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. In this study, we present two retrieval methods based on deep learning. We use deep neural networks (DNNs) to retrieve multipixel estimates of cloud optical thickness and column-mean cloud droplet effective radius simultaneously from multispectral, multipixel radiances. Cloud field data are obtained from large-eddy simulations, and a 3D radiative
5   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 DNN-based retrieval 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 3D radiative transfer effects that would otherwise affect the radiance values.

1 Introduction

10   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

15   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, homogeneous cloud for each pixel of the satellite image, whereas 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,

20   as well as to the horizontal radiative transport that occurs in an inhomogeneous cloud field. 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 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. For finer-resolution imagery, by contrast, retrieval errors from inter-pixel horizontal radiative transport become more important. The radiance observed at each

5 pixel is determined by the spatial arrangement of cloud water in the target pixel and its neighbors. 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 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 3D radiative effects differ for COT and CDER, Iwabuchi and Hayasaka (2003) had to construct different

10 sets of numerous fitting coefficients for COT and CDER, which was an obstacle to generalizing their algorithm.

To consider adjacency effects in a generalized manner, neural networks (NNs) (also known as multilayer perceptrons) are useful, as they have been used in 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 3D radiative effects in the retrieval of cloud properties, because doing so creates a more complex problem.

15 Some studies have already proposed such applications to the problem of 3D clouds. Faure et al. (2001, 2002) showed 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 radiance data at wavelengths of 0.64, 1.6, 2.2, and 3.7 μm and a horizontal resolution of 0.8 km × 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 NNs to retrieve cloud properties

20 (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 × 0.25 km at wavelengths of 0.544, 1.6, and 2.15 μm and 1 km × 1 km at wavelengths of 0.544, 1.6, 2.15, 3.65, and 10.8 μm.

More recently, the deep learning (a kind of machine-learning techniques), which uses deep neural networks (DNNs), has

25 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 to more complex problems. A DNN is more suitable for approximating complex nonlinear functions of many variables because the ability to approximate a function is generally improved by using a deeper NN. Recent advances in computer technology, such as multicore central processing units (CPUs)

30 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, 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 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 CDER is calculated as

$$r_e = \frac{1}{\chi} \left(\frac{3}{4\pi} \frac{\text{LWC}}{\rho_b N} \right)^{\frac{1}{3}},$$ (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 LWC and number density are used to calculate the column-mean effective radius R_e, which is defined as

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

where $\langle . \rangle$ denotes the column mean. 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 will be discussed in Section 4.

Figure 2 shows temporal variations of (a) the domain-mean COT, (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

$$H = \frac{\sigma_\tau^2}{\overline{\tau}^2},$$ (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.

5   ## 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

10   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

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

15   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.

20   # 3 Method

3.1 Design and configuration of DNNs

Each layer in the DNNs consists of multiple network units, each of which receives input signals from the previous layer, computes a weighted sum and add a bias, as follows:

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

25   where x'_k is the kth input signal, w_k is the corresponding weight, and b is the bias. The weights and bias are determined in 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 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 in recent DNN applications.

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 compared to previous studies. Larger input and output vectors allow
5   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 and retrieves cloud properties directly from the radiances at 0.86, 1.64, 2.13, and 3.75 μm.

10   The DNN-2r network is designed to correct the IPA retrieval of COT and CDER that originated from multispectral radiances. The elements of the DNN-2r input vector are the radiances at wavelengths of 0.86 and 2.13 μm, and the COT and CDER estimated by the IPA retrieval for 10×10 pixels at 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 DNN-2r network consists of several
15   fully connected layers in which each unit is connected with all units in the previous layer. 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 trained to predict the correction terms that are added to the data from the shortcut path. Such shortcuts facilitate machine learning, even in cases with many NN layers. In this way, the DNN-2r network can be considered a way to correct the IPA retrievals.

