A Consistent Representation of Cloud Overlap and Cloud Subgrid Vertical Heterogeneity

Raphaël. Lebrun, Jean-Louis Dufresne, Najda Villefranque

JAMES | Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1029/2022MS003592

A Consistent Representation of Cloud Overlap and Cloud Subgrid Vertical Heterogeneity

Raphaël Lebrun[1], Jean-Louis Dufresne[1], and Najda Villefranque[1]

[1]Laboratoire de Météorologie Dynamique/IPSL, CNRS, Sorbonne Université, École Normale Supérieure, PSL Research University, École Polytechnique, Paris, France

Key Points:

- We extend the use of exponential-random overlap (ERO) to represent both overlap and subgrid variability
- The commonly used maximum-random overlap hypothesis can generate cloud covers half too small
- The decorrelation lengths used with ERO are highly dependent on the vertical resolutions of models and observations

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

R. Lebrun,
raphael.lebrun@lmd.ipsl.fr

Citation:

Lebrun, R., Dufresne, J.-L., & Villefranque, N. (2023). A consistent representation of cloud overlap and cloud subgrid vertical heterogeneity. *Journal of Advances in Modeling Earth Systems*, *15*, e2022MS003592. https://doi.org/10.1029/2022MS003592

Received 22 DEC 2022
Accepted 6 APR 2023

Abstract Many global climate models underestimate the cloud cover and overestimate the cloud albedo, especially for low-level clouds. We determine how a correct representation of the vertical structure of clouds can fix part of this bias. We use the 1D McICA framework and focus on low-level clouds. Using Large Eddy Simulations results as reference, we propose a method based on exponential-random overlap that represents the cloud overlap between layers and the subgrid cloud properties over several vertical scales, with a single value of the overlap parameter. Starting from a coarse vertical grid, representative of atmospheric models, this algorithm is used to generate the vertical profile of the cloud fraction with a finer vertical resolution, or to generate it on the coarse grid but with subgrid heterogeneity and cloud overlap that ensures a correct cloud cover. Doing so we find decorrelation lengths are dependent on the vertical resolution, except if the vertical subgrid heterogeneity and interlayer overlap are taken into account coherently. We confirm that the frequently used maximum-random overlap leads to a significant error by underestimating the low-level cloud cover with a relative error of about 50%, that can lead to an error of SW cloud albedo as big as 70%. Not taking into account the subgrid vertical heterogeneity of clouds can cause a relative error of 20% in brightness, assuming the cloud cover is correct.

Plain Language Summary Low-level clouds are the main source of spread in model estimates of climate sensitivity, but climate models resolutions do not allow them to explicitly resolve the geometrical complexity of low-level clouds, which must be parametrized. Most climate models low-level clouds have a cloud cover too small and a cloud albedo too high, which is known as the "too few too bright bias." In this work we determine whether a better representation of the vertical structure of clouds can fix part of this bias. We use high-resolution simulations as references and radiative transfer algorithms to assess the performances of our cloud generation. We use a framework of commonly used overlap assumptions, and we extend it to also represent vertical subgrid variability. When the cloud cover of the scene is known, we show that an appropriate choice of parameters allows a good representation of the vertical structure of clouds and of the cloud albedo. We find the parameters used to model the overlap are highly dependent on the model vertical resolution, and present a way to overcome this dependency when both subgrid scale and interlayer overlap are taken into account consistently.

1. Introduction

The size and the spatial structure of clouds vary by several orders of magnitude (Koren et al., 2008). The size of the horizontal meshes of global and regional atmospheric circulation models typically range from a few kilometers to a few hundred kilometers, and their vertical resolution in the troposphere is typically from ten to several hundred meters. Thus the geometric representation of clouds in these models at scales smaller than those of the mesh sizes must be parametrized, especially to compute the radiative effect of clouds that is of crucial importance for the climate.

The cloud geometry in a model is generally simply described by a horizontal fraction of the layer being cloudy, the remaining part being clear. In the cloudy part, the in-cloud liquid or solid amount of water is often assumed to be uniform, although some improved representations have been proposed (Hogan & Shonk, 2013; Räisänen et al., 2004). The cloud cover CC (defined as the fraction of the area of the entire scene i.e., covered by clouds) and the mean optical depth of the cloudy region (that we will call the cloud optical depth) depend on how the cloud fractions overlap in the vertical. If they overlap maximally, the cloud cover will be minimum and the cloud optical depth maximum, if they overlap minimally, the cloud cover will be maximum and the cloud optical depth

minimal, and if the overlap is random, the cloud cover as well as the cloud optical depth will be somewhere in between.

How the cloud fraction of each atmospheric layer overlap with other layers has been widely studied (Barker et al., 1999; Geleyn & Hollingsworth, 1979; Jakob & Klein, 1999). Many recent studies use an exponential-random scheme approach where the probability of two layers overlapping decreases exponentially with the distance between them (Bergman & Rasch, 2002; Hogan & Illingworth, 2000; Oreopoulos, Cho, & Lee, 2022; Shonk & Hogan, 2010; Tompkins & Di Giuseppe, 2007). The corresponding decorrelation length scale has been estimated from satellite radar and lidar observations (Jing et al., 2016; Li et al., 2015), in-situ observations (Mace & Benson-Troth, 2002), and high resolution model simulations (Neggers et al., 2011). Studies have shown that the decorrelation length can be parametrized as a function of the horizontal wind profile of the column (Di Giuseppe & Tompkins, 2015; Pincus et al., 2005; Sulak et al., 2020).

The vertical subgrid heterogeneity of the cloud fraction has been less investigated. Atmospheric model cloud schemes calculate the cloud fraction as the fraction of the volume of a layer that contains clouds, C_v, but radiation is primarily sensitive to the areal cloud fraction C_a, which is the fraction of the area of the layer covered by clouds. We will keep those definitions for the rest of the study. Often implicitly, these two fractions are assumed to be equal, that is, the clouds are assumed to be homogeneous on the vertical in each layer. This can seem logical on the first order given the area/depth ratio of the layers, however, recent studies show that this may introduce significant biases, as the distribution of cloud water can be vertically heterogeneous in layers as thin as 100 m (Brooks et al., 2005; Jouhaud et al., 2018), and that C_a is typically greater than C_v by about 30% (Neggers et al., 2011). A direct consequence of not taking into account this difference is that, for a given volume cloud fraction and mean liquid water content, the areal cloud fraction of the clouds is too small and the water content per unit of cloud fraction (and therefore the cloud albedo) too large.

Considering these results, we address the following questions: can we use ERO to statistically represent the vertical structure of cloud scenes, only using a small number of aggregated quantities, to simulate precisely radiative fluxes? How does this representation depend on the vertical resolution? What is the radiative error that is induced when the subgrid vertical structure of the clouds is not explicitely resolved and hence not seen by radiation? To answer them we propose an overlap model that ensures consistency between the overlap between cloudy layers and the representation of subgrid vertical heterogeneity. Indeed, we contend that both are intended to represent the same characteristic of clouds, their vertical distribution, and that the distinction between the two depends on the vertical resolution of the atmospheric model, which can vary. The cloud generator and formalism we present are able to take into account both the vertical resolution and the subgrid variability of the cloud fraction. As done in McICA radiative transfer calculations (Pincus et al., 2003), we neglect the 3D effects and keep the classical plane parallel assumption (each vertical profile consists of a stack of horizontally infinite and homogeneous slabs) in our 1D approach. Assuming that the volume cloud fraction and water content are known on a coarse vertical grid consisting in a single column, typical of an atmospheric model, we developed an algorithm to generate an ensemble of subcolumns to statistically represent the heterogeneity of clouds. As cumulus clouds are prone to the "too-few too-bright" bias in global models (Konsta et al., 2022), as well as marine boundary layer clouds being the main source of uncertainty in tropical cloud feedbacks simulated by general circulation models (Bony & Dufresne, 2005), we focus our study specifically on cumulus cloud scenes, and use Large Eddy Simulations (LES) as references.

The manuscript is organized as follows: in Section 2, we consider the ERO as a Markov process and show its ability to represent the vertical distribution of the cloud fraction over a wide range of scales that includes both the subgrid scale and the overlap between layers. In Section 3 we study cloud scenes with known cloud covers, and compute the overlap parameters and decorrelation lengths that should be used with ERO on finer grids to reproduce those cloud covers, and doing so we assess the radiative impact of ERO on the SW cloud albedo of the generated subcolumns. We also study the effects of different simplifying assumptions. Section 4 focuses on reproducing those results directly on the coarse grid, taking into account both the interlayer overlap and the subgrid scale, assuming again that the cloud cover is known. The implication for cloud parameterization in atmospheric models and for how to estimate the decorrelation lengths are presented in Section 5.

2. Statistical Representation of the Cloud Fraction Vertical Distribution

The model explored here is the so-called ERO model of Hogan and Illingworth (2000). We will only look at single-layer cumulus cloud fields so the "random" part of the model, which concerns cloudy layers that are

separated by clear layers, will not be studied. The "exponential" part of the model states that the cloud fraction of two cloudy layers of cloud fractions C_1 and C_2 is:

$$C_{1,2} = \alpha C_{1,2,\max} + (1-\alpha) C_{1,2,rand}$$

where $C_{1,2,max}$ is the combined cloud fraction of the two layers in case they overlap maximally:

$$C_{1,2,\max} = \max(C_1, C_2)$$

and $C_{1,2,rand}$ is the combined cloud fraction of the two layers in case they overlap randomly:

$$C_{1,2,rand} = C_1 + C_2 - C_1 C_2$$

In this model, "exponential" refers to the fact that α can be parametrized with an exponential function (see further). This model has been used in two different manners in radiative transfer parameterizations: either in a deterministic way, to compute the overlap matrix that is used to distribute downwelling and upwelling fluxes from clear and cloudy regions of a layer into clear and cloudy regions of an adjacent layer (TripleClouds, Shonk & Hogan, 2008), or in a probabilistic manner, to generate a sample of vertical profiles that preserve, when averaged, the principal characteristics of the cloud scene (the areal cloud fraction and the liquid water content in each layer), and upon which radiative transfer is simulated under the plane-parallel homogeneous assumption (McICA, Pincus et al., 2003). In this study, we generate samples of vertical profiles, as done by other cloud generators. The main difference with other cloud generators where the profiles are generated on the vertical grid of the host model, we aim at generating profiles at any vertical resolution, including finer vertical resolutions.

We consider that a single atmospheric column consists of $\mathcal{N}$ vertical cloudy layers. From this column we assume the volume cloud fraction of each layer, $(C_v)_{k=1\ldots\mathcal{N}}$, is known. We consider the ERO model as a Markovian process and deduce the relationship between the overlap parameter α and the cloud cover CC of the whole scene. We then use the same result to deal with subgrid vertical heterogeneity.

2.1. ERO as a Markovian Process: A Sequence of Conditional Probabilities

Using a certain overlap scheme in an atmospheric column to generate a cloud fraction distribution from top to bottom can be interpreted as a Markovian process as it is a sequence of overlapping or non-overlapping events. It is then possible to compute its outcome as a sequence of conditional probabilities, as done by Bergman and Rasch (2002). To remain as general as possible, we do not specify what type of cloud fraction C is being used in this section, as any profile can be used.

