~~multipixel radiances. Cloud field data are obtained~~ multi-pixel radiances. The first DNN method corrects traditional bi-spectral retrievals based on the plane-parallel homogeneous cloud assumption using the reflectances at the same two wavelengths. The other DNN method uses so-called convolutional layers and retrieves cloud properties directly from the reflectances at four wavelengths. The DNN methods are trained and tested on cloud fields from large-eddy simulations ~~, and~~ used as input to a 3D radiative transfer model ~~is employed~~ to simulate upward radiances~~from clouds. The cloud and radiance data are used to train and test the DNNs. The proposed~~ . The second DNN-based retrieval, sidestepping the bi-spectral retrieval step through convolutional layers, is shown to be more accurate~~than the existing look-up table approach that assumes plane-parallel, homogeneous clouds. By using convolutional layers, the DNN method estimates cloud properties robustly, even for optically thick clouds, and can correct the~~ . It reduces 3D radiative transfer effects that would otherwise affect the radiance values and estimates cloud properties robustly even for optically thick clouds.

1 Introduction

Clouds play an important role in determining the radiation budget of the Earth. To understand how, it is necessary to know ~~about~~ the global distribution of cloud properties such as optical thickness (COT) and cloud droplet effective radius (CDER). These particular cloud properties are retrieved globally by optical remote sensing from various satellites. A standard method for COT and CDER retrieval is the bi-spectral method that is used to produce the Moderate Resolution Imaging Spectroradiometer (MODIS) cloud product (Nakajima and King, 1990; Platnick et al., 2003). This method uses solar reflection measurements at

two wavelengths, one with and the other without absorption by water droplets. The nonabsorbing wavelength is selected in the visible or near-infrared part of the spectrum, whereas the absorbing one is in the shortwave infrared (SWIR) part, typically around 1.6, 2.1, or 3.7 μm. The method is based on the independent pixel approximation (IPA) assuming plane-parallel, horizontally and vertically homogeneous cloud for each pixel of the satellite image ~~, whereas the~~ because of high computational cost for simulation of three-dimensional (3D) radiative transfer. The observed cloud radiances result from three-dimensional (3D) radiative transfer in the cloud field~~. The radiances are influenced by horizontal and vertical inhomogeneities within clouds, as well as to the~~ , which means that the radiances are not only affected by the vertical cloud structure, but also by net horizontal radiative transport ~~that occurs in an inhomogeneous cloud field~~in inhomogeneous cloud fields. Previous studies have pointed out that cloud inhomogeneities and 3D radiative effects produce large errors in the retrieved cloud properties (Iwabuchi and Hayasaka, 2002, 2003; Zhang and Platnick, 2011; Zhang et al., 2012). Studies using observational data have confirmed the dependency of such retrieval errors on both the cloud ~~state and~~ horizontal and vertical inhomogeneity and the sun–cloud–satellite viewing geometry (Liang et al., 2009; Liang and Girolamo, 2013; Grosvenor and Wood, 2014).

Satellite image data with relatively coarse resolution does not contain sufficient information about in-pixel inhomogeneity. Although statistical bias correction is possible (Iwabuchi and Hayasaka, 2002), it is still difficult to perform error correction on each pixel, especially if unresolved in-pixel inhomogeneity is the major source of error. Zhang et al. (2016) recently described a novel method to correct the effect of in-pixel cloud inhomogeneity using subpixel reflectance variability. For finer-resolution imagery, by contrast, retrieval errors from inter-pixel horizontal radiative transport become more important. The radiance observed at each pixel is determined by the spatial arrangement of cloud water in the ~~target pixel~~ pixel of interest (target pixel) and its neighbors. The 3D radiative effects operate on horizontal scales that are determined mainly by cloud thickness and solar zenith angle. When the sun is oblique (i.e., with a solar zenith angle of 60° or larger), the maximum horizontal scale for 3D radiative effects is roughly 15–20 times larger than cloud thickness (Marshak and Davis, 2005). This necessitates consideration of the adjacent cloud effects when estimating the cloud properties at the target pixel. Iwabuchi and Hayasaka (2003) attempted to correct the horizontal transport effect by using multispectral, multipixel radiances ~~to retrieve~~ for retrieving COT and CDER. They fitted a polynomial function of the multispectral radiances at the target and adjacent pixels to the IPA radiances at the target pixel. ~~Because~~ Since 3D radiative effects ~~differ for COTand CDER,~~ vary with COT, CDER, cloud geometrical thickness, cloud top roughness, and sun–cloud–satellite geometry, Iwabuchi and Hayasaka (2003) had to construct different sets of ~~numerous fitting coefficientsfor COT and CDER, which was an obstacle to generalizing their algorithm~~fitting coefficients, thus limiting the applicability of the technique in practice. In addition, their method was based on linear regression, which is not flexible to account for any nonlinear 3D radiative transfer effects.

To consider adjacency effects in a generalized manner, neural networks (NNs) (also known as multilayer perceptrons) are useful, ~~as they have been used in~~ and they have thus been applied to cloud detection and retrieval. Minnis et al. (2016) used an NN recently to estimate the COT of ice clouds from MODIS multispectral infrared radiances. ~~Using an NN is considered a better way to achieve high accuracy when accounting for~~ An NN is expected to improve cloud retrieval accuracy in presence of 3D radiative effects ~~in the retrieval of cloud properties, because doing so creates a more complex problem. Some~~ because of the complexity of the problem. Therefore, a few studies have already proposed such applications to the problem of 3D clouds.

~~Faure et al. (2001, 2002) showed~~ Faure et al. (2001) demonstrated the feasibility of ~~using~~ NNs to retrieve ~~cloud properties (i.e.,~~ mean optical thickness, mean effective radius, fractional cloud cover, and subpixel-scale cloud inhomogeneity ~~) from multispectral and multipixel~~ from multispectral radiance data at wavelengths of 0.64, 1.6, 2.2, and 3.7 μm ~~and a horizontal~~ for a pixel resolution of 0.8 km $\times$ 0.8 km. ~~Their results show that NN retrieval can be improved by using the radiances of adjacent pixels . Cornet et al. (2004) also showed the feasibility of using~~ Faure et al. (2002) improved NN cloud property retrievals of 1D inhomogeneous clouds by considering multi-spectral radiance (at 0.64, 1.6, 2.2, and 3.7 μm) from a collection of pixels adjacent to the pixel of interest. Cornet et al. (2004) used NNs to retrieve cloud properties (i.e., mean optical thickness, mean effective radius, fractional cloud cover, inhomogeneity parameters of optical thickness and effective radius, and cloud-top temperature) from multispectral and multiscale radiance data. They used horizontal resolutions of 0.25 km $\times$ 0.25 km at wavelengths of 0.544, 1.6, and 2.15 μm and 1 km $\times$ 1 km at wavelengths of 0.544, 1.6, 2.15, 3.65, and 10.8 μm. Their method was adapted to MODIS in Cornet et al. (2005). Evans et al. (2008) used an NN to estimate mean and standard deviations of COT from multiangular reflectances of boundary layer clouds.

More recently, ~~the~~ deep learning (a ~~kind of~~ machine-learning ~~techniques~~technique), which uses deep neural networks (DNNs), has become a useful tool in various applications. Deep learning involves training a DNN that has three or more layers with a network structure that is more complex than that used previously. An advantage of deep learning is automatic feature extraction: features ~~are extracted hierarchically , thereby extending applicability~~ in training datasets are learned hierarchically in the DNNs, although it is not easy to trace how the features emerge. Nevertheless, DNNs extend the applicability of NN to more complex problems. ~~A DNN is more~~ In addition, input and output parameters can easily be added, and structures can be modified - another advantage of deep learning. In combination, these advantages make DNNs more flexible under varying conditions. They are more suitable for approximating complex nonlinear functions of many variables because the degree of nonlinearity increases with the number of layers, and the ability to approximate a function ~~is generally improved by using~~ generally improves for a deeper NN. Recent advances in computer technology, such as multicore central processing units (CPUs) and general-purpose graphics processing units (GPGPUs), have facilitated calculations involving the large training datasets that are required for DNNs. In addition, a number of DNN optimization algorithms have been proposed in the past few years.

The present study is aimed at using a DNN approach to retrieve the COT and CDER of inhomogeneous clouds, and at testing the feasibility of a multispectral, multipixel approach based on DNNs. For training and testing, we use 3D cloud-field data generated by large-eddy simulation (LES) and radiances generated by a 3D radiative transfer model. The outline of this paper is as follows. Section 2 explains the cloud-field data and radiative-transfer simulations that are used to generate the training and test datasets. Section 3 describes the designs and configurations of our DNNs and the preprocessing methods. Section 4 presents results of performance comparisons for cloud retrieval using DNNs, ~~IPA~~IPA-based look-up table (LUT), and a simple NN. Finally, Section 5 concludes the paper with a discussion on the merits of DNN-based cloud retrieval.