20   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 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 shortcuts, the DNN-4w network has two convolutional layers that consist of units that compute the convolutions. In the first convolutional layer, convolutions operate
25   on 5×5 pixels surrounding the center pixel, with 100 different profiles of filter weight 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

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

30   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 concerned, we expect that using a convolutional layer will allow

5   the DNN to learn patterns that characterize the 3D radiative effects among the target pixel and those adjacent to it. We expect the DNN-4w network firstly to correct the 3D radiative transfer effects and then to transform the signals to COT and CDER with the possibility of additional corrections of the 3D effect in this latter part.

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

10 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 on. However, in some cases,

15 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

20 performance and is widely used in many applications.

3.2 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-

25 LES cloud-field data, $1,977,440$ samples 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.

30 To construct an efficient DNN, it is worth investigating the relationships between the input and output variables. In the DNN preprocessing, the cloud properties are transformed using

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

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

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

5   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 test datasets include $10,000$
10   samples.

4   Results

In this section, we illustrate the ability of DNNs to retrieve cloud properties and we compare this with the corresponding abilities of existing methods. Values of COT and CDER are retrieved from test datasets by using DNNs and IPA. 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,
15   COT and CDER are estimated from look-up tables 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 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 the associated range defined by the look-up tables, the COT/CDER value is forced to be the lower or upper limit, as appropriate.

20   ### 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 $60°$. Cross sections taken at $y = 14.56$ km are shown in Fig. 6 with additional DNN-2r retrieval results. The sunny (left-hand) side of the COT fluctuation peak is directly illuminated by the Sun. For pixels on that side, the radiances calculated by 3D radiative transfer are brightened (illuminating effect), which results in the overestimation (resp. underestimation) of IPA retrievals of COT (resp.
25   CDER). For pixels on the opposite (right-hand) side, the radiances are darkened (shadowing effect) and IPA retrieval of COT (resp. CDER) is underestimated (resp. overestimated). These illuminating and shadowing effects have considerable influence on the IPA retrieval. A phase lag appears in the IPA-retrieved horizontal COT distribution because of this illuminating and shadowing; the IPA error in the COT is particularly large for optically thick parts. In contrast, the DNN retrieves COT values that are close to the true values assumed in this test, successfully correcting the phase lag. However, minor errors are still
30   present in the DNN-retrieved COT.

As for the retrieved values of CDER, the DNN ones are obviously better than the IPA 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.

5 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 and DNN-4w relative errors are plotted against the true COT and CDER

10 values. The IPA-retrieved COT error and its standard deviation are particularly large for a SZA of $60°$, at which the 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. Compared to the IPA, 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

15 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 IPA retrieval.

The DNN also yields better CDER retrievals than does the IPA, 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,

20 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 column-mean CDER. However, overestimation of CDER in the IPA retrieval is

25 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. 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.

30 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 $60°$. Only 16 of 100 filters are shown here, and each filter weight can be either positive or negative. The patterns in some filters 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 have higher weights in pixels along

35 the solar azimuth direction, which suggest 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 3D radiative effects to recover the information about local cloud properties. However, it is difficult at present 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 Faure et al. (2001). Originally, this NN had two hidden layers with 10 units each. 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 Faure et al. (2001), and the outputs are COT and CDER.

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 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 SCALE-LES cloud-field data whose horizontal resolution was 280 m. Both DNNs outperformed IPA-based retrieval in relation to 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 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 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.

5   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). In addition, we considered cloud thickness of only less than 0.9 km, as shown in Fig. 1. Therefore, it would 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

10   remote sensing is expected to become more common. However, using DNNs with actual satellite data will require training 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.

Acknowledgements. SCALE-LES was developed by the SCALE interdisciplinary team at the RIKEN Advanced Institute for Computational

15   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).

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, doi:10.1029/2003JD004186, 2004.

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, doi: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, doi: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, doi: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, doi: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, doi: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, doi: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, doi:10.1029/2008GL037124, 2009.

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

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, doi: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, doi: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, doi:10.1016/0022-4073(86)90088-9, 1986.

5   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, doi: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, doi: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,

10   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, doi: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, doi:10.1029/2000JD900333, 2000.

15   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, doi:10.2151/sola.2014-003, 2014.