In a single atmospheric column of $\mathcal{N}$ vertical layers, let us consider a 1D subcolumn. We want to articulate how the overlap used for the whole atmospheric column translates to a subcolumn. We use the term cell to refer to the different layers of each subcolumn, and suppose a cell can have only one of two discrete states, cloudy or clear-sky. If $\vec{\mathrm{S}} = (\mathrm{S}_k)_{k=1\ldots\mathcal{N}}$ is the random variable representing the cloud states of the subcolumn, with the random variables $\mathrm{S}_k \in \{0,1\}$ (whether the cell is cloudy or not), and k is the vertical index, with $k=1$ at the top of the column, the probability of a certain state $\vec{\mathrm{S}} = (s_k)_{k=1\ldots\mathcal{N}} \in [0,1]^{\mathcal{N}}$ is given by:

$$P\left(\vec{\mathrm{S}}\right) = \prod_{k=1}^{\mathcal{N}} P(S_k = s_k | S_{k-1} = s_{k-1}) \tag{1}$$

where $S_0 = 0$ (i.e., there is no cloud above the cloud block considered here). We use the classic upper case notation S_k for the random variables and the lower case notations s_k for their realizations.

For any cell k in the subcolumn, the probability to have $s_k = 1$ is the cloud fraction of the layer, meaning $P(s_k = 1) = C_k$. We'll call $P(S_k = s_k | S_{k-1} = s_{k+1})$ a *transition probability*, it is the probability that in a subcolumn, cell k is in the state s_k, knowing the cell $k-1$ is in the state s_{k-1}. Since s_k is either 0 or 1, there are only four possible types of transition between two cells, and being able to compute their probabilities for each layer gives

the probability of any combination of states for the subcolumn. Moreover, for each cell k, two out of the four transition probabilities are dependant, as a cell is either cloudy or clear-sky:

$$\begin{cases} P(\mathrm{S}_k = 0|\mathrm{S}_{k-1} = 1) = 1 - P(\mathrm{S}_k = 1|\mathrm{S}_{k-1} = 1) \\ P(\mathrm{S}_k = 1|\mathrm{S}_{k-1} = 0) = 1 - P(\mathrm{S}_k = 0|\mathrm{S}_{k-1} = 0) \end{cases}. \tag{2}$$

Therefore, it is enough to know for instance the two transition probabilities $P(\mathrm{S}_k = 1|\mathrm{S}_{k-1} = 1)$ and $P(\mathrm{S}_k = 0|\mathrm{S}_{k-1} = 0)$ for each layer k to compute the probability of any given state of overlap for the subcolumn, using Equation 1.

The transition probability $P(\mathrm{S}_k = 1|\mathrm{S}_{k-1} = 1)$ is the probability that both cells of the subcolumn are cloudy, knowing that the cell $k-1$ is already cloudy. By definition, we have:

$$P(\mathrm{S}_k = 1|\mathrm{S}_{k-1} = 1) = \frac{P(\mathrm{S}_k = 1 \cap \mathrm{S}_{k-1} = 1)}{P(\mathrm{S}_{k-1} = 1)} \tag{3}$$

where $(\mathrm{S}_k = 1 \cap \mathrm{S}_{k-1} = 1)$ is the event with both cells cloudy. If we assume an ERO we have:

$$P(\mathrm{S}_k = 1|\mathrm{S}_{k-1} = 1) = \alpha P_{\max}(\mathrm{S}_k = 1|\mathrm{S}_{k-1} = 1) + (1-\alpha)P_{rand}(\mathrm{S}_k = 1|\mathrm{S}_{k-1} = 1) \tag{4}$$

where P_{max} and P_{rand} are the corresponding transition probabilities, in a subcolumn, for maximum overlap and random overlap between two consecutive layers of the atmospheric column. By definition of random overlap the probability of cell k being cloudy is independent of the conditions of cell $k-1$:

$$P_{rand}(\mathrm{S}_k = 1|\mathrm{S}_{k-1} = 1) = P_{rand}(\mathrm{S}_k = 1|\mathrm{S}_{k-1} = 0) = P_{rand}(\mathrm{S}_k = 1) = C_k \tag{5}$$

The transition probability in a subcolumn for maximum overlap can be obtained using Equation 3: if $C_{k-1} < C_k$: $P_{max}(\mathrm{S}_k = 1|\mathrm{S}_{k-1} = 1) = 1$, and on the contrary if $C_{k-1} \geq C_k$: $P_{\max}(\mathrm{S}_k = 1|\mathrm{S}_{k-1} = 1) = \frac{C_k}{C_{k-1}}$. As a result:

$$P_{\max}(\mathrm{S}_k = 1|\mathrm{S}_{k-1} = 1) = \frac{\min(C_{k-1}, C_k)}{C_{k-1}} \tag{6}$$

and (Equation 4) becomes:

$$P(\mathrm{S}_k = 1|\mathrm{S}_{k-1} = 1) = \alpha \frac{\min(C_{k-1}, C_k)}{C_{k-1}} + (1-\alpha)C_k \tag{7}$$

Let us compute $P_{max}(\mathrm{S}_k = 0|\mathrm{S}_{k-1} = 0)$ in the same way, and we get:

$$P_{\max}(\mathrm{S}_k = 0|\mathrm{S}_{k-1} = 0) = \frac{1 - \max(C_{k-1}, C_k)}{1 - C_{k-1}}$$

and therefore:

$$P(\mathrm{S}_k = 0|\mathrm{S}_{k-1} = 0) = \alpha \times \frac{(1 - \max(C_{k-1}, C_k))}{1 - C_{k-1}} + (1-\alpha)(1 - C_k) \tag{8}$$

These equations are applicable only for non overcast cloudy layers (i.e., $C \in]0,1[$), because of the terms $1/C_k$ and $1/(1-C_k)$.

Having computed the transition probabilities between different cloud states of the cells, we can now use them to generate subcolumns. The details of the implementation are presented in Appendix A, along with the main difference with the work of Räisänen et al. (2004), from which our algorithm is very much inspired. Thanks to Equations 2, 7 and 8 we can now compute the different transition probabilities for each layer k, knowing α. Then using Equation 1 we can compute the probability to generate any cloud state combination for a subcolumn, for any ERO parameter $\alpha \in [0, 1]$.

2.2. The Relationship Between the Overlap Parameter α and the Cloud Cover

In a similar fashion as the work done by Barker (2008a, 2008b), we are now going to establish the relationship between the overlap parameter α and the cloud cover CC, assuming ERO.

To obtain the formal expression of the cloud cover from the previous equations, it is easier to compute the probability of having no cloud for a whole subcolumn. Indeed $P_\emptyset$ corresponds to transition probabilities "clear-sky/clear-sky" of the form $P(0|0)$. The probability to generate a fully clear-sky subcolumn can be seen as a first order Markov chain probability and therefore computed as the product of conditional probabilities, as seen in the previous section:

$$P_\emptyset = \prod_{k=1}^{\mathcal{N}} P(\mathrm{S}_k = 0 | \mathrm{S}_{k-1} = 0) \tag{9}$$

Using Equation 8 we get:

$$P_\emptyset\left(\alpha, (C_k)_{1\ldots\mathcal{N}}\right) = \prod_{k=1}^{\mathcal{N}} \left[\frac{\alpha * (1 - \max(C_{k-1}, C_k))}{1 - C_{k-1}} + (1-\alpha)(1 - C_k) \right] \tag{10}$$

Given this equation, if we know the overlap parameter α, the cloud cover is:

$$CC_{ERO} = 1 - P_\emptyset\left(\alpha, (C_k)_{1\ldots\mathcal{N}}\right) \tag{11}$$

On the other hand if the cloud cover CC is known, we can then determine the overlap parameter α that matches the cloud cover CC:

$$\alpha = f_\emptyset^{-1}(1 - CC) \tag{12}$$

where

$$f_\emptyset \quad : \quad \alpha \in [0,1] \rightarrow f_\emptyset(\alpha) = P_\emptyset\left(\alpha, (C_k)_{1\ldots\mathcal{N}}\right)$$

For a given $(C_k)_{k=1\ldots\mathcal{N}}$ profile (with $C_k \in]0,1[$ for each layer) and knowing CC, the function $f_\emptyset$ is strictly increasing, so $f_\emptyset^{-1}$ exists. We compute α with a dichotomy method using a tolerance $\epsilon = 10^{-5}$. As we are only considering contiguous cumulus clouds in this study, we only encounter $\alpha \in [0, 1]$ (see further), but the equations remain applicable for minimum overlap and $\alpha < 0$, which can be found for non-contiguous clouds (Oreopoulos, Cho, & Lee, 2022).

Equation 12 gives the expression of α for a given cloud cover CC and cloud fraction profile $(C_k)_{k=1\ldots\mathcal{N}}$. Equation 10 allows us to compute CC if we know the overlap parameter α and the profile $(C_k)_{1\ldots\mathcal{N}}$. Therefore for any given profile $(C_k)_{1\ldots\mathcal{N}}$ and given the ERO model, it is equivalent to know CC or α (or the decorrelation length, see further).

2.3. Vertical Subgriding

We are now going to use the same method but to define how to generate a sample of subcolumns with a higher vertical resolution starting from an atmospheric column with a coarse vertical resolution. We start from such a single column of N coarse layers from which we know the vertical volume cloud fraction distribution $\left\{\widehat{C_v}\right\}_{k=1\ldots N}$, and we generate subcolumns with n times more vertical layers, $\mathcal{N} = (N \times n)$. We introduce the hypothesis that at every coarse layers of the atmospheric column, the volume cloud fraction is the same for all the n sublayers:

$$\forall \quad l \in \mathcal{L}_k^n, \quad C_{v,l} = \widehat{C}_{v,k}$$

where $\mathcal{L}_k^n$ is the ensemble of n sublayers within the coarse layer k.