2 Data

2.1 SCALE-LES cloud-field data

Three-dimensional cloud-field data are generated using an LES model known as SCALE-LES (Sato et al., 2014, 2015; Nishizawa et al., 2015). ~~The~~ A double-moment bulk scheme is used for the cloud microphysics. The cloud liquid-water mass mixing ratio and number density are obtained at each grid point in the domain. Figure 1 shows examples of such cloud-field data for two types of boundary-layer cloud: closed-cell and open-cell. These cloud types are simulated for polluted (closed) and clean (open) aerosol conditions (Sato et al., 2014). Clouds are optically thick in the closed case, whereas they are optically thin with large precipitation rates in the open case. Each case consists of 60 time steps at 1-min intervals. The horizontal size of the LES scenes is 28 km × 28 km. The resolution for the x- and y-axis of the cloud field is originally 35 m. For the z-axis, the resolution is 5 m at the bottom of the atmosphere, and it is coarse (less than 60 m) for the upper layers. 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. After the averaging, the finest resolution for z-axis is 40 m. The CDER is calculated as

$$r_e = \frac{1}{\chi}\left(\frac{3}{4\pi}\frac{\mathrm{LWC}}{\rho_b N}\right)^{\frac{1}{3}}, \tag{1}$$

where χ is a constant depending on width of the droplet size distribution, LWC is the liquid water content, ρ_b is the density of water, and N is the droplet number density.

As shown in Fig. 1, the extinction coefficient and CDER in both cases tend to increase with height from the cloud base toward the cloud top, although the IPA retrieval assumes a homogeneous cloud. The CDER has a particularly inhomogeneous vertical structure in the closed-cell case. In the open-cell case, the CDER spatial variability is high in general, particularly so in the uppermost core parts of cells. ~~In this study , the column-mean~~ 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, where the vertically-averaged LWC and number density are used to ~~calculate~~ define the column-mean effective radius R_e~~, which is defined as~~ :

$$R_e = \frac{1}{\chi}\left(\frac{3}{4\pi}\frac{\langle \mathrm{LWC}\rangle}{\rho_b \langle N\rangle}\right)^{\frac{1}{3}}, \tag{2}$$

where $\langle . \rangle$ denotes the ~~column mean~~ mean over cloud column. Note the similarity between the definition of R_e in Eq. (2) and that of r_e in Eq. (1). ~~The~~ R_e ~~is considered as droplet size representative for each cloud column, and retrieval performance of R_e~~ represents droplet size of a cloud column. The retrieval performance for this parameter will be discussed in Section 4. It should be pointed out that there are other possibilities for column-average CDER (Miller et al., 2016).

Figure 2 shows temporal variations of (a) the domain-mean COT (specified at 0.55 μm throughout this paper), (b) the domain-mean column-mean CDER, (c) the cloud fraction, and (d) the inhomogeneity index H~~. Throughout this paper, we take the COT to be that at a wavelength of 0.55 μm. The horizontal inhomogeneity index H is~~ , defined as

$$H = \frac{\sigma_\tau^2}{\overline{\tau}^2}, \tag{3}$$

where σ_τ is the standard deviation of the COT and $\overline{\tau}$ is the mean COT. The coefficient of COT variation, $\sqrt{H}$, has been used often in previous studies (Szczap et al., 2000; Liang et al., 2009; Liang and Girolamo, 2013). Clouds in the closed-cell case are optically thick and horizontally homogeneous, covering almost the entire sky and giving a high cloud fraction. Therefore, as can be seen in Fig. 2(a, b), the domain-averaged COT and CDER remain almost constant over the entire period. In contrast, clouds in the open-cell case are distributed sparsely, meaning that the inhomogeneity index H is larger than that in the closed-cell case and increases gradually over time. The domain-averaged CDER is larger in the open-cell case than it is in the closed-cell case.

2.2 Radiative transfer simulations

A Monte Carlo 3D radiative transfer model known as MCARaTS (Monte Carlo Radiative Transfer Simulator; Iwabuchi (2006)) is used to simulate the cloud radiances. The radiances reflected in the zenith direction are calculated for solar zenith angles (SZAs) of 20° and 60° at wavelengths of 0.86, 1.64, 2.13, and 3.75 μm. The aerosol optical properties are derived using the one-dimensional RSTAR6b radiative transfer code (Nakajima and Tanaka, 1986, 1988). The aerosol optical thickness is assumed to be 0.2, and the rural aerosol model is used (Hänel, 1976). A correlated k-distribution is used for gaseous absorption by H_2O, CO_2, O_3, N_2O, CO, CH_4, and O_2 molecules (Sekiguchi and Nakajima, 2008). Rayleigh scattering by air molecules is included in the scattering process. The particle size distribution of water cloud droplets is expressed as a log-normal *volume* (V) distribution

$$\frac{\mathrm{d}V}{\mathrm{d}\ln r} = C\exp\left[-\frac{1}{2}\left(\frac{\ln r - \ln r_{\mathrm{mod}}}{\ln s}\right)^2\right], \tag{4}$$

where r is the particle radius, C is the maximum value of the volume distribution at mode radius r_{mod}, and s is the width of the distribution. In this study, we assume $s = 1.5$. The CDER r_{e} is related to r_{mod} by $r_{\mathrm{e}} = r_{\mathrm{mod}}\exp(-1/2\times(\ln s)^2)$. The χ parameter in Eqs. (1) and (2) is determined as $\chi = r_{\mathrm{vol}}/r_e = \exp(-\ln^2 s) = 0.84$, where r_{vol} is the volume mean radius. The scattering properties of water cloud droplets are calculated using the Lorenz–Mie theory (Bohren and Huffman, 1983). For simplicity, the underlying surface is approximated as black.

3 Method

3.1 ~~Design and configuration of DNNs~~Deep learning techniques

~~Each layer in the DNNs~~ This section briefly explains fundamentals of deep learning. This study uses Chainer, an NN framework developed by Tokui et al. (2015), for implementating DNN cloud retrieval methods. Chainer is used in a wide variety of research fields because it covers common functions and algorithms for constructing and training DNNs and provides easy access to efficient GPU-based computation. It is easy to test different DNN structures and learning techniques, and there is high degree of freedom for choice. Interested readers may refer to Tokui et al. (2015) for details. Several other deep learning frameworks have been developed for general purposes and are publicly available.

The DNNs consists of multiple network ~~units~~layers, each of which consists of units. Each unit receives input signals from the previous layer ~~,~~ and generates an output signal. If each unit in a layer is connected with all units in the previous layer, that layer is called fully connected. The unit computes a weighted sum and ~~add a bias ,~~ adds a bias as follows:

$$x = \sum_k w_k x'_k + b, \tag{5}$$

where x'_k is the kth input signal, w_k is the corresponding weight, and b is the bias. The weights and the bias are determined ~~in~~ at the training stage. The result x is usually transformed by a function known as the activation function to obtain an output signal. In this study, we use a rectified linear function (Nair and Hinton, 2010) defined as

$$f(x) = \max(0, x) \tag{6}$$

for the activation function. ~~Of~~ Among the various activation functions used for NNs, this rectified linear function is relatively simple, leads to good learning efficiency, and is ~~the one used most commonly~~ most commonly used in recent DNN applications.

Image recognition is often implemented through convolutional NNs because they enable pattern recognition in an image. A convolutional layer consists of units that compute the convolution on the input image. For multichannel images, multiple structural filters operate within a convolutional layer. A convolutional signal x_m of the mth output channel at a pixel is represented as

$$x_m = \sum_l \sum_k w_{k,l,m} x'_{k,l,m} + b_m, \tag{7}$$

where $x'_{k,l,m}$ and $w_{k,l,m}$ are the input signal and the corresponding filter weight for the kth pixel (the target or an adjacent pixel) and lth input channel. The number of required filters is the product of the number of input channels (wavelengths) and output channels. The summation over k in Eq. (7) operates only on the target pixel and its neighbors. Unlike a fully connected layer, a convolutional layer has the following two characteristics: 1) the input and output signals of a convolutional layer are sparsely connected, and 2) the filter profiles are defined independently for input channels but are shared among all pixels; the filter profile does not depend on pixel location in the input image.

Current deep learning techniques are flexible to use a DNN structure by connecting multiple fully-connected and convolutional layers in a complicated way, as shown in the next subsection. During the training, the DNN parameters are optimized to minimize the so-called loss function. In this study, the loss function is the sum of the squared residuals between the DNN output and ideal data in the training dataset. For the DNN optimization, we use the Adam (Adaptive moment estimation; Kingma and Ba (2014)) algorithm, which automatically determines the learning rate at each training step using the mean and variance of the loss function. An NN is expected to deliver meaningful and accurate retrievals for the dataset that it was trained on. However, in some cases, the NN can be overfitted to the training dataset, thereby losing its ability to generalize, and performing appreciably worse for other data. Such overfitting is a serious issue in NNs. In the present study, we use the dropout technique (Srivastava et al., 2014) to overcome this problem. The dropout technique removes randomly selected units from the NN at each step in the training stage, decreasing the number of degrees of freedom of the NN, thus avoiding overfitting. An

NN trained with the dropout technique can work like ensemble estimation that uses many different independently trained NNs. Dropout results in better performance and is widely used in many applications.

3.2 Design and configuration of DNNs

The DNNs used in this study are designed to estimate COT and column-mean CDER simultaneously at multiple pixels from multipixel, multispectral radiances. This is a unique ~~point~~ approach compared to previous studies. Larger input and output vectors allow more degrees of freedom for the features to be learned in the DNNs. Two types of DNN were constructed:

1. DNN-2r (with IPA retrieval and two wavelengths) that corrects IPA retrievals based on 0.86 and 2.13 μm radiances using the radiances at those same wavelengths (0.86 and 2.13 μm);

2. DNN-4w (with four wavelengths) that uses ~~the so-called convolutional layer~~ convolutional layers and retrieves cloud properties directly from the radiances at 0.86, 1.64, 2.13, and 3.75 μm.

It should be noted that only solar radiation is considered in the present study, which requires that the thermal radiation at 3.75 μm be corrected during pre-processing.