20   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, doi: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, doi:10.1016/j.jqsrt.2008.07.013, 2008.

25   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.

30   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.

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, doi:10.1029/2011JD016216, 2011.

35   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, doi:10.1029/2012JD017655, 2012.
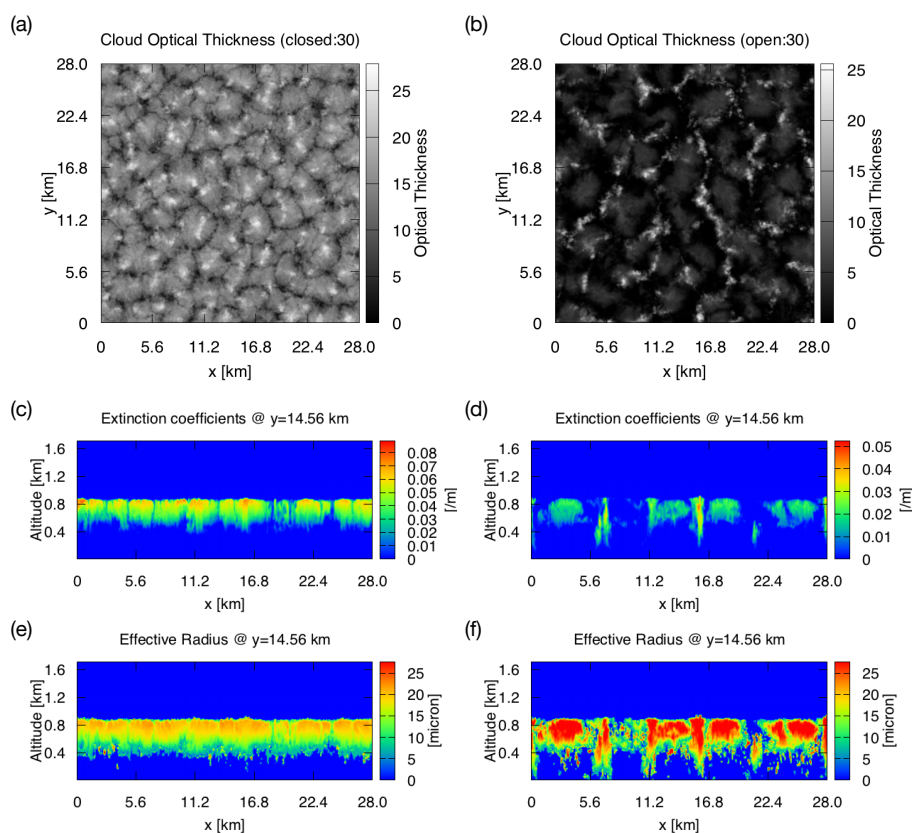