We then compute, like done previously, the probability $P_\emptyset$ to generate a clear-sky subcolumn. As the cloud fraction in a single coarse layer is uniform, the intralayer transition probability $P(S_l = 0 | S_{l-1} = 0)$ (Equation 8) between sublayers inside the same coarse layer k simplifies as:

$$P(S_l = 0 | S_{l-1} = 0) = P_{intra,k} = \alpha + (1-\alpha)\left(1 - \widehat{C}_{v,k}\right) \tag{13}$$

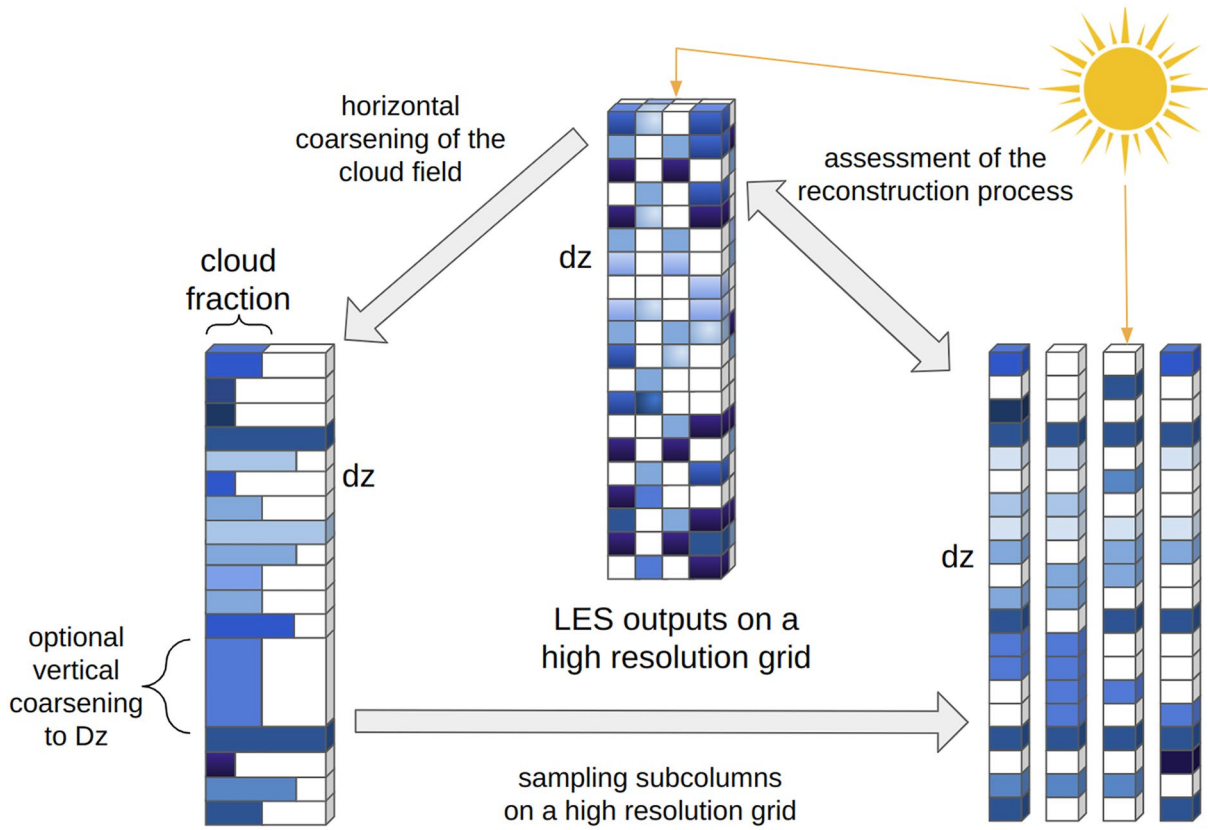

Figure 1. Method used to develop and assess our cloudy subcolumns sampling. The different shades of blue indicate different Liquid Water Contents, and white indicates clear-sky. The Large Eddy Simulations (LES) cloud field of resolution $dx = dy = dz = 25$ m is horizontally averaged into a single column and possibly averaged vertically to a coarse resolution $Dz > dz$. We then sample N_s subcolumns with a vertical resolution dz using the exponential-random overlap algorithm, and then assess the process by comparing the sample's cloud fraction profile and top of the atmosphere SW cloud albedo to the ones of the original LES. At each altitude of the sample, the Liquid Water Content (LWC) is homogeneous. The total LWC at each altitude is conserved, as well as the cloud fraction C_v.

For two adjacent cells that belong to two adjacent coarse layers, $C_{v,k}$ and $C_{v,k-1}$ can be different and the interlayer overlap transition probability, $P(S_k = 0|S_{k-1} = 0) = P_{inter,k}$ is given by Equation 8. Finally, $P_\emptyset$ is given by:

$$\begin{aligned} P_\emptyset(\alpha, N, n, C_v) &= \prod_{k=1}^{N}\left[P_{inter,k}\prod_{1}^{n-1}P_{intra,k}\right] \\ &= \prod_{k=1}^{N}\left[\left[\alpha + (1-\alpha)\left(1-\hat{C}_{v,k}\right)\right]^{n-1}\right] \\ &\times\left[\frac{\alpha * \left(1-\max\left(\hat{C}_{v,k-1},\hat{C}_{v,k}\right)\right)}{1-\hat{C}_{v,k-1}} + (1-\alpha)\left(1-\hat{C}_{v,k}\right)\right] \end{aligned} \tag{14}$$

Like done previously, we can compute the cloud cover generated by a given overlap parameter α, or if the cloud cover of the scene is known, we can inverse this equation using Equation 12 to compute the overlap parameter α that generates the same cloud cover. The next section shows the results of this subgriding: both its impacts on the cloud fraction profiles and the radiative properties of the ERO samples.

3. Evaluating α and the Cloud Generation

As done in many previous works such as Larson et al. (2002), Neggers, Duynkerke, and Rodts (2003), Neggers, Jonker, and Siebesma (2003), Neggers et al. (2011), we use LES as reference cases to assess our ERO algorithm. To test the algorithm presented in the previous section, different shallow cumulus cloud cases have been used. We mostly studied the Atmospheric Radiation Measurement Southern Great Plains site cloud case (Brown

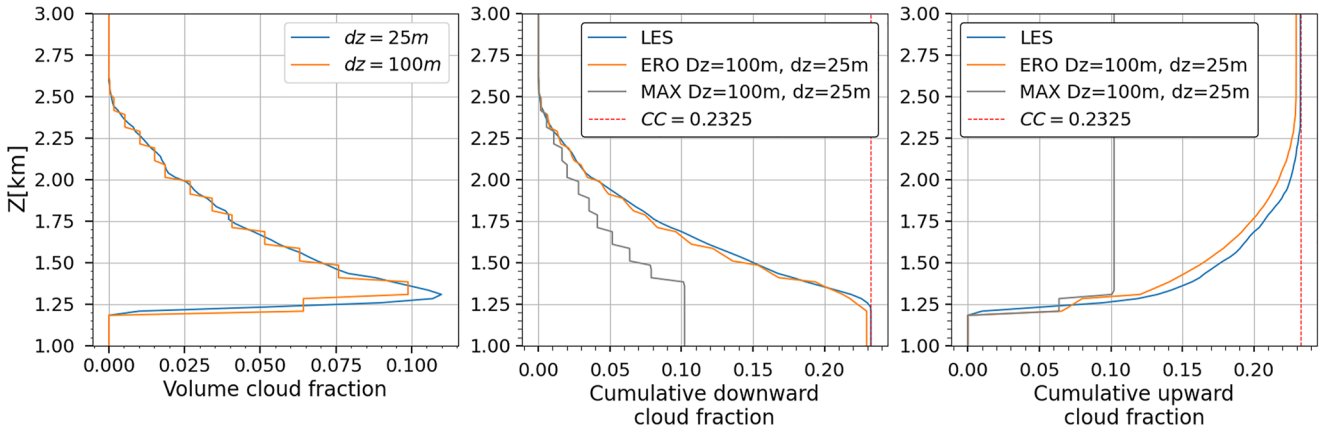

Figure 2. Vertical distribution of the volume cloud fraction (left), of the downward cumulative cloud fraction (middle) and upward cumulative cloud fraction (right, see text). On the middle and right panels are compared the profiles from the Large Eddy Simulations (LES) (blue) and those obtained with two overlap models: maximum overlap (gray) and exponential-random overlap (orange). The red dotted line shows the cloud cover CC of the scene. Both samples were made using the same initial single column with a vertical resolution $Dz = 100$ m and have the same final vertical resolution $dz = 25$ m than the LES. The data presented is the ARM cloud case (time step $h = 10$).

et al., 2002) showing the development of shallow cumulus convection over land, as well as two marine, trade-winds cumulus cloud cases Barbados Oceanographic Meteorological Experiment (BOMEX) (Siebesma et al., 2003) and Rain in Cumulus over the Ocean (RICO) (van Zanten et al., 2011), and another case of continental cumulus Small Cumulus Micro–physics Study (SCMS) (Neggers, Duynkerke, & Rodts, 2003). For each case we use the corresponding LES results obtained with the atmospheric non-hydrostatic model Mesoscale Non–Hydrostatic model (Lafore et al., 1998; Lac et al., 2018), and all these simulations represent a 6.4 km × 6.4 km × 4 km domain with a $dx = dy = dz = 25$ m resolution. The LES were generated using periodic boundary conditions and are representative of horizontally infinite domains for cumulus clouds. For each LES simulation we average all the cells of each vertical levels. Thus we obtain a single atmospheric column with the same vertical resolution dz, or a lower vertical resolution Dz, if we also perform vertical averaging, as shown in Figure 1. For each of these single columns we know, by means of the LES, the cloud cover CC, as well as the cloud fraction and the liquid water content for each layer. Doing so we go from a highly detailed 3D simulation to a single column, and we lose the horizontal cloud structure. Using this single column we then sample subcolumns with the ERO algorithm presented in the previous section. Finally, we assess this generation by comparing the statistical properties and solar albedo of the subcolumns with those of the LES.

3.1. Testing ERO and Subgriding Assuming the Overlap Parameter has a Vertically Constant Value

To assess the ERO generation process we first test the assumption that it is sufficient to use a single overlap parameter α for the whole cloud scene. We use an atmospheric column with a coarser vertical grid than the LES ($Dz = 100$ m for the coarse resolution, $dz = 25$ m for the LES, see Figure 2, left panel), and then use subgriding with the method presented in Section 2.3 to generate a sample of N_s subcolumns with a higher vertical resolution. The overlap parameter α used to generate this sample is computed with Equations 11, 12 and 14 to ensure the same cloud cover as the original scene (a similar approach is taken by Barker (2008a, 2008b)). Here and for the rest of the study, $N_s \simeq 6.5 \times 10^4$ subcolumns have been generated. For this number, the cloud cover of the LES is reproduced with a standard deviation 2.10^{-3}, and it has been verified that the standard deviation is decreasing like $1/\sqrt{N_s}$, where N_s is the number of subcolumns generated, as predicted by the central limit theorem. As a first test, we assess how the cumulative downward and upward cloud fractions at altitude z varies as a function of this altitude (Figure 2). As in Barker (2008a) and Oreopoulos, Cho, Lee, Lebsock, and Zhang (2022), we define the cumulative downward cloud fraction at altitude z as the areal fraction covered by clouds at altitude z when looking downward from TOA, and the cumulative upward cloud fraction at altitude z as the areal fraction covered by clouds at altitude z when looking upward from the ground.