The DNN-2r network is designed to correct the IPA retrieval of COT and CDER ~~that originated from multispectral radiances~~from the bispectral retrieval for 3D effects. The elements of the DNN-2r input vector are the radiances at the wavelengths of 0.86 and 2.13 μm~~, and~~ . Prior to applying DNN-2r, the COT and CDER ~~estimated by the IPA retrieval~~ are estimated by IPA for 10×10 pixels at ~~280-m~~ 280 m resolution. Thus, the input vector has $400 = (10 \times 10 \times (2+2))$ elements. Figure 3 shows the DNN-2r structure schematically; the COT and CDER distributions are estimated at 8×8 pixels at the center of the input field, and the output vector has $128 = (8 \times 8 \times 2)$ elements. The reason for including margin pixels in the input field is to take into account the 3D radiative effects from the surroundings of the cloud field. The DNN-2r network consists of several fully connected layers~~in which each unit is connected with all units in the previous layer~~. In the first layers, radiances and IPA-estimated cloud properties are merged to obtain $8 \times 8 \times 2$ elements (two elements per pixel for 8×8 pixels). The final part of DNN-2r consists of two independent groups of layers that finally estimate the COT and CDER. As in the residual network designed by He et al. (2015), the DNN-2r network has what are known as shortcuts, which allow residuals to be learned. The NN ~~should be~~ is trained to predict the correction terms ~~that are~~ to be added to the data from the shortcut path ~~. Such shortcuts facilitate machine learning ,~~ represented in Fig. 3 as bypassing route of data. Such shortcuts make machine learning possible even in cases with many NN layers. In this way, the DNN-2r network can be considered a way to correct the IPA retrievals.

The DNN-4w structure is shown schematically in Fig. 4. The input comprises radiance distributions at four wavelengths (0.86, 1.64, 2.13, and 3.75 μm) and 10×10 pixels of ~~280-m~~ 280 m resolution. Thus, the input vector has $400 = (10 \times 10 \times 4)$ elements. Unlike in DNN-2r, the COT and CDER distributions in DNN-4w are predicted at the center of 6×6 pixels of the input field, and the output vector has $72 = (6 \times 6 \times 2)$ elements. ~~As well as~~ In addition to shortcuts, the DNN-4w network has two convolutional layers~~that consist of units that compute the convolutions~~. In the first convolutional layer, convolutions operate on 5×5 pixels surrounding the center pixel, with 100 different profiles of filter ~~weight~~ weights for each wavelength. ~~The number of filters is a product of the numbers of input channels (wavelengths) and output channels.~~ There are 400 filters in

the first convolutional layer because the number of input wavelengths is 4 and that of output channels is 100. A convolutional signal x_m of the mth output channel at a pixel is represented as

$$x_m = \sum_l \sum_k w_{k,l,m} x'_{k,l,m} + b_m,$$

where $x'_{k,l,m}$ and $w_{k,l,m}$ are the input signal and the corresponding filter weight for the kth pixel (the target or a adjacent pixel) and lth input channel. The summation over k in Eq. (7) operates not all but only for the target pixel and adjacent pixels. As shown in Fig. 4, the activation function is and is not applied to the signal x_m in the first and second, respectively, convolutional layer. Unlike a fully connected layer, a convolutional layer has the following two characteristics: 1) the input and output signals of a convolutional layer are sparsely connected, and 2) the filter profiles are defined independently for input channels (wavelengths) but are shared among all pixels; the filter profile does not depend on pixel location in the input image. A convolutional NN can detect a specific pattern in an image and is commonly used with high performance in image recognition. In the problem with which the present study is concernedconvolutional layer, but not in the second. By using these convolutional layers, we expect that using a convolutional layer will allow the DNN to learn patterns that characterize the DNN learns image patterns that capture the 3D radiative effects among between the target pixel and those adjacent to it. We expect the its surroundings. The DNN-4w network firstly to correct the corrects for 3D radiative transfer effects and then to transform transforms the signals to COT and CDER with the possibility of additional corrections of the 3D effect in this latter partcorrections at this stage.

Chainer, an NN framework developed by Tokui et al. (2015), is used to construct the DNNs. Chainer is used in a wide variety of research fields because it covers common functions and algorithms for constructing DNNs and provides easy access to efficient GPU-based computation. In the training, the DNN parameters are optimized to minimize the loss function, which is the sum of the squared residuals between the DNN output and ideal data in the training dataset. For this optimization, we use the Adam (Adaptive moment estimation; Kingma and Ba (2014)) algorithm, which automatically determines the learning rate at each training step using the mean and variance of the loss function.

An NN is expected to deliver meaningful and accurate retrievals for the dataset that it was trained onThe above two DNN structures were obtained from various trial-and-error experiments. Different DNN structures were also tested. For example, we tested a DNN similar to DNN-2r but with four wavelengths, and one similar to DNN-4w but with only two wavelengths. However, in some cases, the NN can be overfitted to the training dataset, thereby losing its ability to generalize and performing appreciably poorer for other data. Such overfitting is a serious problem in NNs. In the present study, we use the dropout technique (Srivastava et al., 2014) to overcome this problem. The dropout technique removes randomly selected units from the NN at each step in the training stage, decreasing the number of degrees of freedom of the NN and avoiding overfitting. An NN trained with dropout can work like ensemble estimation that uses many different NNs that were trained independently. Dropout results in better performance and is widely used in many applications. DNN-2r and DNN-4w performed best. There is room for improvement in DNN structures, which should be investigated in the future.

3.3 Generation of the training and test datasets

A training dataset is necessary for machine learning. In this study, the training dataset is generated as follows. The zenith radiances are calculated using MCARaTS with 10^5 model photons incident on each pixel, which results in Monte Carlo noise of approximately 1%. Such noise can be interpreted as measurement noise in the present problem. From two cases of SCALE-LES cloud-field data, $1,977,440$ samples (10×10 areas) are chosen randomly for the training datasets. As shown in Fig. 2, the 25th to 75th percentile ranges for COT are 0–5 and 11–15 for the open- and closed-cell cases, respectively. With a DNN, a variety of training data is important for better generalization performance. To increase the variety of the COT training data, one half was generated from original cloud data, whereas the other half was generated from artificially modified cloud fields in which the cloud extinction coefficients were multiplied by numbers chosen randomly from the range 0.5–1.5. The cloud extinction coefficients of all pixels within a single cloud scene were multiplied by the same number.

~~To construct an efficient DNN~~ Although DNN can generally approximate nonlinear functions, it is ~~worth investigating the relationships between the~~ 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. In the DNN preprocessing, the cloud properties are transformed using

$$F(\tau) = \frac{(1-g)\tau}{1+(1-g)\tau}, \tag{8}$$

$$G(r_e) = \sqrt{r_e}, \tag{9}$$

where g is the asymmetry parameter. As a representative value for water droplets, we set $g = 0.86$ for preprocessing purposes only. After the above transformations, all the DNN input and output data, including the radiances and cloud properties, are normalized as

$$z'_{i,j} = \frac{z_{i,j} - \overline{z_j}}{\sigma_j}, \tag{10}$$

where $z_{i,j}$ is the jth element of an input or output vector in the ith sample, and $\overline{z_j}$ and σ_j are the mean and standard deviation, respectively, of the jth element over the all samples. This is referred to as z-score normalization and is known to improve the efficiency of a DNN (Kotsiantis et al., 2006; Nawi et al., 2013).

The test dataset used for evaluation should be independent of the training dataset. In the present study, ~~we generate the test datasets in the same way as we do the training dataset, but with different random selections~~the training and test datasets include different randomly selected locations within the cloud fields, but the statistics of cloud properties are nearly identical in the training and test datasets. The test datasets include $10,000$ samples. As for computational cost, the training requires significant computation time, for which even one GPU helps considerably. Once the DNNs are trained, the retrievals using the present DNNs are generally very quick because they entail only very few simple manipulations of numerical data.

4 Results

In this section, we illustrate the ability of DNNs to retrieve cloud properties, and we compare ~~this~~ it with the corresponding abilities of existing methods. ~~Values~~ The values of COT and CDER are retrieved from test datasets by using DNNs and IPA retrievals. The retrieved values are compared to the true values in the test datasets, and the retrieval errors at each pixel are evaluated. In the IPA retrieval, COT and CDER are estimated from ~~look-up tables~~ LUTs of radiances at the wavelengths of 0.86 and 2.13 μm. These wavelengths are used in the MODIS product for retrieving cloud properties over oceans (Platnick et al., 2003). ~~Also in the IPA retrieval, the~~ The lower and upper limits for COT are zero and 150, respectively, and those for CDER are zero and 55 μm, respectively. If any radiance ~~strays beyond~~ falls outside the associated range defined by the ~~look-up tables~~LUTs, the COT/CDER value is forced to be the lower or upper limit, as appropriate.