Figure 1. Examples of cloud properties in (a,c,e) closed-cell and (b,d,f) open-cell cases, taken from the 30th timestep of SCALE-LES simulation data. (a,b) Horizontal distributions of COT, (c,d) vertical cross sections of extinction coefficients, and (e,f) vertical cross sections of CDER.
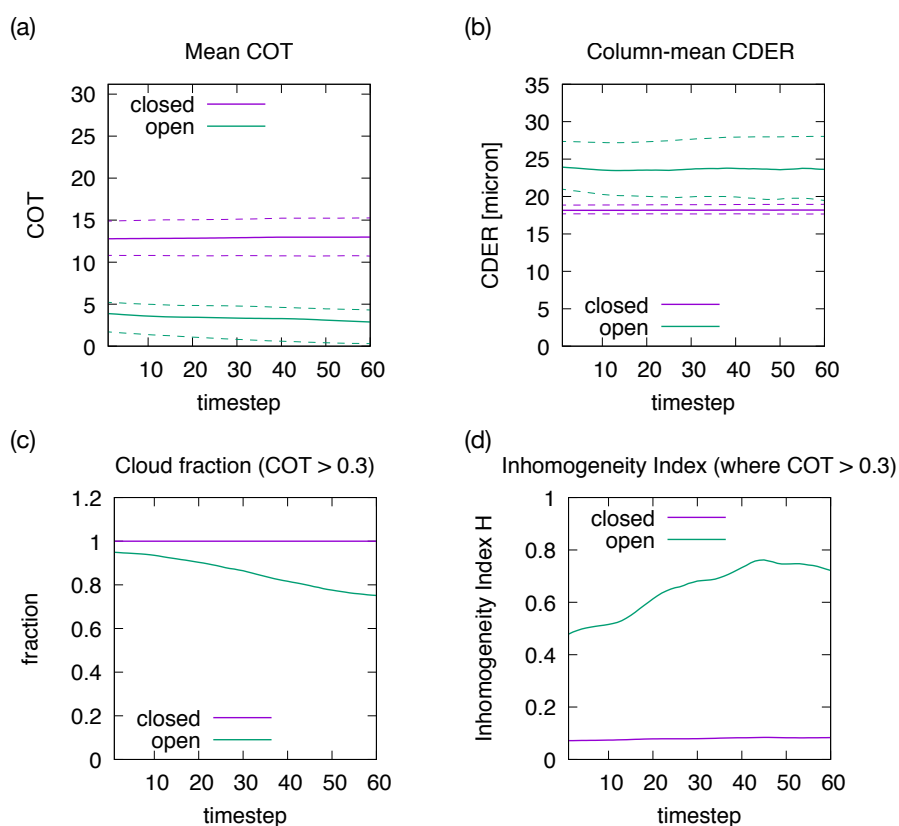
Figure 2. Temporal variations of (a) COT, (b) column-mean CDER, (c) cloud fraction, and (d) inhomogeneity index H. Solid lines shows mean values, and dashed lines show the 25th and 75th percentiles. The cloud fraction and H are computed for pixels with COT > 0.3.