The blue lines (Figure 2, middle and right panels) are the cumulative cloud fraction profiles of the original LES, with a cloud cover of 0.2325. The gray lines are obtained using a maximum overlap assumption, and show a cloud cover of only ∼10%. Since the scene consists of a single cloud block, this corresponds to models using

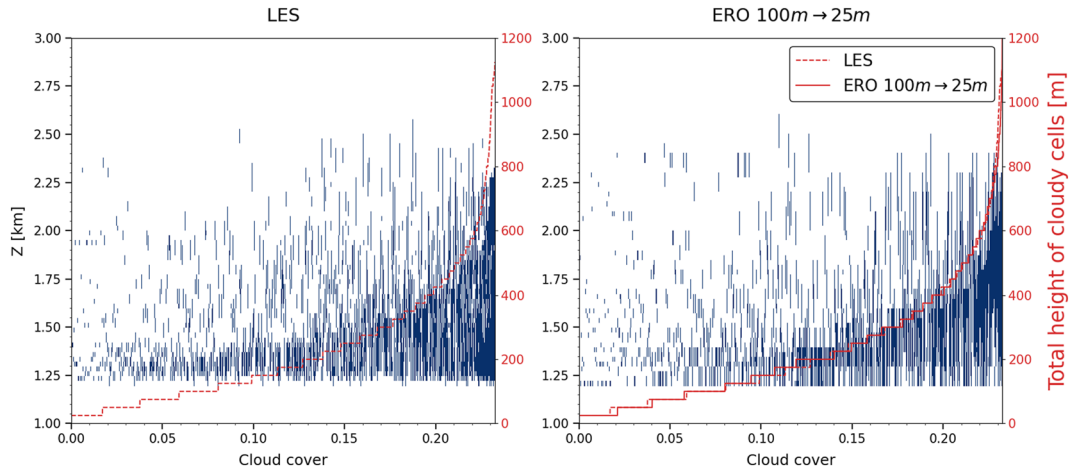

Figure 3. The cloudy subcolumns of the Large Eddy Simulations (LES) scene (left) are sorted along the number of cloudy cells in each subcolumns (dashed red). On the right the cloudy subcolumns out of a $N_s \simeq 6.5 \times 10^4$ sample of subcolumns generated with exponential-random overlap (ERO) sorted in the same way (solid red for the number of cloudy cells of the ERO profile). The number of cloudy cells of the LES has been reproduced in dashed to compare it better with that of the ERO generation. The field used is the 10th hour of the ARM case.

the classical maximum-random overlap and assuming the cloud fraction is vertically uniform within each coarse layer. The orange line is computed with ERO to match the cloud cover of the LES ($\alpha = 0.921$), with a very close cloud cover of 0.231 for that sample. The two plots on the right show that the ERO sampled subcolumns not only have the same cloud cover than the LES, but also a close cumulative cloud fraction at each altitude. The abrupt changes in the cumulative cloud fraction of the sampled subcolumns are a consequence of the hypothesis of a constant volume cloud fraction C_v in each coarse cell. Indeed, for the generation without vertical coarsening of the previous section ($Dz = dz = 25$ m), the vertical distribution of the cumulative cloud fraction is almost indiscernible to that of the LES (not shown).

To go further, Figure 3 shows the cloudy subcolumns of the same scene (cloudy cells in blue) sorted along the number of cloudy cells in each subcolumn (red). The left panel shows the cloudy subcolumns of the original LES, and the right panel shows the same plot for the sample of subcolumns generated by ERO. As the vertical distribution of cloudy cells is very close, it shows the ERO generation not only reproduces the cloud cover of the original scene, but also the distribution of the cloud fraction between the subcolumns.

We then assess the radiative characteristics of the sample by comparing the shortwave (SW) radiative properties of the LES and that of the ERO sample. We compute the mean albedo of the cloudy subcolumns (i.e., we do not consider any clear-sky subcolumns) for different cloud scenes using a path-tracing Monte Carlo code from Villefranque et al. (2019). It tracks photon paths throughout a virtual atmosphere, explicitly simulating the radiative processes such as scattering, absorption, and surface albedo. When a photon hits the top of the atmosphere (TOA), the algorithm adds its weight to a TOA counter (for reflection toward space), to a ground counter when it touches the ground (for ground absorption, here we put the ground albedo at zero), or to an atmospheric counter when it is absorbed (by liquid water or a gas). As the generated sample has no horizontal structure, we use the Independent Column Approximation along the Monte Carlo algorithm (McICA, Pincus et al., 2003). For different cloud scenes (one every hour), Figure 4 shows the cloud albedo obtained using different sampling hypotheses and using the original LES scenes, as well as the total albedo of the scenes, and their cloud cover. For each value of the coarse resolution Dz, a new overlap parameter has been computed: the different ERO scenes hence have the same cloud covers.

The maximum overlap assumption (gray line) shows a much higher cloud albedo (top panel) since it produces cloud scenes with less cloud cover (lower panel) and hence brighter clouds. The maximum overlap assumption therefore contributes to the too-few too-bright bias. Using ERO produces a much closer cloud albedo, and the coarse resolution of the initial atmospheric single column has little impact: the relative difference with the cloud albedo of the homogeneous LES is ∼1.5% when starting with a 25 m vertical resolution and only ∼2.5% when starting with a 200 m vertical resolution, for the scenes between 6 and 12 hr. The middle panel shows the total albedo of the cloud scenes, taking into account the clear-sky part. For the ERO samples, the difference with the

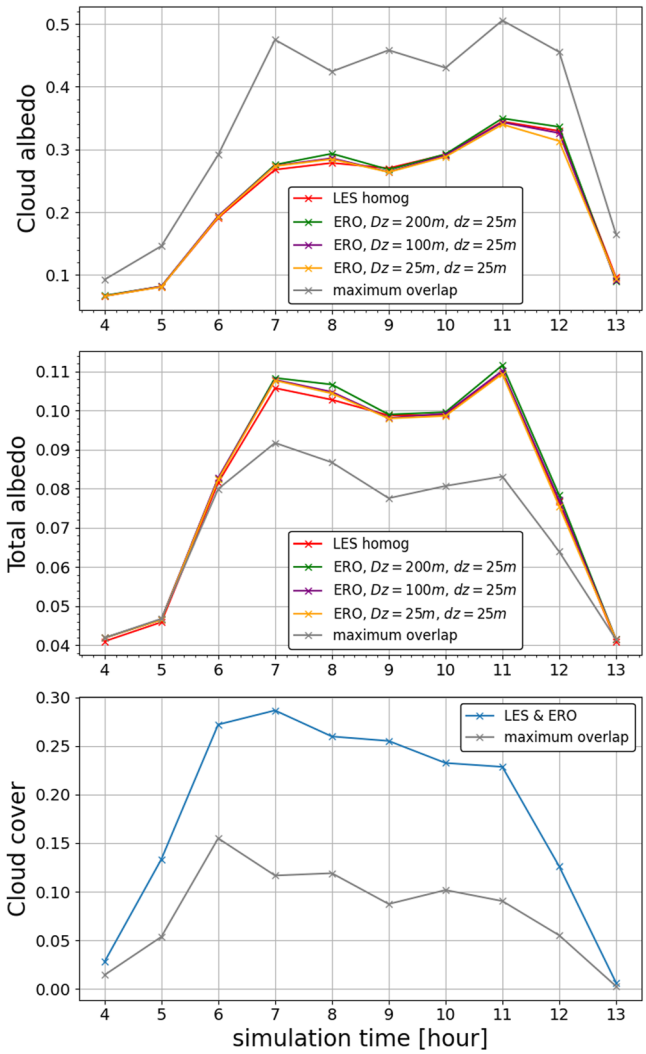

Figure 4. Cloud albedo (top panel), total albedo (middle panel) and cloud cover (lower panel) for the Large Eddy Simulations (in red), for exponential-random overlap starting from columns with different coarse resolutions Dz, and for maximum overlap with the coarse resolution $Dz = 100$ m (in gray). The albedo of each scene is computed using a Monte-Carlo algorithm under the Indenpendent Column Approximation, for the ARM cloud case scenes (time steps $h \in [4, 13]$). The surface albedo is set at zero, Dz is the vertical resolution of the coarse atmospheric single column and dz that of the reconstructed sample. In all scenes the in-cloud Liquid Water Content is homogeneous for each layer. For each computation, 10^6 realizations were made, with a Monte-Carlo standard deviation of the cloud albedo of 10^{-6}.

LES values follow the differences of the cloud albedo, as the scenes have the same cloud cover, and so also the same clear-sky cover. For the maximum overlap sample, the total albedo is lower than that of the LES. The smaller cloud covers overcome the brighter cloud albedo.

3.2. Analysis of the Overlap Parameter α

In Section 2 we established the relationship between the overlap parameter α and the cloud cover CC and used it in 3.1 to determine α from the cloud cover CC diagnosed from LES results. In this section we analyze the overlap parameters computed this way and compare them to the overlap parameters computed using other methods. For two different cloudy atmospheric layers at the altitudes z_k, z_l the overlap parameter $\alpha_{k,l}$ and a decorrelation length L_α are usually related to each other via the following relation (Bergman & Rasch, 2002; Hogan & Illingworth, 2000; Mace & Benson-Troth, 2002; Oreopoulos, Cho, & Lee, 2022):

$$\alpha_{k,l} = \exp\left(-\int_{z_k}^{z_l} \frac{\mathrm{d}z}{L_\alpha(z)}\right) \tag{15}$$

If the decorrelation length L_α is constant on the vertical (which is generally assumed), it becomes:

$$\alpha_{k,l} = e^{-|z_l - z_k|/L_\alpha} \tag{16}$$

The decorrelation length (and hence the overlap parameter of a scene) is often computed by fitting an exponential function to the profile of the overlap parameter dependance to the separation distance $|z_k - z_l|$ (Hogan & Illingworth, 2000; Oreopoulos & Norris, 2011), according to Equation 16. Figure 5 shows the variations of the overlap parameters α computed at different times of the day of the ARM simulations, with three different methods. The overlap parameter $\alpha_{LES,fit}$ is computed by fitting an exponential function to the profile of the overlap parameter on our LES simulations with Equation 16. This profile was obtained by computing the mean overlap parameter for each possible separation distance by using $C_a = \alpha C_{\max} + (1 - \alpha)C_{rand}$. The overlap parameter $\alpha_{25,Dz}$ corresponds to the overlap parameter computed using Equation 14 to reproduce the cloud cover CC with vertical subgriding at a resolution $dz = 25$ m, starting from a column with a vertical resolution $Dz = 100$ m. The overlap parameter $\alpha_{LES,loc}$ is the mean of the local consecutive overlap parameters $\alpha_{k,k-1}$ on the LES simulations at $dz = 25$ m.

Three simulation times (hours 4, 5, 13) show poorly consistent values, caused by a smaller cloud cover of those scenes when the cloud layer is developping in the morning and dissipating at the end of the day. Without these three time steps, for the hours 6 to 12, the mean values of those overlap parameters are $\bar{\alpha}_{25,Dz} = 0.915$, $\bar{\alpha}_{LES,loc} = 0.916$ and $\bar{\alpha}_{LES,fit} = 0.866$. The equivalent decorrelation lengths are $\bar{L}_{\alpha,25,Dz} = 291$ m, $\bar{L}_{\alpha,loc} = 298$ m and $\bar{L}_{\alpha,fit} = 205$ m. Barker (2008a) refers to the decorrelation length that reproduces CC the effective decorrelation length. The values computed locally on the LES and the ones computed for ERO are close and stable during the day, when the exponential fit shows much wider variations. In the BOMEX case however (with the same resolutions), the overlap parameter daily averages are closer to each other: we find $\bar{\alpha}_{25,Dz} = 0.87$, $\bar{\alpha}_{loc} = 0.88$ and $\bar{\alpha}_{fit} = 0.85$, and equivalently $\bar{L}_{\alpha,25,Dz} = 179$ m, $\bar{L}_{\alpha,loc} = 195$ m and $\bar{L}_{\alpha,fit} = 153$ m. The decorrelation lengths that are computed here ($L_\alpha = 200$–300 m) are comparable to those computed in the literature with similar LES simulations (Neggers et al., 2011; Sulak et al., 2020; Villefranque et al., 2021). However such values of the decorrelation lengths can appear small when the corresponding overlap parameters are close to 1. That is because of the high vertical resolution used here, $dz/L_\alpha \ll 1$ causing overlap parameters to be close to 1. The difference with decorrelation lengths in the literature that take into account the overlap of whole atmospheric

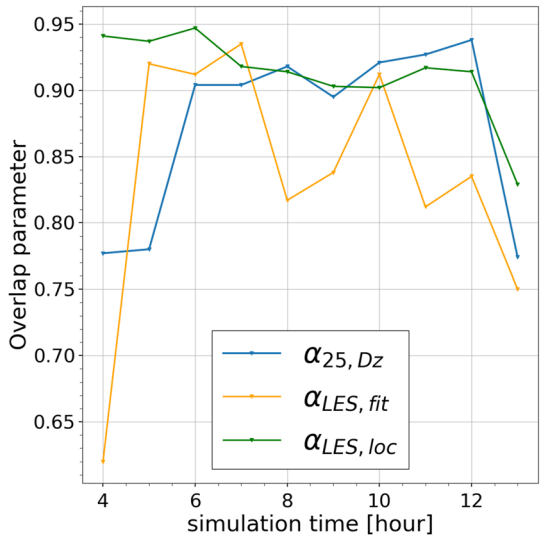

Figure 5. Overlap parameters computed with three different methods (see text) at each time step of the Large Eddy Simulations simulations. The data used are the ARM cloud fields.

columns, as obtained by satellite observations, and used in global models, is further discussed in Section 5.