4.1 Retrieval results for DNN-2r and DNN-4w

Figure 5 shows examples of the IPA and DNN-4w retrieval results for an open-cell case with a SZA of 60° and a viewing zenith angle of 0°. Cross sections ~~taken~~ at $y = 14.56$ km are shown in Fig. 6 with additional DNN-2r retrieval results. In Fig. 6, the relative error of estimated cloud properties are also shown. On this cross section, the RMSE of estimated cloud properties are shown in Table 1. The sunny (left-hand) side of the COT fluctuation peak is directly illuminated by the Sun. This is noticeable at locations around 11 km and 16.5 km. For pixels on that side, the radiances calculated by 3D radiative transfer are brightened (~~illuminating~~ illumination effect), which results in ~~the overestimation (resp. an~~ overestimation (underestimation) of ~~IPA retrievals of COT (resp. CDER)~~ COT (CDER) by IPA retrievals. For pixels on the opposite (right-hand) side, the radiances are darkened (shadowing effect)~~and IPA retrieval of COT (resp.~~ , and COT (CDER) is underestimated (~~resp. overestimated) . These illuminating~~ overestimated) by IPA. These illumination and shadowing effects have considerable influence on the IPA retrieval ~~. A phase lag appears~~ (Várnai and Davies, 1999; Várnai and Marshak, 2002; Marshak et al., 2006). These effects lead to a spatial offset in the IPA-retrieved horizontal COT distribution~~because of this illuminating and shadowing; the IPA error in the COT~~ , and the COT IPA retrieval error is particularly large for optically thick ~~parts. In contrast, the DNN retrieves~~ pixels. This can be ascribed to weaker sensitivity of radiance to COT when COT is larger; a small difference in radiance leads to large difference in retrieved COT. Compared to the IPA retrieval, the DNNs retrieve COT values that are ~~close~~ closer to the true values assumed in this test~~, successfully correcting~~ . The DNN-4w successfully corrects the phase lag ~~.~~ as shown in Fig. 6c. However, minor errors are still present in the DNN-retrieved COT.

As for the retrieved values of CDER, the DNN ones are obviously better than the ~~IPA~~ IPA-retrieved ones. The CDER is noticeably overestimated at pixels for which the shadowing effect decreases the radiance, a fact that we attribute to the strong nonlinear dependence of CDER on radiance. As a result, a positive bias appears in the IPA-retrieved CDER, which also shows an appreciable fluctuation at small horizontal scales because SWIR radiances are sensitive to cloud-top variability at such scales (Iwabuchi and Hayasaka, 2003). In contrast, the DNN-retrieved CDER is generally highly accurate, although small-scale fluctuations of CDER are not very well reproduced.

The COT and CDER retrieval errors are evaluated for all the test datasets, and the mean and standard deviation of the relative errors are calculated in bins that are equally spaced in the logarithm of COT and CDER. The results are evaluated using $360,000$ pixels for each SZA. In Fig. 7, the IPA retrieval and DNN-4w relative errors are plotted against the true COT and CDER values. The IPA-retrieved COT error and its standard deviation are particularly large for a SZA of $60°$, ~~at which the~~ where radiative roughening causes the 3D radiance to deviate from the IPA radiance. Both the COT and CDER retrieval errors are reduced considerably by using the DNN, which suggests that the DNN is well trained to correct the 3D radiative transfer effects. The DNN mean bias errors are generally closer to zero than ~~are the IPA ones~~those from the IPA retrieval. Compared to the IPA retrieval, the DNN retrieves COT better, even at optically very thick pixels. In particular, the COT error is markedly reduced for true COT values greater than 5 and for an SZA of $60°$. At pixels with small COT (1 or less), the DNN overestimates COT, although the errors are still smaller than those ~~for~~ from the IPA retrieval.

The DNN also yields better CDER retrievals than does the IPA retrieval, with much smaller variability of CDER errors. For SZAs of $20°$ and $60°$, the IPA-retrieved CDER tends to be overestimated over almost the entire range of CDER. The IPA retrieval shows a particularly large bias when the true CDER is small, although very few data are available for CDER values less than 15 μm, as shown in Fig. 2. This overestimation of CDER can be partly attributed to the neglect of vertical inhomogeneity in the IPA retrieval. The reflected SWIR radiances (2.13 μm) give information about the cloud microphysical status only near the cloud top (Platnick, 2000), and the IPA-retrieved CDER is associated primarily with the CDER near the cloud top (Nakajima et al., 2010; Zhang et al., 2012; Nagao et al., 2013). IPA-retrieved CDER thus tends to be larger than column-mean CDER, whereas DNNs are by design trained to ~~estimate the~~ learn the relationship between the column-mean CDER and radiances by taking into account the vertical inhomogeneity. However, ~~overestimation of~~ this vertical inhomogeneity effect on the IPA-retrieved CDER does not seem to be the main cause of the large positive bias. Overestimation of CDER in the IPA retrieval is mainly observed at the shadowed pixels, as shown in Figs. 5 and 6. The IPA retrieval also shows large values of standard deviation of the relative errors, particularly for small values of CDER. Figures 5(a) and 5(b) show that the CDER tends to be smaller at pixels with small COT, where the shadowing tends to reduce the SWIR radiance. A small radiance perturbation due to 3D effects may result in a large error in the retrieved CDER because of the weaker sensitivity of SWIR radiance to CDER in cases of small COT. However, the DNN-retrieved values of column-mean CDER are close to the true values.

Table 2 shows the relative RMSE of estimated COT and CDER by the IPA and DNN retrievals for open-cell and closed-cell cases. In both cases, the retrieval accuracies for DNNs are obviously improved compared to the IPA retrieval. An exception appears for COT in closed-cell case when SZA is $20°$; COT RMSE of 24% for DNN is larger than 16% for IPA. The DNN-4w is better than DNN-2r in both cases.

Figure 8 shows selected examples of the trained (5×5)-pixel filters of the first convolutional layer used in DNN-4w for a SZA of $60°$. It is noted that the convolutional layer is designed to correct the 3D radiative effects appeared in radiances (Fig. 4). Only 16 out of 100 filters are shown here, and each filter weight can be either positive or negative. At the beginning, filters are initialized with white noise. Once the DNN has been well-trained, this noise is replaced with features that emerge from patterns in the input images. The patterns in some filters (e.g., first and eleventh ones from the left) are nearly symmetrical around the

center pixel with various spatial profiles, which suggests that they extract features that characterize the relationship between the center pixel and those adjacent to it. For example, isotropic smoothing and second-order central difference operators have such a symmetrical pattern. Also, several filters (e.g., fifth one from the left) have higher weights in pixels along the solar azimuth direction, which ~~suggest~~ suggests a feature related to the ~~solar~~ direct beam that operates along that direction. In our design of DNN-4w, different filters operate at each wavelength independently, whereas most of the obtained filters show strong correlations among wavelengths. These patterns suggest that the combination of filter patterns in the DNN ~~works to correct~~ corrects 3D radiative effects ~~to recover the information about local cloud properties~~and thus recovers the local cloud property information. However, it is ~~difficult at present~~ presently difficult to understand which combinations of filter patterns perform such corrections in the DNN, or indeed how they do so.

4.2 Comparison with previous work using a neural network

It is of interest to compare the performance of our present DNN with that of the NN used previously ~~by~~ as the second method in Faure et al. (2001). Originally, this NN had two hidden layers with 10 units each ~~,~~and it used the 0.8 km × 0.8 km area-averaged radiances at four wavelengths (0.64, 1.6, 2.2, 3.7 μm) at the target pixel and eight adjacent pixels (called as "with ancillary data"). It is described in the section 3.3 (2) in the original paper. However, in this comparison, we construct an NN with 512 units in each layer to allow more degrees of freedom. The NN inputs for the present study are the radiances at four wavelengths (0.86, 1.64, 2.13, and 3.75 μm) at the target pixel and eight adjacent pixels, as in the second method in Faure et al. (2001), and the outputs are COT and CDER. In this comparison, the resolution of the pixels is 280 m. Data used for training and test for the NN are from the same original datasets used for DNNs.

It is of interest to compare the performance of our present DNN with that of the NN used previously as the second method in Faure et al. (2001), section 3.3 (2). Originally, this NN had two hidden layers with 10 units each and it used 0.8 km × 0.8 km area-averaged radiances at four wavelengths (0.64, 1.6, 2.2, 3.7 μm) at the target pixel and eight adjacent pixels (denoted as "ancillary data"). Rather than reproducing their NN exactly, we construct an NN with 512 units in each layer for the sake of this comparison to allow more degrees of freedom. As NN inputs, we use the radiances at four wavelengths (0.86, 1.64, 2.13, and 3.75 μm) at the target pixel and eight adjacent pixels. As in Faure et al. (2001), the outputs are COT and CDER, but with a pixel resolution of 280 m. Data for training and testing the NN are from the same original datasets used for DNNs.

Figure 9 shows comparisons of the NN and our DNNs. For a SZA of 20°, the COT is well retrieved for true COTs of 10–50 for both the NN and DNNs. When the true COT is less than 10, the COT values from the NN and DNN-4w retrievals are overestimated more for optically thinner clouds, although DNN-2r gives better estimates. The COT estimated by the NN tends to be underestimated when the true COT is larger than 50, whereas DNN-2r and DNN-4w yield better retrievals in this range. For an SZA of 60°, the DNN retrievals of COT are generally better than the NN retrievals. The COT retrievals by the NN tend to be overestimated (resp. underestimated) for optically thin (resp. thick) clouds. This suggests that 3D radiative effects with low sun are not well modeled in the current NN because it uses only 3 × 3 pixels, whereas the DNNs use 10 × 10 pixels. Moreover, the multiple convolutional layers in the DNNs are more powerful for representing the complex 3D radiative effects compared to the layers in the NN. In general, the DNN-2r retrievals show large error variability, with the largest standard

deviation among the three methods. The CDER is well retrieved by all three methods (NN and DNNs) when the true CDER is larger than 10 μm, although overestimating smaller CDERs is common among the three methods.