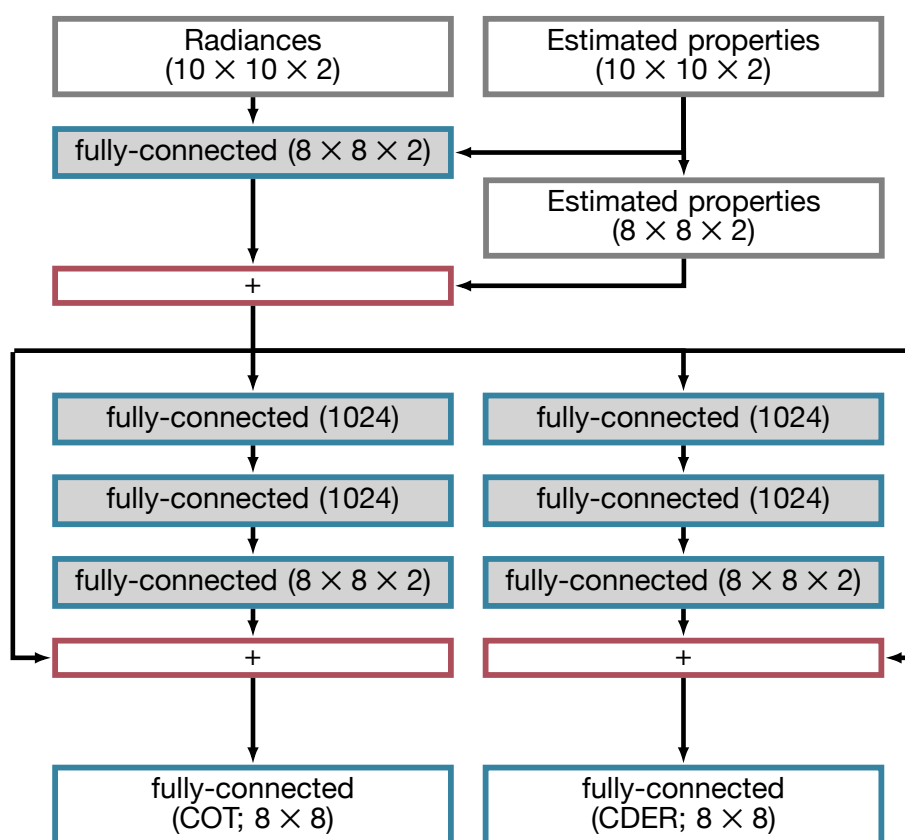
Figure 3. Structure of the DNN-2r network. Blue rectangles denote fully-connected layers, and a red rectangle denotes the addition of two vectors. A gray background indicates that the layer use the activation function. The numbers of units in each layer are shown in parentheses.
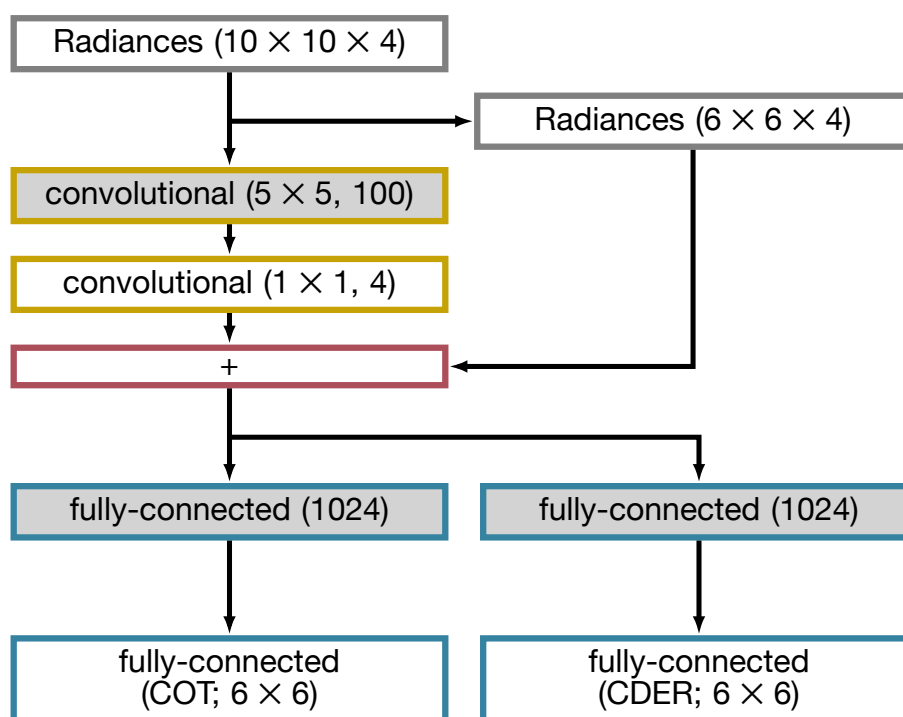
Figure 4. The same as Fig. 3 but for the DNN-4w network. Yellow rectangles denote the convolutional layers, for which the numbers in parentheses