We have also computed the overlap parameter α using ERO like done previously but on the individual largest clouds of the studied scenes, and found very similar results than for the total scene. For instance, for the scene ARM ($h = 10$) when taking into account the 45 clouds that account for 99% of the cloud cover (out of 67 individual clouds in the scene), the mean overlap parameter over the different clouds is $\alpha_{25,Dz} = 0.913$ (with a standard deviation of 0.07), which is equivalent to a decorrelation length of 275 m.

4. Using ERO to Model Subgrid Heterogeneity and Overlap Coarse Vertical Layers

To summarize the previous section, if we know the overlap parameter $\alpha_{25,Dz}$ or the cloud cover of the scene, the volume cloud fraction C_v and the liquid water content LWC for an atmospheric column of vertical resolution Dz, we are able to generate a sample of subcolumns with a higher vertical resolution (25 m, the same as the LES) with properties that are close to the LES so that the cloud albedo of the scene only differs by a few percent (about 2% on the whole day for the ARM and the BOMEX cases). But in this approach, the radiative computations are made on a high resolution vertical grid, not on the coarse one. In this section we will focus on how to adapt the method to deal directly with coarse grids, without having to use a finer mesh. To do so we will characterize how the subgrid properties of clouds should be computed on the coarse grid, and then how they should be combined vertically so that both the vertical cloud structure, the cloud cover and in fine the cloud albedo remain close enough to the high-resolution reference case.

4.1. Subgrid Properties on the Coarse Grid

Defining subgrid properties on the coarse vertical grid requires to distinguish two cloud fractions, the areal cloud fraction C_a and the volume cloud fraction C_v (Del Genio et al., 1996; Jouhaud et al., 2018). C_v represents the volume fraction of the layer that contains clouds (i.e., where liquid or solid water particles are present), whereas C_a represents the areal fraction of the layer covered by clouds when looking from above or below. In other words, C_a is the vertical projection of C_v, and it is C_a that is used by radiation codes in GCMs and remote sensing.

At the LES grid scale, we have assumed that a grid cell's cloud state is either clear-sky or cloudy. Therefore $C_v = C_a$ when the single column has the same vertical resolution as the LES. This is no longer the case when the vertical grid is coarsened, and ERO can be used to compute C_a, knowing C_v. For that we consider an atmospheric cloudy column of coarse vertical resolution $Dz = n \times dz$. If C_v is known and vertically uniform within each coarse layer, we are back in the configuration we were in Section 3 when using subgriding, with $C_{v,k} = \widehat{C}_{v,k}$. We can then compute the subgrid areal cloud fraction $C_{a,sg,k}$ as the cloud cover of a single coarse layer, by using Equation 14, but setting to zero the volume cloud fractions above and below the coarse layer considered ($N = 1$):

$$C_{a,sg,k} = 1 - (1 - C_{v,k})(\alpha_{sg} + (1 - \alpha_{sg})(1 - C_{v,k}))^{n-1} \quad (17)$$

where α_{sg} is the overlap parameter used here to compute this subgrid areal cloud fraction. Although other choices are possible, we choose here to use $\alpha_{sg} = \alpha_{25,Dz}$. If the cloud cover CC is known but not $\alpha_{25,Dz}$ we can compute it by inverting Equation 12. The next figure illustrates the performance of that equation.

The top panels of Figure 6 show the profile of C_a obtained using the LES original data, using Equation 17, and also assuming maximum overlap within each layer, for two coarse resolutions (left panel at $Dz = 100$ m and right panel $Dz = 200$ m). The overlap parameter $\alpha_{25,Dz}$ is computed for the two resolutions to ensure the same cloud cover ($\alpha_{25,100} = 0.921$, $\alpha_{25,200} = 0.911$), and then used to compute $C_{a,sg}$ with Equation 17. The maximum overlap assumption (gray) does a poor job representing the areal cloud fraction profile, and leads to a relative error of 30%–50%. It shows the error made when neglecting subgrid variability, that is, assumming $Ca = C_v$ on the coarse grid. For this assumption, the coarser the vertical resolution, the larger the error. Using Equation 17 allows a better representation of the areal cloud fractions, even if a substantial error remains. For all methods, the largest

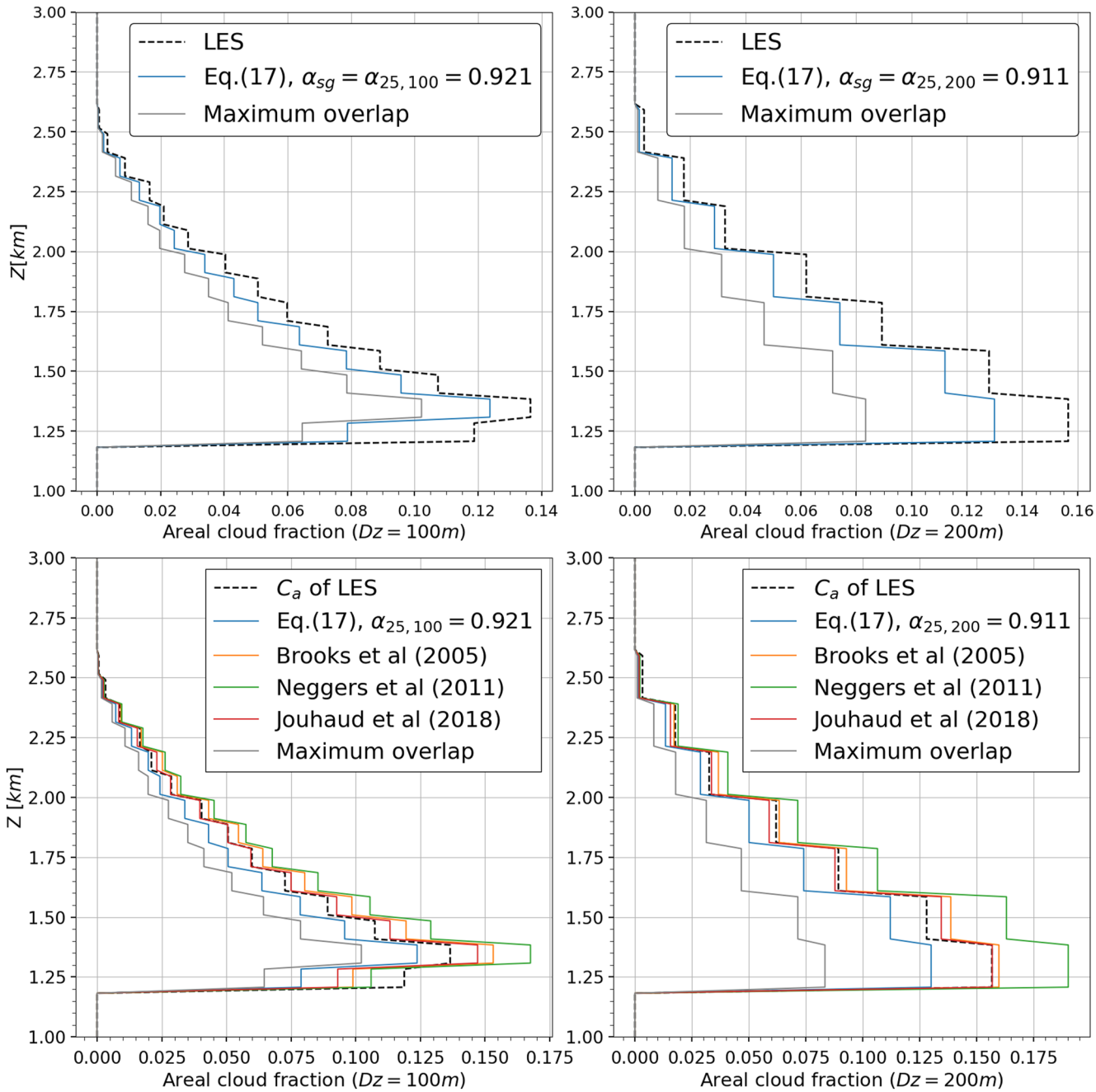

Figure 6. Vertical distribution of the areal cloud fraction C_a obtained with Large Eddy Simulations (LES) full resolution results or with different approximations with a coarse vertical resolution of 100 m (left panels) or 200 m (right panels). The top panels compare the LES (dashed black) with exponential-random overlap using Equation 17 and $\alpha_{sg} = \alpha_{25,Dz}$ (blue) as well as the maximum overlap sample (gray). The bottom panels also compare Equation 17 with other parametrizations found in the literature. The cloud case is ARM ($h = 10$).

error corresponds to the lower layer which is the bottom of the cloud. On this layer, the volume cloud fraction C_v decreases steeply, which makes the hypothesis of a constant C_v inaccurate.

To go further we also compare the performance of Equation 17 with that of other references in the literature (Figure 6, bottom panels). Neggers et al. (2011) and Jouhaud et al. (2018) have both been developed using LES data of small cumulus with $C_v \simeq 0.1$, including the ARM and BOMEX cases, and are therefore comparable to our method. Brooks et al. (2005) develops a lidar and radar-based parametrization of C_a using C_v, with the possibility

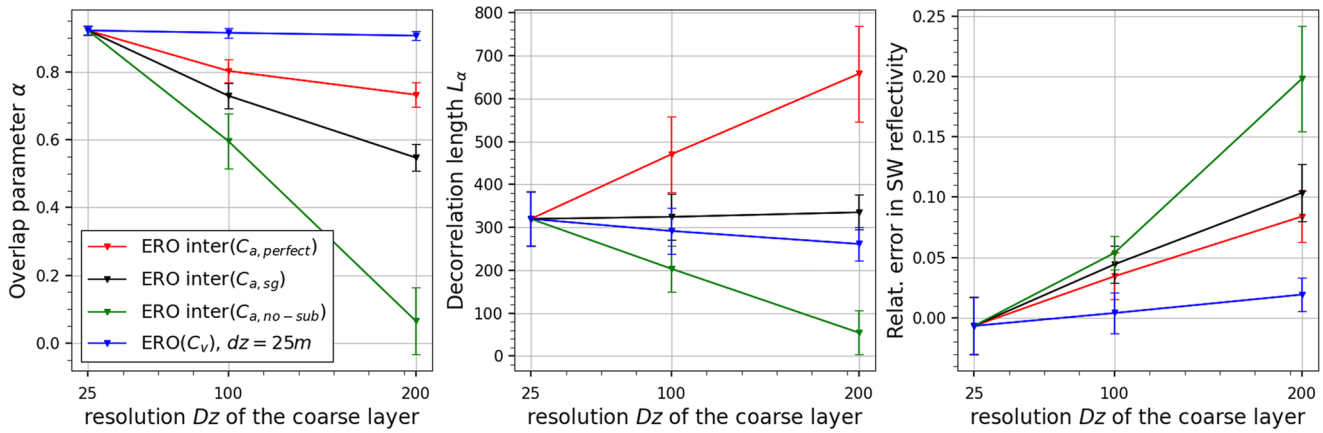

Figure 7. Overlap parameters (left) and decorrelation lengths (middle) for the ARM simulations (hours 6 to 12), for different coarse resolutions Dz and for different reconstructions using exponential-random overlap (see text). The daily mean value is shown. The overlap parameters are computed to match the cloud cover of the Large Eddy Simulations (LES). The right panel shows the corresponding relative error in SW cloud albedo at TOA compared to that of the LES when using those overlap parameters to generate the scenes. For each plot, the standard deviation due to the different simulation times is shown as an error bar.

to take into account wind shear (not used here), and is valid on a wider range of cloud covers and situations. Brooks et al. (2005) and Jouhaud et al. (2018) show the smallest errors with C_a of the LES.