5 Conclusions

In this study, the feasibility of a multispectral, multipixel approach to retrieving COT and CDER ~~using~~ with a deep learning technique has been investigated. Two types of DNN were constructed: 1) DNN-2r that corrects IPA retrievals using the reflectances at two wavelengths, and 2) DNN-4w that uses convolutional layers and retrieves cloud properties directly from the reflectances at four wavelengths. Both DNNs retrieve multipixel estimates of COT and CDER simultaneously from multispectral, multipixel radiances. The DNNs were trained and evaluated ~~by using~~ with SCALE-LES cloud-field data ~~whose horizontal resolution was~~ at a horizontal resolution of 280 m. Both DNNs outperformed IPA-based retrieval in ~~relation to~~ terms of accuracy, and showed better ability to represent 3D radiative effects compared to that of an NN used in previous work. The CDER retrievals of both DNNs were considerably better than the corresponding IPA retrieval. Whereas the IPA retrieval appreciably overestimated the CDER at pixels that were affected by shadowing, the DNNs successfully corrected such 3D effects. The DNN-4w network was generally more accurate than the DNN-2r network. Information that was lost in the IPA retrieval when the radiances came from ~~look-up tables~~ LUTs made for plane-parallel clouds limited the ability of the DNN-2r network to correct those retrievals sufficiently well. In contrast, the DNN-4w network does not use an IPA retrieval in its input, and therefore is more robust at retrieving cloud properties. In addition, multipixel information and convolutional layers were shown to be efficient in improving cloud retrievals with 3D radiative effects taken into account.

In the DNN-4w that we tested, we excluded 3D radiative transfer effects that occurred at horizontal scales greater than ~~approximately 1.5 km (5~~560 m (2 pixels). In addition, we ~~considered~~ only considered a cloud thickness of ~~only~~ less than 0.9 km, as shown in Fig. 1. ~~Therefore, it would~~ It would therefore be interesting to test the sensitivity and performance of the algorithm for input vectors for wider areas (more pixels) of cloud. This is because 3D radiative transfer effects are known to operate on horizontal scales that are determined mainly by cloud thickness and solar zenith angle (Marshak and Davis, 2005). In the future, the application of DNNs to cloud remote sensing is expected to become more common. However, using DNNs with actual satellite data will require training using realistic cloud fields for various types of cloud. Incorporating more parameters (e.g., sun–cloud–satellite geometry, surface albedo, aerosols, spectral and spatial specifications of sensors) into the method will also be necessary to handle the complexities of such measurement data. It is an advantage of this method that it is easy to incorporate such parameters into DNN. An appropriate DNN architecture for addition of input parameters should be investigated in the future.

Acknowledgements. SCALE-LES was developed by the SCALE interdisciplinary team at the RIKEN Advanced Institute for Computational Science (AICS), Japan. Some of the results in this paper were obtained using the K supercomputer at RIKEN AICS. The authors are grateful to Dr. Yousuke Sato of RIKEN for providing the SCALE-LES simulation data. Some of the results of radiative transfer calculations in this

paper were obtained using the server and visualization systems of the Graduate School of Simulation Studies, University of Hyogo. We would also like to thank the OpenCLASTR project for the use of the RSTAR radiative transfer model. This work was partly supported by a Grant-in-Aid for Scientific Research (B) (KAKENHI Grant No. 15H03729) of the Japan Society for the Promotion of Science (JSPS). Sebastian Schmidt was supported through NASA grant NNX15AQ19G (Remote Sensing Theory).

References

Bohren, C. F. and Huffman, D. R.: Absorption and scattering of light by small particles, Wiley, 1983.

Cornet, C., Isaka, H., Guillemet, B., and Szczap, F.: Neural network retrieval of cloud parameters of inhomogeneous clouds from multispectral and multiscale radiance data: Feasibility study, Journal of Geophysical Research: Atmospheres, 109, https://doi.org/10.1029/2003JD004186, 2004.

Cornet, C., Buriez, J.-C., Riédi, J., Isaka, H., and Guillemet, B.: Case study of inhomogeneous cloud parameter retrieval from MODIS data, Geophysical research letters, 32, https://doi.org/10.1029/2005GL022791, 2005.

Evans, K. F., Marshak, A., and Várnai, T.: The potential for improved boundary layer cloud optical depth retrievals from the multiple directions of MISR, Journal of the Atmospheric Sciences, 65, 3179–3196, 2008.

Faure, T., Isaka, H., and Guillemet, B.: Neural network retrieval of cloud parameters of inhomogeneous and fractional clouds: Feasibility study, Remote Sensing of Environment, 77, 123–138, https://doi.org/10.1016/S0034-4257(01)00199-7, 2001.

Faure, T., Isaka, H., and Guillemet, B.: Neural network retrieval of cloud parameters from high-resolution multispectral radiometric data: A feasibility study, Remote Sensing of Environment, 80, 285–296, https://doi.org/10.1016/S0034-4257(01)00310-8, 2002.

Grosvenor, D. and Wood, R.: The effect of solar zenith angle on MODIS cloud optical and microphysical retrievals within marine liquid water clouds, Atmospheric Chemistry and Physics, 14, 7291–7321, https://doi.org/10.5194/acp-14-7291-2014, 2014.

Hänel, G.: The properties of atmospheric aerosol particles as functions of the relative humidity at thermodynamic equilibrium with the surrounding moist air, Advances in geophysics, 19, 73–188, 1976.

He, K., Zhang, X., Ren, S., and Sun, J.: Deep residual learning for image recognition, arXiv preprint arXiv:1512.03385, 2015.

Iwabuchi, H.: Efficient Monte Carlo Methods for Radiative Transfer Modeling, Journal of the Atmospheric Sciences, 63, 2324–2339, https://doi.org/10.1175/JAS3755.1, 2006.

Iwabuchi, H. and Hayasaka, T.: Effects of cloud horizontal inhomogeneity on the optical thickness retrieved from moderate-resolution satellite data, Journal of the Atmospheric Sciences, 59, 2227–2242, 2002.

Iwabuchi, H. and Hayasaka, T.: A multi-spectral non-local method for retrieval of boundary layer cloud properties from optical remote sensing data, Remote Sensing of Environment, 88, 294–308, https://doi.org/10.1016/j.rse.2003.08.005, 2003.

Kingma, D. and Ba, J.: Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980, 2014.

Kotsiantis, S., Kanellopoulos, D., and Pintelas, P.: Data preprocessing for supervised leaning, International Journal of Computer Science, 1, 111–117, 2006.

Liang, L. and Girolamo, L. D.: A global analysis on the view-angle dependence of plane-parallel oceanic liquid water cloud optical thickness using data synergy from MISR and MODIS, Journal of Geophysical Research: Atmospheres, 118, 2389–2403, https://doi.org/10.1029/2012JD018201, 2013.

Liang, L., Di Girolamo, L., and Platnick, S.: View-angle consistency in reflectance, optical thickness and spherical albedo of marine water-clouds over the northeastern Pacific through MISR-MODIS fusion, Geophysical Research Letters, 36, https://doi.org/10.1029/2008GL037124, 2009.

Marshak, A. and Davis, A.: 3D radiative transfer in cloudy atmospheres, Springer Science & Business Media, 2005.

Marshak, A., Platnick, S., Várnai, T., Wen, G., and Cahalan, R. F.: Impact of three-dimensional radiative effects on satellite retrievals of cloud droplet sizes, Journal of Geophysical Research: Atmospheres, 111, https://doi.org/10.1029/2005JD006686, 2006.

Miller, D. J., Zhang, Z., Ackerman, A. S., Platnick, S., and Baum, B. A.: The impact of cloud vertical profile on liquid water path retrieval based on the bispectral method: A theoretical study based on large-eddy simulations of shallow marine boundary layer clouds, Journal of Geophysical Research: Atmospheres, 121, 4122–4141, https://doi.org/10.1002/2015JD024322, http://dx.doi.org/10.1002/2015JD024322, 2015JD024322, 2016.

Minnis, P., Hong, G., Sun-Mack, S., Smith, W. L., Chen, Y., and Miller, S. D.: Estimating nocturnal opaque ice cloud optical depth from MODIS multispectral infrared radiances using a neural network method, Journal of Geophysical Research: Atmospheres, 121, 4907–4932, https://doi.org/10.1002/2015JD024456, 2016.

Nagao, T. M., Suzuki, K., and Nakajima, T. Y.: Interpretation of multiwavelength-retrieved droplet effective radii for warm water clouds in terms of in-cloud vertical inhomogeneity by using a spectral bin microphysics cloud model, Journal of the Atmospheric Sciences, 70, 2376–2392, https://doi.org/10.1175/JAS-D-12-0225.1, 2013.

Nair, V. and Hinton, G. E.: Rectified linear units improve restricted boltzmann machines, in: Proceedings of the 27th International Conference on Machine Learning (ICML-10), pp. 807–814, 2010.

Nakajima, T. and King, M. D.: Determination of the optical thickness and effective particle radius of clouds from reflected solar radiation measurements. Part I: Theory, Journal of the Atmospheric Sciences, 47, 1878–1893, 1990.

Nakajima, T. and Tanaka, M.: Matrix formulations for the transfer of solar radiation in a plane-parallel scattering atmosphere, Journal of Quantitative Spectroscopy and Radiative Transfer, 35, 13–21, https://doi.org/10.1016/0022-4073(86)90088-9, 1986.