denote the filter size and the number of output channels. The number of filters is determined by multiplying the numbers of input and output channels.
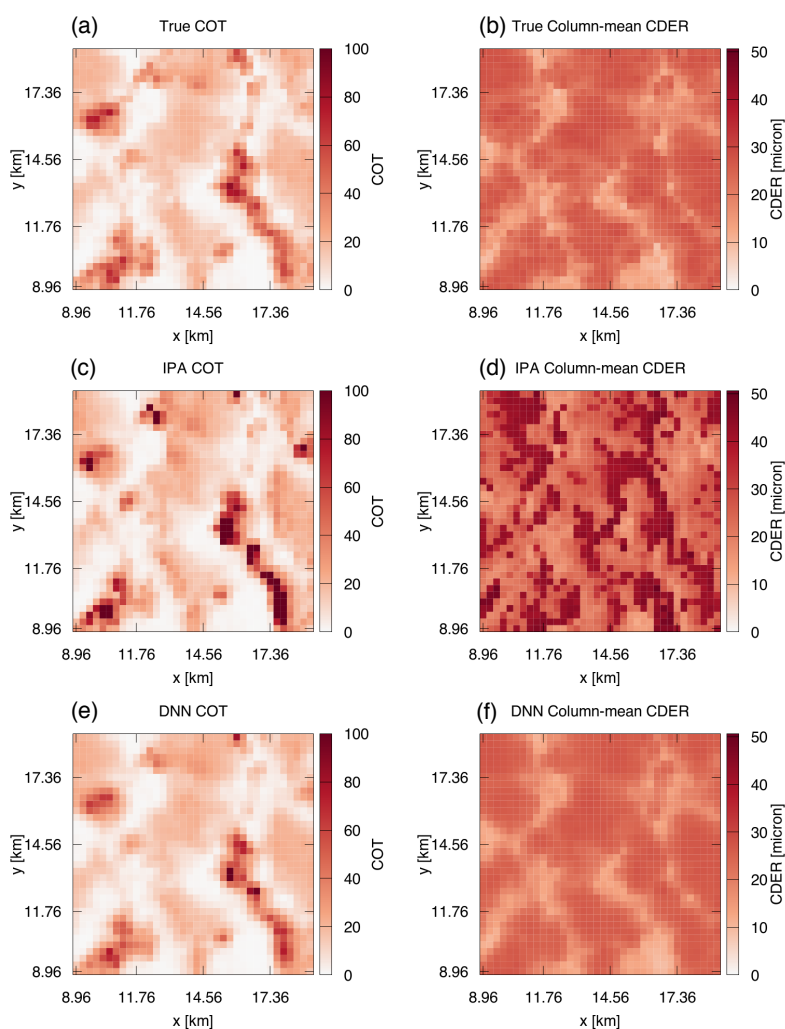
Figure 5. Examples of estimated (a,c,e) COT and (b,d,f) CDER of IPA and DNN retrievals for a view zenith angle of $0°$. The sun is located on the left-hand side with an SZA of $60°$. (a,b) True (reference) values of COT and CDER, (c,d) IPA retrievals, and (e,f) DNN-4w retrievals.
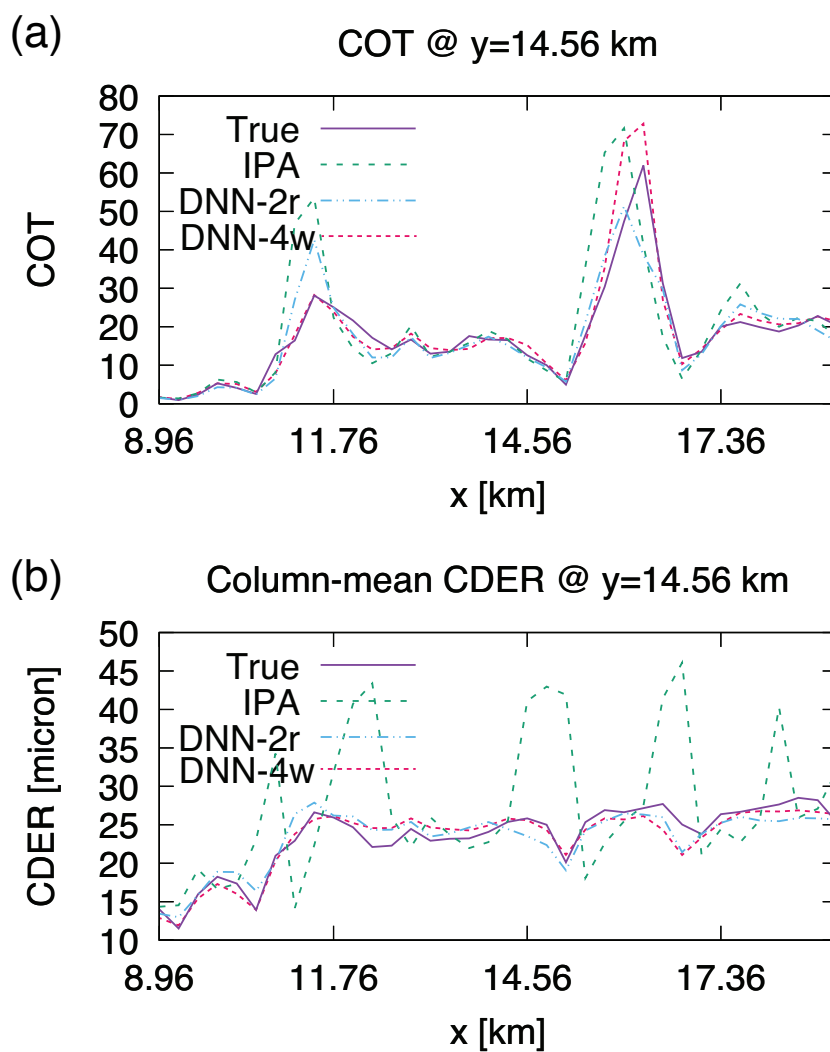