Our approach favors an accurate cloud cover on the whole vertical extent of the cloud layer. Results show that with this approach we tend to underestimate the areal cloud fraction of the coarse layers. This is because the overlap parameter α has been computed to match the cloud cover CC of the whole scene, not the areal cloud fraction C_a of each coarse layer. When only used for the subgrid scale it creates too small an areal cloud fraction. This underestimation is still much smaller than when considering maximum overlap. The gap in areal cloud fraction caused by using our method is similar to those caused by other approximations of the literature, but with an opposite sign in the difference. Our underestimation of C_a was already visible in Figure 2 on the panel showing "cumulative downward cloud fraction." The only difference between using subgriding or not is the hypothesis C_v = constant in each coarse layers, so we can conclude than the underestimation of our method comes from this hypothesis.

4.2. Interlayer Overlap

We now consider that the vertical profile of the areal cloud fraction $(C_{a,sg})_z$ that takes into account the subgrid heterogeneity on the coarse grid, is known. We have to define the overlap of the coarse layers, and we again choose to define it to ensure the conservation of the cloud cover CC. To compute the subgrid areal cloud fraction profile $(C_{a,sg})_z$ in the previous section, we were using the first part of Equation 14, which represents the subgrid overlap. Here we use the second part of the equation, which represents the interlayer overlap, using the unknown interlayer overlap α_{inter}. This corresponds to using Equation 10 on the coarse grid with $(C_{a,sg})_z$ to produce the cloud cover:

$$CC = 1 - \prod_{k=1}^{N}\left[\frac{\alpha_{inter}(1 - \max(C_{a,sg,k}, C_{a,sg,k-1}))}{1 - C_{a,sg,k-1}} + (1 - \alpha_{inter})(1 - C_{a,sg,k})\right] \quad (18)$$

The overlap parameter α_{inter} can be computed as in the previous sections, by inverting Equation 18 to constrain the cloud cover CC:

$$\alpha_{inter} = f_{\emptyset}^{-1}(1 - CC) \quad (19)$$

4.3. Generating Subcolumns on the Coarse Grid

To summarize the previous steps, we can now compute the overlap parameter $\alpha_{25,Dz}$ with Equation 12, the areal subgrid cloud fractions $(C_{a,sg})_z$ using Equation 17 with $\alpha_{sg} = \alpha_{25,Dz}$, and then the overlap parameter α_{inter} using Equations 18 and 19 in order to overlap these coarse layers to produce the cloud cover CC. The corresponding decorrelation length can be computed with Equation 16 and Dz as the separation distance. However, at this stage,

there is no evidence of a formal link between the two overlap parameters $\alpha_{25,Dz}$ and α_{inter} or decorrelation lengths, or of a dependence to the vertical resolution. In any case, we have not found one.

For $\alpha_{25,Dz}$ and the corresponding decorrelation length (Figure 7, blue lines, left and middle panels), we find that they depend little on the starting coarse resolution Dz on this 25–200 m range, with mean values $\overline{\alpha_{25,Dz}} = 0.915$ and $\overline{L_{\alpha,25,Dz}} = 291$ m. Using this overlap and Equation 17 we then compute the subgrid profile $(C_{a,sg})_z$, as well as the interlayer overlap parameter α_{inter} using Equations 18 and 19.

For the overlap parameter α_{inter}, we find that it varies with the resolution Dz but the corresponding decorrelation length varies little from $\overline{L_{\alpha,sg}} = 326$ m (Figure 7, black lines, left and middle panels). The decorrelation lengths show small variation whether we generate the subcolumns on the fine or coarse grid, and depends little on the resolution of the coarse grid (Figure 7, middle panel, blue and black lines). When it comes to radiative effects (Figure 7, right panel), the error made on the SW cloud albedo is still small even when computed on the coarse grid (black line) rather than on the finer grid (blue line).

4.4. Analysis and Comparisons of Interlayer Overlap for Different Estimations of the Areal Cloud Fraction

Here we investigate, using Equations 18 and 19, how the overlap parameter α_{inter} and the decorrelation length should vary to keep the correct value of the cloud cover for different estimations of the areal cloud fraction C_a in Equation 18, instead of $C_{a,sg}$. First we consider the extreme case where no subgrid heterogeneity is considered (Figure 7, green lines), meaning the subgrid areal cloud fraction equals the volume cloud fraction $(C_{a,no-sub})_z = (C_v)_z$ on the coarse grid. When the starting coarse resolution is $Dz = 25$ m, we are already at the finest resolution of the simulations (which means the coarse grid can not be finer), and all the reconstructions are the same. As shown in Figure 6, $C_{v,z} < C_{a,z}$ for any z, so to generate the same cloud cover, the overlap when no subgrid is taken into account has to be closer to random (i.e., α closer to 0), hence $\alpha_{inter,no-sub} < \alpha_{inter,sg}$. For $Dz = 200$ m, the interlayer overlap without subgriding is already almost fully random. We then consider the case where the subgrid reconstruction takes perfectly into account the subgrid heterogeneity and reproduces perfectly the areal cloud fraction profile $(C_{a,perfect})_z$ (Figure 7, red lines). We then compute the interlayer overlap corresponding to this profile with Equations 18 and 19. The same reason applies to explain the difference with the interlayer overlap parameters computed for the areal subgrid cloud fraction profile: as shown in Figure 6, $C_{a,sg}$ approaches $C_{a,perfect}$ in such a way that for any altitude, we have $C_{a,perfect} > C_{a,sg} > C_{a,no-sub}$. Hence to conserve the same cloud cover, the different overlap parameters follow $\alpha_{inter,perfect} > \alpha_{inter,sg} > \alpha_{inter,no-sub}$.

The middle panel of Figure 7 shows the corresponding decorrelation lengths, computed from each overlap parameter α with $L_\alpha = -dz/\ln(\alpha)$, where dz is the vertical resolution of the target grid. When doing overlap on the coarse grid, the final resolution is $dz = Dz$ (red, black and green lines). When doing ERO on the finer grid, the final resolution is $dz = 25$ m (blue lines). We see that for interlayer overlap, the decorrelation lengths have a strong dependence to the resolution when overlapping coarse layers of which the areal fraction is either perfect $(C_{a,perfect})_z$ or determined assuming no subgrid heterogeneity $(C_{a,no-sub})_z$, with important variations. This is not the case when the areal cloud fraction $C_{a,sg}$ is computed using a consistent representation of cloud heterogeneity on both subgrid scale and interlayer overlap (black) or when reconstructing on the finer grid (blue). Numerical tests were made on artificial cloud scenes with constant cloud fractions and various cloud covers, as well as on the same LES clouds with double the vertical extent to go up to 400 m coarse resolutions, and this appears to be a consistent result: strong dependence of the decorrelation lengths with the coarse resolution when overlapping $(C_{a,perfect})_z$ and $(C_{a,no-sub})_z$, but a small dependence to the resolution of the decorrelation length when overlapping $C_{a,sg}$. This dependence of L_α with Dz has already been mentioned by Hogan and Illingworth (2000) and Räisänen et al. (2004), but does not seem to be taken into account in the literature when generating cloudy subcolumns from GCMs or for observational simulators (Bodas-Salcedo et al., 2011; Pincus et al., 2005; Swales et al., 2018).

4.5. Cloud Albedo Dependence on the Vertical Cloud Structure

We have shown in Section 3.1 that by using ERO and a subgrid overlap parameter on a finer grid (Figure 4 and blue lines of Figure 7) we can reproduce the cloud albedo of those scenes with a 2% relative error. In the previous section we show that it is also possible to take into account the subgrid scale directly on the coarse grid by choosing to compute the surface cloud fraction as a bulk subgrid property using the volume cloud fraction and a subgrid overlap parameter. Overlapping this computed areal subgrid cloud fraction leads to a relative error in cloudy albedo

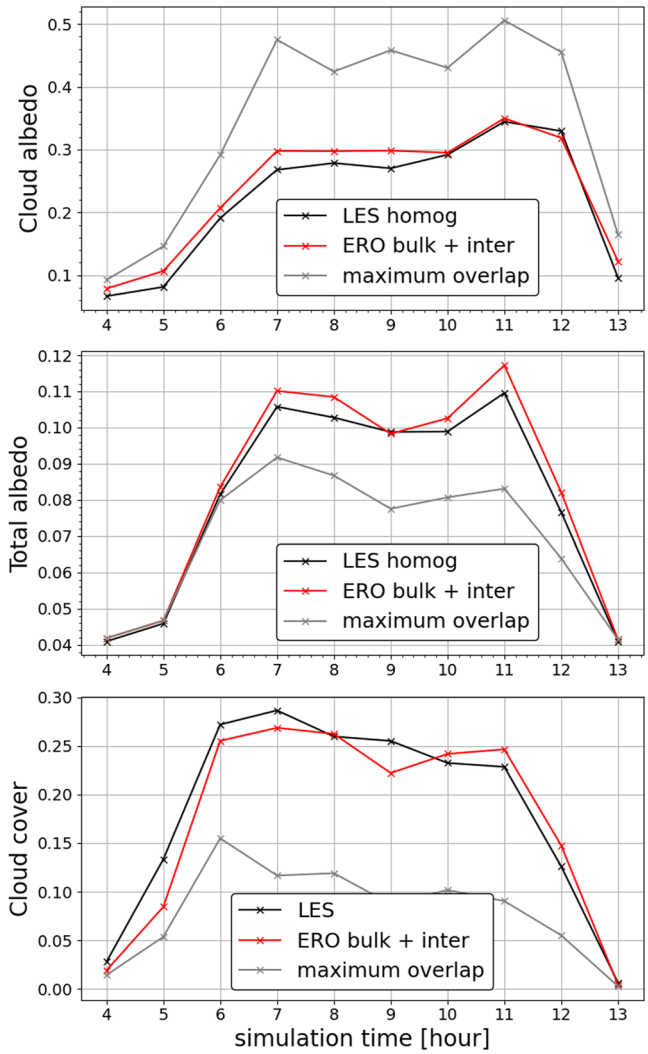

Figure 8. Cloud albedo (top panel), total albedo (middle panel) and cloud cover (lower panel) for the Large Eddy Simulations (in black), our reconstruction using exponential-random overlap (in red) and a maximum overlap reconstruction (gray). The constant decorrelation length used here both for the subgrid computation of the areal cloud fraction profile and its interlayer overlap is $L_\alpha = 309$ m. The scenes are the ARM case (time steps $h \in [4, 13]$). In all scenes the Liquid Water Content is homogeneous for each layer.

of $\simeq$ 10% for coarse resolutions of 100 and 200 m (Figure 7, black line). If this subgrid computation were perfect to take into account the subgrid scale, it would lead to a slightly improved 5%–8% relative error in cloud albedo for coarse resolutions of 100 and 200 m (Figure 7, red line). Finally, even without taking into account any subgrid scale by overlapping $(C_{a,no-sub})_z$ on the coarse grid, we can approach the albedo of the LES scenes within a 20% relative error (for a resolution of 200 m, Figure 7, green line) if the cloud cover is reproduced. As all the generations shown in Figure 7 have the same cloud cover and mean liquid water path as the LES simulations, the difference in cloud albedo are all due to vertical subgrid heterogeneity. If the conservation of the cloud cover is of first order importance for the cloud albedo, the subgrid scale information contained in the cloud fraction profile can have a significant impact on the cloud albedo as well, up to 20%. Numbers in this section are computed on 7 scenes from the ARM cloud case, but similar results were also found consistently in several other cases, see Figures S1–S3 in Supporting Information S1.