Nakajima, T. and Tanaka, M.: Algorithms for radiative intensity calculations in moderately thick atmospheres using a truncation approximation, Journal of Quantitative Spectroscopy and Radiative Transfer, 40, 51–69, https://doi.org/10.1016/0022-4073(88)90031-3, 1988.

Nakajima, T. Y., Suzuki, K., and Stephens, G. L.: Droplet growth in warm water clouds observed by the A-Train. Part I: Sensitivity analysis of the MODIS-derived cloud droplet sizes, Journal of the Atmospheric Sciences, 67, 1884–1896, https://doi.org/10.1175/2009JAS3280.1, 2010.

Nawi, N. M., Atomi, W. H., and Rehman, M. Z.: The effect of data pre-processing on optimized training of artificial neural networks, Procedia Technology, 11, 32–39, 2013.

Nishizawa, S., Yashiro, H., Sato, Y., Miyamoto, Y., and Tomita, H.: Influence of grid aspect ratio on planetary boundary layer turbulence in large-eddy simulations, Geoscientific Model Development, 8, 3393–3419, https://doi.org/10.5194/gmd-8-3393-2015, 2015.

Platnick, S.: Vertical photon transport in cloud remote sensing problems, Journal of Geophysical Research: Atmospheres, 105, 22 919–22 935, https://doi.org/10.1029/2000JD900333, 2000.

Platnick, S., King, M. D., Ackerman, S. A., Menzel, W. P., Baum, B. A., Riédi, J. C., and Frey, R. A.: The MODIS cloud products: Algorithms and examples from Terra, IEEE Transactions on Geoscience and Remote Sensing, 41, 459–473, 2003.

Sato, Y., Nishizawa, S., Yashiro, H., Miyamoto, Y., and Tomita, H.: Potential of retrieving shallow-cloud life cycle from future generation satellite observations through cloud evolution diagrams: A suggestion from a large eddy simulation, SOLA, 10, 10–14, https://doi.org/10.2151/sola.2014-003, 2014.

Sato, Y., Nishizawa, S., Yashiro, H., Miyamoto, Y., Kajikawa, Y., and Tomita, H.: Impacts of cloud microphysics on trade wind cumulus: which cloud microphysics processes contribute to the diversity in a large eddy simulation?, Progress in Earth and Planetary Science, 2, 1, https://doi.org/10.1186/s40645-015-0053-6, 2015.

Sekiguchi, M. and Nakajima, T.: A k-distribution-based radiation code and its computational optimization for an atmospheric general circulation model, Journal of Quantitative Spectroscopy and Radiative Transfer, 109, 2779–2793, https://doi.org/10.1016/j.jqsrt.2008.07.013, 2008.

Table 1. The RMSE of estimated COT and CDER by the IPA retrieval and DNNs for a view zenith angle of 0° at $y = 14.56$ km in Fig. 5.

Retrieval method	COT	CDER [micron]
IPA retrieval	11.36	9.33
DNN-2r	5.64	1.64
DNN-4w	4.40	1.37

Srivastava, N., Hinton, G. E., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting., Journal of Machine Learning Research, 15, 1929–1958, 2014.

Szczap, F., Isaka, H., Saute, M., Guillemet, B., and Ioltukhovski, A.: Effective radiative properties of bounded cascade nonabsorbing clouds: Definition of the equivalent homogeneous cloud approximation, Journal of Geophysical Research: Atmospheres, 105, 20 617–20 633, 2000.

Tokui, S., Oono, K., Hido, S., and Clayton, J.: Chainer: a Next-Generation Open Source Framework for Deep Learning, in: Proceedings of Workshop on Machine Learning Systems (LearningSys) in The Twenty-ninth Annual Conference on Neural Information Processing Systems (NIPS), http://learningsys.org/papers/LearningSys_2015_paper_33.pdf, 2015.

Várnai, T. and Davies, R.: Effects of cloud heterogeneities on shortwave radiation: Comparison of cloud-top variability and internal heterogeneity, Journal of the Atmospheric Sciences, 56, 4206–4224, 1999.

Várnai, T. and Marshak, A.: Observations of three-dimensional radiative effects that influence MODIS cloud optical thickness retrievals, Journal of the Atmospheric Sciences, 59, 1607–1618, 2002.

Zhang, Z. and Platnick, S.: An assessment of differences between cloud effective particle radius retrievals for marine water clouds from three MODIS spectral bands, Journal of Geophysical Research: Atmospheres, 116, https://doi.org/10.1029/2011JD016216, 2011.

Zhang, Z., Ackerman, A. S., Feingold, G., Platnick, S., Pincus, R., and Xue, H.: Effects of cloud horizontal inhomogeneity and drizzle on remote sensing of cloud droplet effective radius: Case studies based on large-eddy simulations, Journal of Geophysical Research: Atmospheres, 117, https://doi.org/10.1029/2012JD017655, 2012.

Zhang, Z., Werner, F., Cho, H.-M., Wind, G., Platnick, S., Ackerman, A., Di Girolamo, L., Marshak, A., and Meyer, K.: A framework based on 2-D Taylor expansion for quantifying the impacts of subpixel reflectance variance and covariance on cloud optical thickness and effective radius retrievals based on the bispectral method, Journal of Geophysical Research: Atmospheres, 121, 7007–7025, 2016.

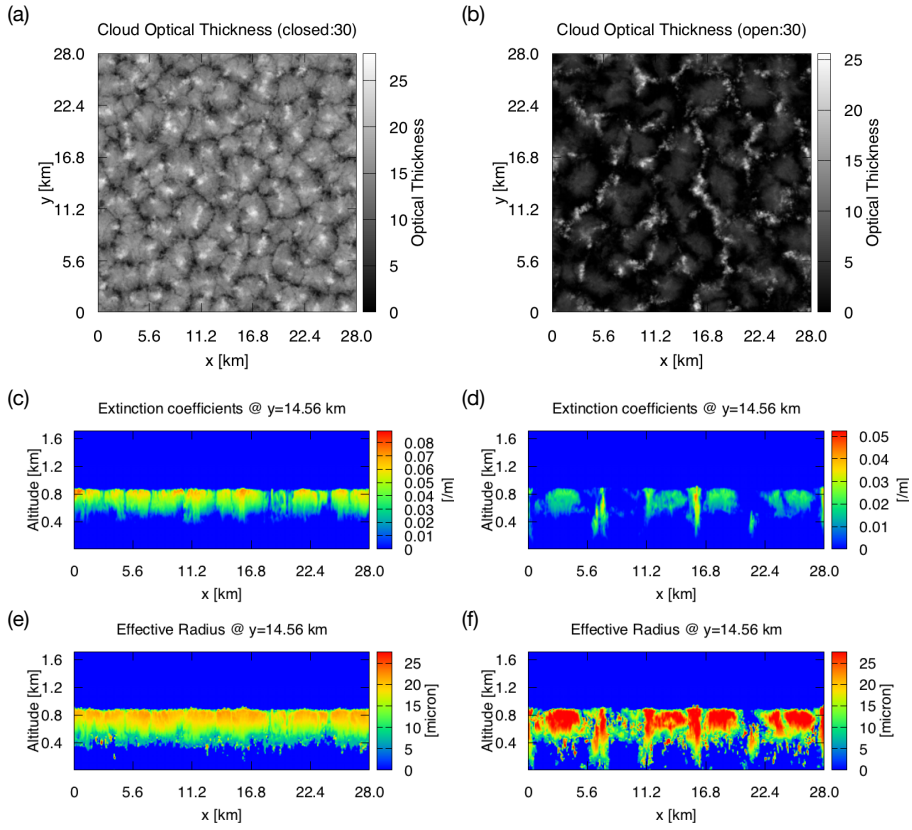

Figure 1. Examples of cloud properties in (a,c,e) closed-cell and (b,d,f) open-cell cases, taken from the 30th timestep of SCALE-LES simulation data. (a,b) Horizontal distributions of COT, (c,d) vertical cross sections of extinction coefficients, and (e,f) vertical cross sections of CDER.

Table 2. The relative RMSE of estimated COT and CDER by the IPA retrieval and DNNs for each cases of cloud field.

Retrieval method	open-cell, COT	closed-cell, COT	open-cell, CDER	closed-cell, CDER
IPA retrieval, SZA: 20°	30.7%	16.0%	38.3%	51.2%
DNN-2r, SZA: 20°	24.3%	24.6%	8.1%	8.9%
DNN-4w, SZA: 20°	21.6%	23.4%	5.5%	6.7%
IPA retrieval, SZA: 60°	74.8%	50.9%	51.1%	55.2%
DNN-2r, SZA: 60°	33.2%	37.4%	7.6%	8.4%
DNN-4w, SZA: 60°	26.6%	18.7%	6.5%	7.3%