Figure 6. Examples of horizontal distribution of estimated (a) COT and (b) CDER by the IPA and DNNs at $y = 14.56$ km in Fig. 5. The sun is located on the left-hand side with an SZA of $60°$.
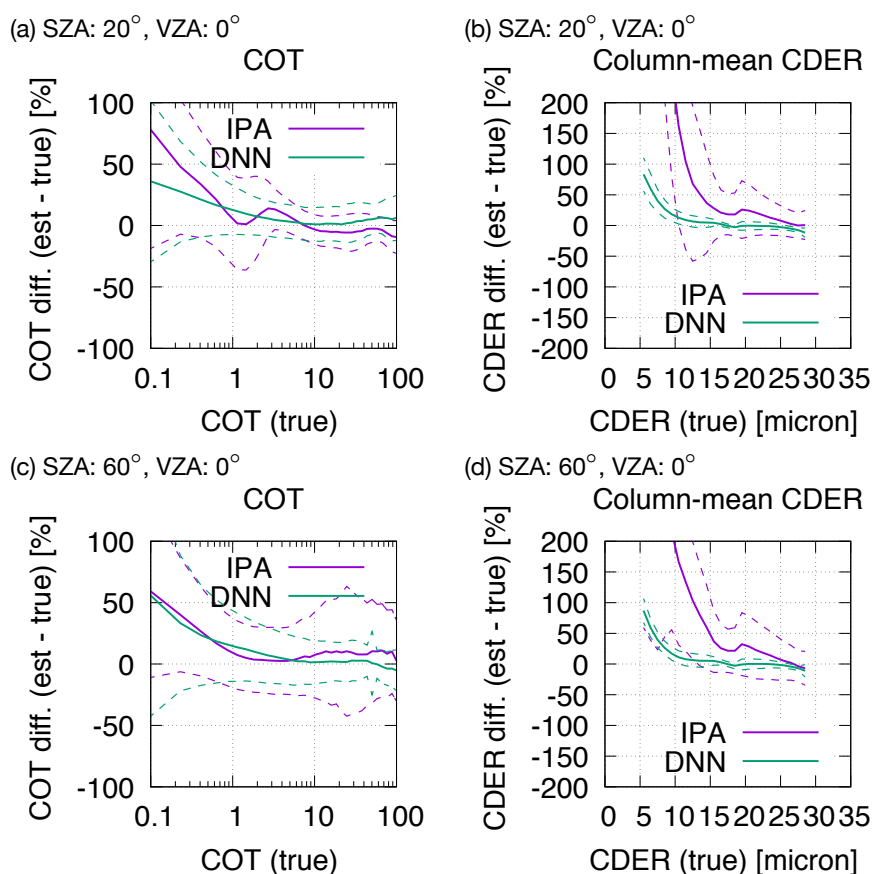
Figure 7. Comparison of retrieval errors of DNN-4w and IPA for SZAs of (a,b) 20° and (c,d) 60°. The horizontal axes show the true values of either COT or column-mean CDER. The vertical axes show the relative error of the estimated cloud property. The solid and dashed lines denote mean errors and means plus/minus standard deviations of error, respectively.