5. Implications

In this last section we address some more global implications of our method, especially on the use and estimate of the decorrelation lengths, as well as the radiative impact of LWC horizontal heterogeneity, which had not been taken into account in this paper until now.

5.1. How to Generate the Cloud Vertical Profile

The starting point of the developments in Section 3 and 4 was to determine how to correctly represent the cloud cover and the SW cloud albedo of a cloud scene in the context of ERO. We have shown in Section 3 that by defining the appropriate decorrelation length $L_{\alpha,25,Dz}$ we can generate a cloud scene with the correct cloud cover and a close SW cloud albedo. This can be done on a new grid with higher vertical resolution (25 m here) as long as the initial coarse resolution and the final resolution are both taken into account in the computation of the overlap. This can also be done directly on the coarse grid without losing much accuracy on the cloud albedo by taking into account both the subgrid scale and the interlayer overlap (Section 4.3). So far we have assumed that the cloud cover is known, whereas in general we are trying to determine the cloud cover. So we have to reverse the previous problem and address the following question: how to create the right cloud cover and the right cloud albedo from the information given by a column with a coarse vertical grid? In this context, an important result of Section 4.3 is that if we consistently account for subgrid heterogeneity and coarse layer overlap, then the decorrelation lengths used for the subgrid and the overlap are almost the same and they depend weakly on the vertical resolution, as we can see on Figure 7.

The procedure for reconstructing a cloud scene that we propose is as follow: given any volume cloud fraction profile $(C_v)_z$ at resolution Dz and the decorrelation length L_α for a reference resolution (here $dz = 25$ m), the subgrid heterogeneity is taken into account by computing a profile of the areal cloud fraction $(C_{a,sg})_z$ with Equation 17, with $n = Dz/dz$ in the equation. The same decorrelation length $L_{\alpha,}$ allows to overlap these coarse layers and to compute the cloud cover (Equation 18). As we can see on Figure 7 for the case studied here, $L_{\alpha,25,Dz} \simeq 291$ m and $L_{\alpha,sg} \simeq 326$ m, so for both steps of this reconstruction we choose to use the unique decorrelation length that is the mean of the two: $\bar{L}_\alpha = 309$ m. We find similar results than those shown on Figure 7 for three other cumulus cloud cases simulated by the same LES, with $L_{\alpha,25,Dz}$ and $L_{\alpha,sg}$ relatively independent of the resolution Dz. For the RICO case we have $\bar{L}_\alpha = 217$ m, for BOMEX $\bar{L}_\alpha = 202$ m and for SCMS $\bar{L}_\alpha = 273$ m (see Figures S1–S3 in Supporting Information S1). Here a different decorrelation length has been computed for

each cloud case. The determination of this decorrelation length in a more general case is beyond the scope of this study. As it can be seen on Figure 8, the scenes generated with this method show a good reproduction of the cloud cover, cloud albedo and total albedo, with relative errors compared to the LES of only −10%, 11%, and −3% respectively, which is significantly better than the errors caused by the maximum-random assumption. We also see from this figure that the maximum overlap causes a "too few too bright" bias here, with a cloud cover too small and a cloud albedo too large. But the two errors do not compensate and the total albedo of the scenes is underestimated. Increasing the liquid water content seen in the radiative computations to balance the mean radiative flux at TOA could correct the value of total albedo but in the same time would also worsen the "too bright" part of the bias. Similar results are found for the three other cloud cases and can be found in Supporting Information S1 on Figures S4 to S6.

5.2. Variations of the Decorrelation Length With the Measurement Resolution

Decorrelation lengths used in GCMs are often derived from observational data from active remote sensing (Jing et al., 2016; Oreopoulos & Norris, 2011; Oreopoulos, Cho, & Lee, 2022). As shown in the previous section, the vertical resolution of the grid on which we generate the cloud scene can have a significant impact on the values of overlap parameters and decorrelation lengths. This may also be applied to the vertical resolution at which those instruments measure cloud fraction profiles, their overlap and hence decorrelation lengths. At the vertical resolution of those instruments, for example, 480 m for CloudSat, a layer is identified as entirely cloudy even if the cloud does not fully extend on the vertical of the layer. Hence the measured profile is the areal cloud fraction $(C_a)_z$ for a coarse layer of thickness $Dz = 480$ m. Combining Equations 17–19, we can compute overlap parameters in various situations, including when dealing with different vertical resolutions. This can be used to compare overlap parameters given by observational measures with different resolutions.

We will consider that two different instruments I_1 and I_2 have the vertical resolutions dz_1 and dz_2, which is finer, with $dz_1 = n \times dz_2$. We suppose they observe the same cloud scene and detect the same cloud cover. Those instruments give us access to two sets of data statistically representing the same cloud scene: $(C_{a,1})_z$, $L_{\alpha,1}$, and $(C_{a,2})_z$, $L_{\alpha,2}$, where $L_{\alpha,i}$ are the decorrelation lengths corresponding to the measured areal cloud fraction profiles. Using the cloud fraction profile with finer vertical resolution $C_{a,2}$ we can use interlayer ERO with $L_{\alpha,2}$ on blocks of n fine layers to compute the corresponding areal cloud fraction profile at the resolution dz_1, $c'_{a,1}$. Knowing the cloud cover CC, we can then compute with Equation 19, the decorrelation length $L'_{\alpha,1}$ that would generate CC with this profile. We can compare $L_{\alpha,1}$ and $L'_{\alpha,1}$ now that they refer to similar resolutions. For the ARM simulations used on Figure 7, let us consider I_1 with resolution $dz_1 = 200$ m and I_2 with resolution $dz_2 = 25$ m. This example is studied in Section 4.4, where we analyzed the evolution of L_α with the vertical resolution for a perfect estimation of the areal cloud fraction profile. I_2 would measure a decorrelation length $L_{\alpha,2} = 320$ m, while I_1 would measure $L_{\alpha,1} = 658$ m (Figure 7 middle panel, in red). We get a factor 2 on the estimation of the decorrelation length in this case. The vertical extent of the studied clouds is too small to be able to compute the decorrelation length in the case of the vertical resolution of CloudSat at 480 m, but an even larger effect is expected.

The decorrelation lengths computed from observations with a low vertical resolution (a couple hunder meters) are often much larger than the ones computed in this study, with $L_\alpha \sim 2$ km (Barker, 2008a; Hogan & Illingworth, 2000; Jing et al., 2016; Oreopoulos & Norris, 2011; Willèn et al., 2005). This difference can then partly be explained by the difference in vertical resolution, as the decorrelation lengths shown here are comparable to those computed in the literature with LES simulations with similar vertical resolutions (Neggers et al., 2011; Sulak et al., 2020; Villefranque et al., 2021). The difference in horizontal resolutions (Astin & Di Girolamo, 2014; Naud et al., 2008; Tompkins & Di Giuseppe, 2015) can also impact the overlap, but it is not studied here.

5.3. Considering LWC Distributions

Until now, we focused on the vertical distribution of the cloud fraction and cover, and therefore assumed an homogeneous LWC in each horizontal layer. In this section we add distributions of the LWC between the subcolumns and study its impact on the radiative properties of the generated scenes. The impact of the LWC heterogeneity on the cloud albedo of a scene is well documented and known to be of second order compared to the accurate reproduction of the cloud cover (Barker et al., 1999; Barker & Räisänen, 2005; Oreopoulos et al., 2012). We want to check the ability of our method to reproduce those results, and compare the second order impacts of the LWC horizontal heterogeneity to those of the cloud fraction subgrid vertical heterogeneity shown in Section 4.5. To do