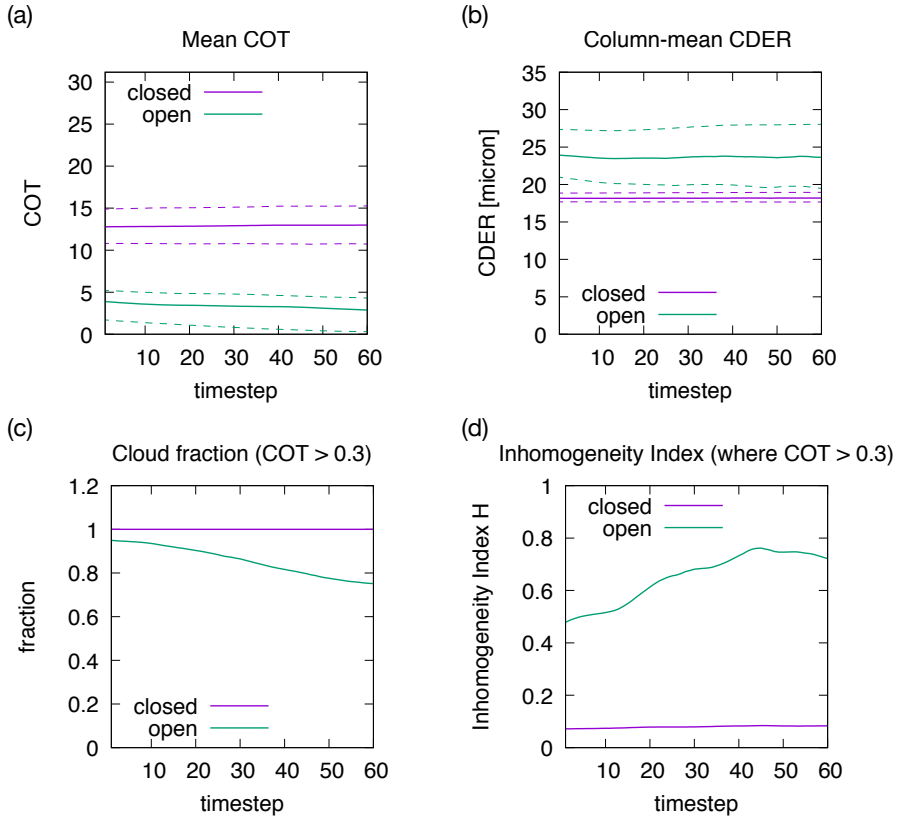

Figure 2. Temporal variations of (a) COT, (b) column-mean CDER, (c) cloud fraction, and (d) inhomogeneity index H. Solid lines shows mean values, and dashed lines show the 25th and 75th percentiles. The cloud fraction and H are computed for pixels with COT > 0.3. A time step corresponds to one minute.

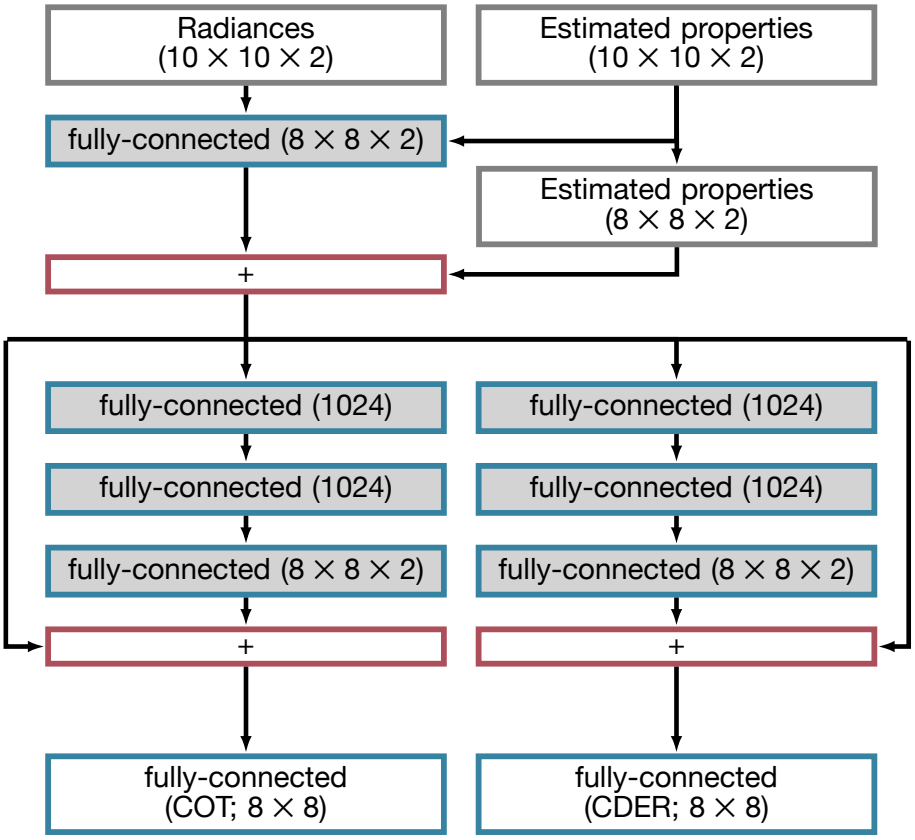

Figure 3. Structure of the DNN-2r network. Blue rectangles denote fully-connected layers, and a red rectangle denotes the addition of two vectors. A gray background indicates that the layer use the activation function. The numbers of units in each layer are shown in parentheses.

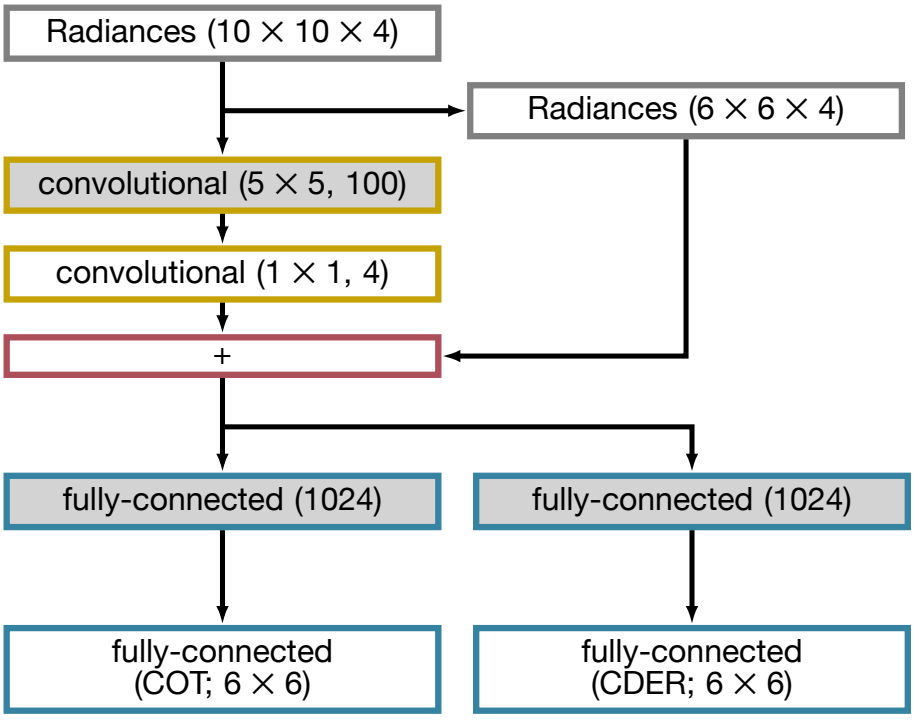

Figure 4. The same as Fig. 3 but for the DNN-4w network. Yellow rectangles denote the convolutional layers, for which the numbers in parentheses denote the filter size and the number of output channels. The number of filters is determined by multiplying the numbers of input and output channels.

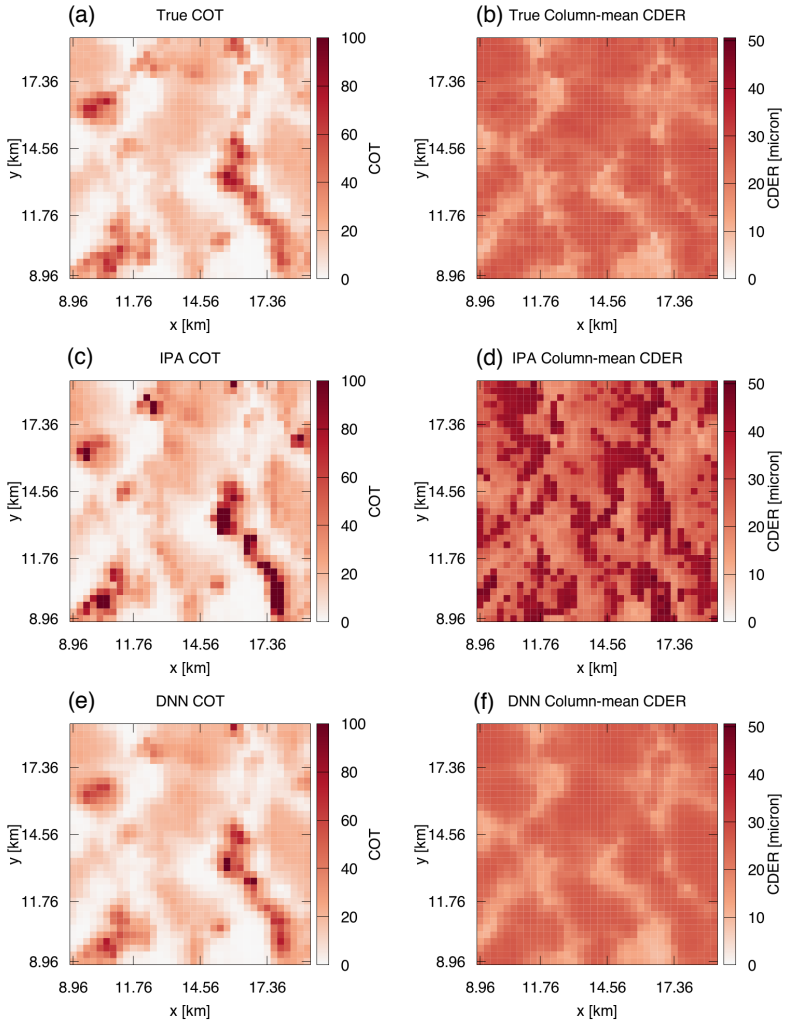

Figure 5. Examples of estimated (a,c,e) COT and (b,d,f) CDER of IPA and DNN retrievals for a view zenith angle of 0°. The sun is located on the left-hand side with an SZA of 60°. (a,b) True (reference) values of COT and CDER, (c,d) IPA retrievals, and (e,f) DNN-4w retrievals.