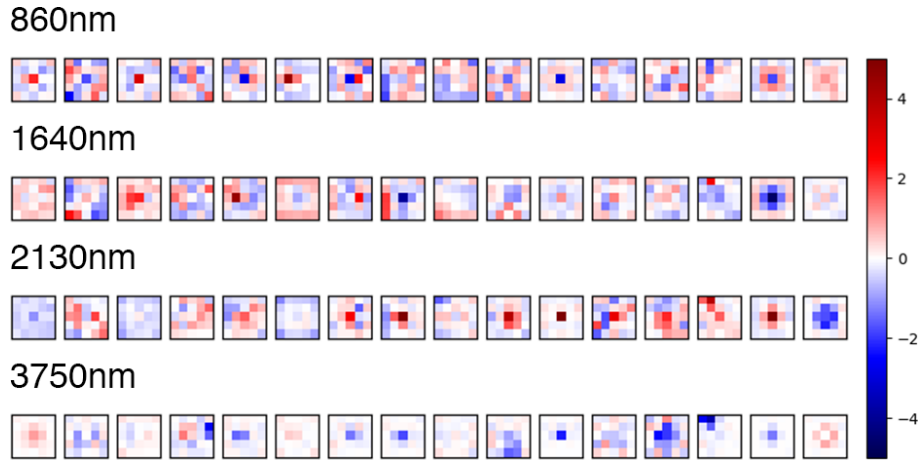
Figure 8. Selected examples of the filter for 5×5 pixels at each wavelength in the first convolutional layer in DNN-4w. Only 16 of 100 filters for each wavelength are shown here. The color shade denotes the filter weight. The sunlight is from the left.
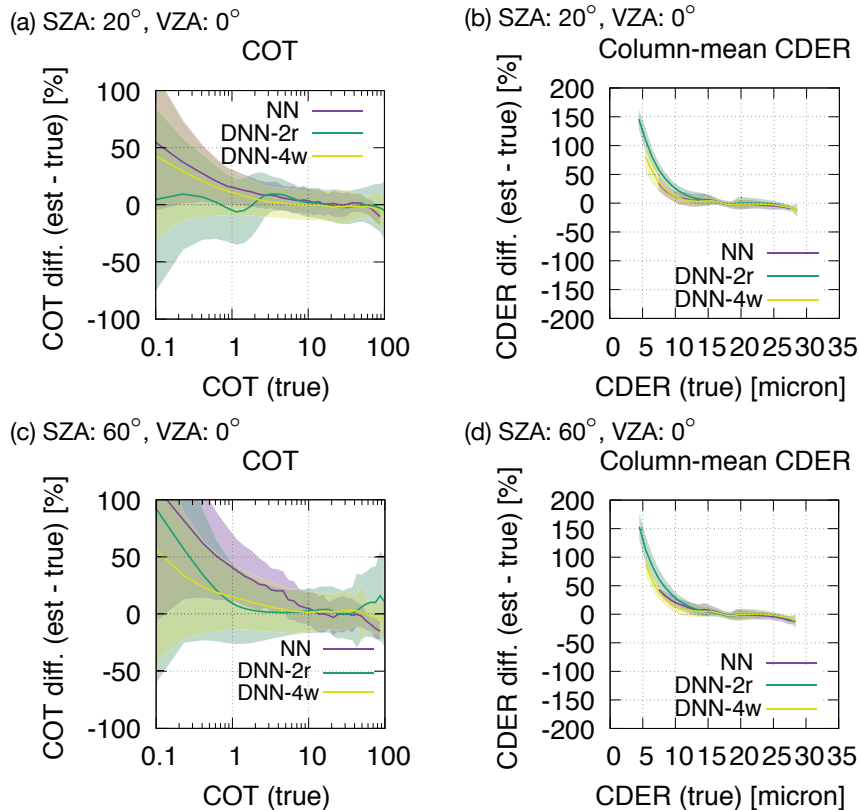
Figure 9. The same as Fig. 7 but for an NN and our DNNs. The solid lines show mean errors, and the shades denote regions of means plus/minus standard deviations.