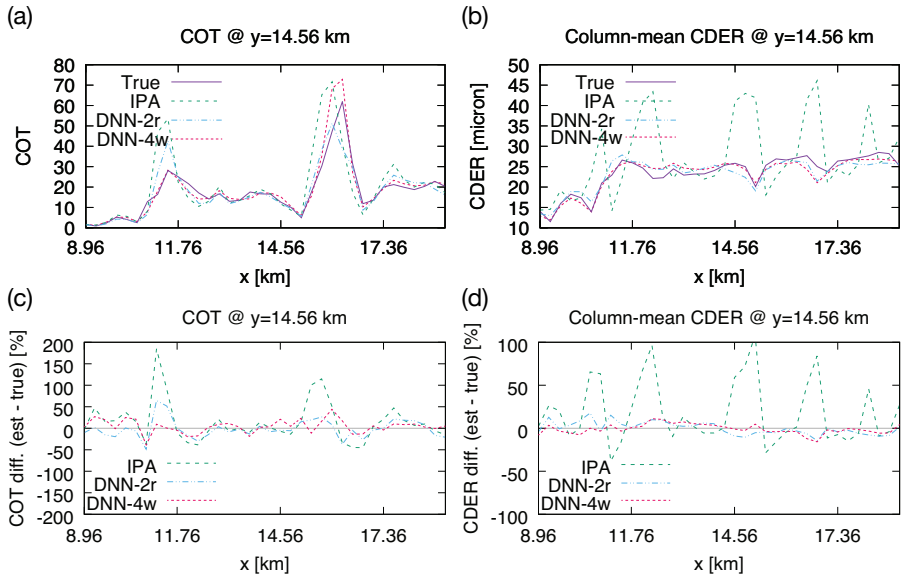

Figure 6. Examples of horizontal distribution of estimated (a) COT and (b) CDER by the IPA retrieval and DNNs for a view zenith angle of 0° at $y = 14.56$ km in Fig. 5. The relative error of the estimated (c) COT and (d) CDER. The sun is located on the left-hand side with an SZA of 60°.

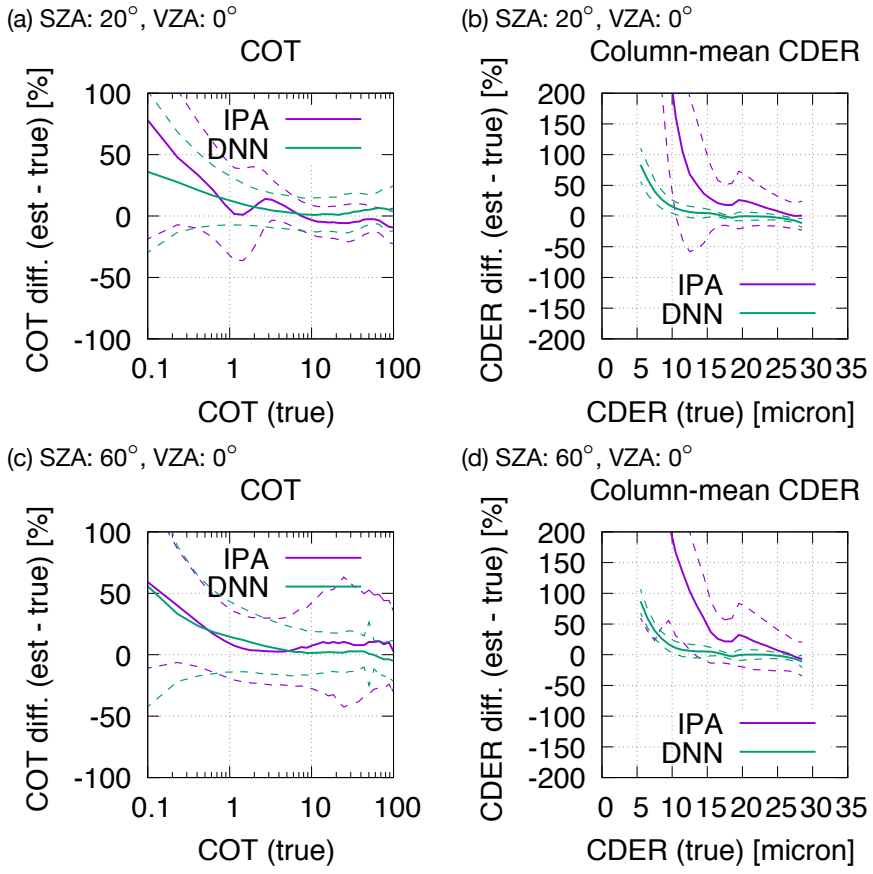

Figure 7. Comparison of retrieval errors of DNN-4w and IPA retrieval for SZAs of (a,b) 20° and (c,d) 60°. The horizontal axes show the true values of either COT or column-mean CDER. The vertical axes show the relative error of the estimated cloud property. The solid and dashed lines denote mean errors and means plus/minus standard deviations of error, respectively.

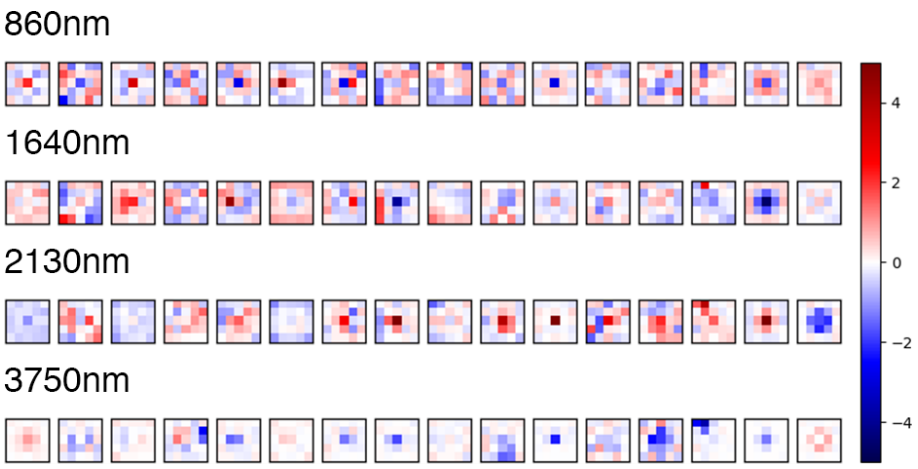

Figure 8. Selected examples of the filter for 5 × 5 pixels at each wavelength in the first convolutional layer in DNN-4w. Only 16 of 100 filters for each wavelength are shown here. The color shade denotes the filter weight. The sunlight is from the left.

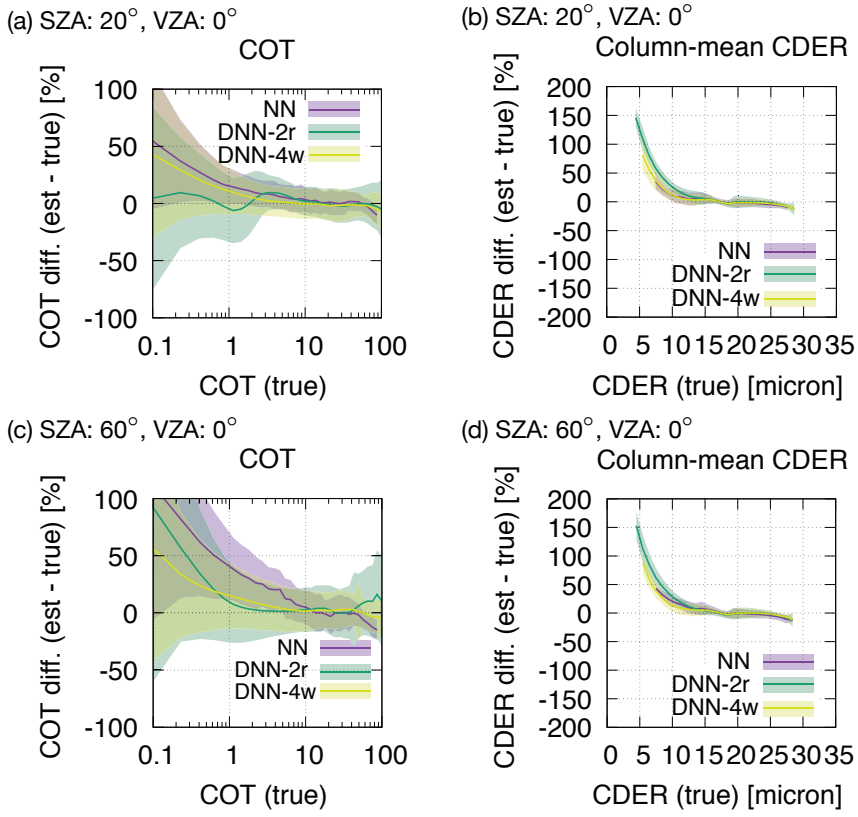

Figure 9. The same as Fig. 7 but for an NN and our DNNs. The solid lines show mean errors, and the shades denote regions of means plus/minus standard deviations.