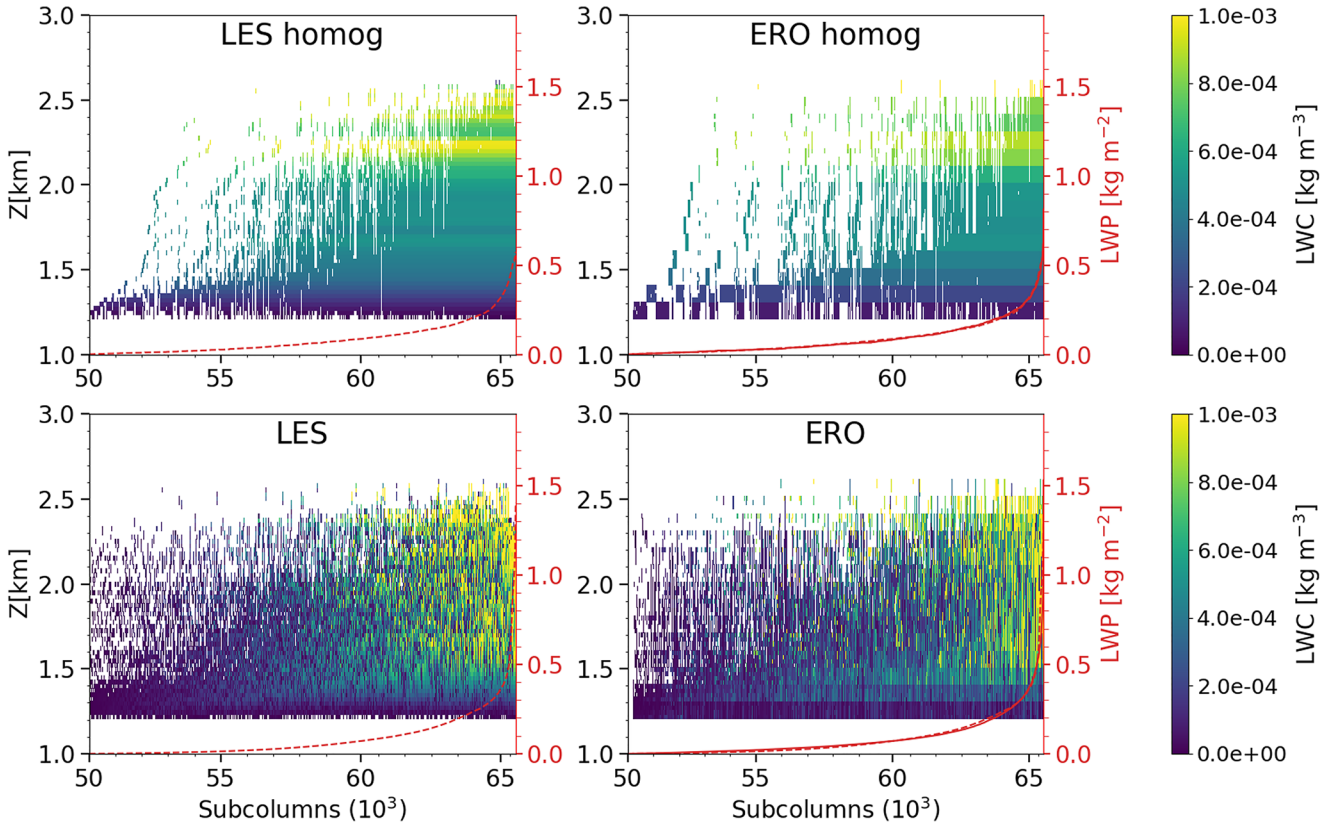

Figure 9. The liquid water content of each scene's cloudy subcolumns in the Large Eddy Simulations (LES) simulations (left panels) and reconstructed using exponential-random overlap (right panels). The subcolumns have been sorted along their Liquid Water Path (LWP) (red lines). The red lines represent the LES (dashed line) and generated (solid line) LWP, the former being represented on the right panels as well to facilitate the comparison. Top panels are homogeneous Liquid Water Content for each layer and whereas it varies in the bottom panels.

so we use ERO with vertical subgriding, assuming that the horizontal distribution of the LWC in each horizontal layer follows the following gamma distribution, as done in Räisänen et al. (2004):

$$f(x,k,\theta) = \frac{x^{k-1}e^{-\frac{x}{\theta}}}{\Gamma(k)\theta^k} \quad \text{for } x > 0 \quad k,\theta > 0$$

where x is the liquid water content in kg/kg, $k\theta$ is the mean of the distribution and $k\theta^2$ its variance, $\Gamma(k)$ is the gamma function, with $\mathrm{Re}(k) > 0$:

$$\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt$$

This distribution can be described by its first two moments. In addition to the first moment, which we have already assumed to be known, the second moment must therefore be specified for each horizontal layer. We have chosen not to take into account the rank correlation here, as its radiative impact was shown to be of a lesser importance for the integrated cloud albedo (Oreopoulos et al., 2012).

We generate the cloud field with LWC distributions from an atmospheric column ($Dz = 100$ m) to a sample of subcolumns with the same vertical resolution as the LES ($dz = 25$ m), and display on Figure 9 the LWC of both scenes' cloudy subcolumns after they have been sorted along their vertical Liquid Water Path (LWP, bottom panels). The equivalent generation with no horizontal heterogeneity of the LWC is shown as a comparison in the top panels. When using LWC distributions, the generated subcolumns shows the same characteristics than the LES: a lot of subcolumns with a small LWP, as well as a LWP increasing with the altitude, and a small number of subcolumns with a high amount of LWP. The generated subcolumns shows demarcations every 100 m that are coming from the coarse vertical resolution of the atmospheric column because the profile $(C_v)_z$ and the LWC

properties are assumed to be constant in each coarse horizontal layer. The LWC heterogeneity also causes more disparity in the LWC values, especially high values, which are smoothed out in the homogeneous plots.

We then quantify the impact of the LWC horizontal distribution on radiative properties. To do so we look at the relative difference of cloud albedo between LES simulations with the exact LWC heterogeneity and their ERO generations with and without LWC heterogeneity. They were generated from the coarse resolution $Dz = 100$ m to the LES vertical resolution $dz = 25$ m like done in Section 3, for the two cases ARM and BOMEX. Introducing LWC horizontal distributions significantly improves the cloudy albedo: the mean relative difference with that of the LES with exact LWC goes from 8.5% to 2.4% for ARM and from 12.7% to 2% for BOMEX. Comparing the LES with exact LWC and their homogeneous versions we find the scenes without LWC horizontal heterogeneity are $\simeq$10% brighter, which confirms the previous findings of Barker et al. (2003), Wu and Liang (2005), and Shonk and Hogan (2010).

Our method is able to reproduce the known impact of LWC horizontal heterogeneity, which is comparable to the impact of the subgrid vertical heterogeneity of the cloud fraction, discussed in Section 4.3.

6. Summary and Conclusion

In this paper we presented a method based on the ERO assumption that allows to statistically represent the vertical structure of cloud scenes at different vertical resolutions. We focus on low-level clouds and show that a single value of the overlap parameter, a fundamental parameter of ERO that is directly related to the decorrelation length, is sufficient to represent the whole cloud scene.

We propose an algorithm to generate the cloud fraction on a high resolution vertical grid for an ensemble of subcolumns using a single low resolution atmospheric column and either the cloud cover or the overlap parameter. Compared to reference LES simulations, the generated cloud scenes show a correct representation of both the downward and upward cumulative cloud fractions. We suggest that they are simple diagnostics that would usefully complement the usual cloud fraction vertical profile when comparing models with observations or when developing models. The generated cloudy albedos are very close to the ones of the original LES cloud scenes, with only a 2% relative error for the best samples.

To avoid having to generate the cloud fraction profile on a high resolution vertical grid, we investigate how to represent both the subgrid variability within coarse layers and the overlap of these coarse layers to ensure correct values of cloud cover and cloud albedo. We demonstrate that, depending on how the subgrid variability is represented, the decorrelation length used to overlap the coarse layers may be highly dependent on their vertical resolution. However, we show that the subgrid variability and the interlayer overlap can be defined in such a way to define a decorrelation length almost independent of the resolution.

We also demonstrate that the decorrelation lengths obtained from remote sensing depend on the vertical resolution of the instruments. For a same cloud scene, the decorrelation length obtained from an instrument with a vertical resolution of 200 m can be two times larger that the one obtained with an instrument with a vertical resolution of 25 m. This may partly explain why the decorrelation lengths obtained by the studies using CloudSat observations are about 7 times larger that those obtained from high resolution models. If the decorrelation length can take into account the distance between cloudy layers to compute the overlap parameters, the thickness of the layers also has to be taken into account when estimating decorrelation lengths, as well as whether the cloud fractions are based on volume or area. We provide a framework that allows to go from one vertical resolution to an other, and this needs more investigations, especially to be adapted to the varying vertical resolutions of GCMs.

To our best knowledge, most current atmospheric models assume a maximum-random overlap of cloud layers or a ERO with a quite large decorrelation length ($\simeq$2–3 km). Those overlap assumptions can lead to an underestimation of the cloud cover by a factor up to two, at least for low-level clouds, and therefore explain a significant part of the underestimation of these clouds that is identified in current climate models (Konsta et al., 2022). Atmospheric models also neglect the subgrid vertical variability of the cloud fraction. This can be explained by the fact that in global models, the layers' horizontal extent is about a hundred times larger than their height ($Dz \simeq 200$ m and $Dx \simeq Dy \simeq$ 20–100 km). But this shape ratio corresponds to the grid's layers, not to the numerous clouds they contain. Their complex geometries create subgrid vertical variability and we show it matters, even at the vertical scale of the layers. This first impacts the areal cloud fraction, that can be underestimated by $\sim$40% when the subgrid heterogeneity is not taken into account. Also, for a low-level cloud scene with a given cloud cover and cloud water path, the cloud albedo can change by about 20% according to what assumptions is made to represent the subgrid vertical variability of the cloud

fraction. A better consideration of subgrid heterogeneity and cloud overlap in the models should allow this bias to be reduced, but would also require a significant revision of the amount of condensed water so that the global albedo does not change too much. This would contribute to reduce the current "too few too bright" bias.

Further work would be needed to extrapolate the results of this work to other types of clouds, especially deep convective clouds, as well as ice and mixed-phase clouds. As we focused on the vertical structure of clouds within the plan parallel approximation, we have not taken into account the solar angle or 3D radiative effects. We computed that averaged over a whole day, the relative 3D effects on the SW cloud albedo are about 7%–18% for the cases used in this study. Further work would be needed to link ERO with a 3D representation of clouds.

Appendix A: Implementation and Difference Between ERO and Räisänen's Cloud Generating Algorithm

For a cloudy block that extends continuously between the vertical layers $[k_{base}, k_{top}]$ (with $\#\left(\left[k_{base}, k_{top}\right]\right) = \mathcal{N}$ our algorithm works as follows:

We generate a sample of N_s subcolumns. The $N_s \times \mathcal{N}$ different cells of this sample are represented by the indices $i \in [1, N_s]$ and $k \in [k_{base}, k_{top}]$. Starting from the top of each subcolumn, the algorithm computes for each cell the cloud state $S_{i,k} \in \{0, 1\}$, which corresponds to whether the cell is cloudy or not, as well as the Liquid Water Content LWC.

For the top cell of the subcolumn i, $S_{i,k_{top}}$ is computed as:

$$S_{i,k_{top}} = \begin{cases} 0 & \text{for} \quad RN1_{i,k_{top}} \leq 1 - C_{k_{top}} \quad (\text{clear}) \\ 1 & \text{for} \quad RN1_{i,k_{top}} > 1 - C_{k_{top}} \quad (\text{cloudy}) \end{cases} \quad i \in [1, N_s] \tag{A1}$$

where $RN1$ are random numbers evenly distributed on [0,1]. Working its way down, the algorithm computes the next cloud states, as follows, for each cell (i, k): let $RN2_{i,k}$ be new random numbers evenly distributed on [0,1].

- **maximum overlap:** if $RN2_{i,k} < \alpha$, the cell is in maximum overlap with the one above $(i, k-1)$. Its cloud state $S_{i,k}$ is computed as:

$$S_{i,k} = c_{i,k-1}(1|1)_{\max} + (1 - C_{i,k-1})(1|0)_{\max}$$

where $(S_k|S_{k-1})_{\max}$ are booleans computed according to the transition probabilities $P_{max}(S_k = s_k|S_{k-1} = s_{k-1})$. The probability $P_{max}(S_k = 1|S_{k-1} = 1)$ is given by Equation 8, and $P_{max}(S_k = 1|S_{k-1} = 0)$ by Equations 2 and 6:

$$P_{\max}(S_k = 1|S_{k-1} = 0) = 1 - P_{\max}(S_k = 0|S_{k-1} = 0) = \frac{\max(C_{k-1}, C_k)}{1 - C_{k-1}} \tag{A2}$$

- **random overlap:** if $RN2_{i,k} > \alpha$, it's in random overlap with the cell above. Its cloud state $S_{i,k}$ is computed as:

$$S_{i,k} = (1|1)_{rand} = (1|0)_{rand}$$

where $(S_k|S_{k-1})_{rand}$ are booleans computed with the transition probability P_{rand} defined by Equation 5.

After this we have generated a cloud field with a cloud cover of CC, with a standard deviation decreasing as $1/\sqrt{N_s}$, and with conservation of the initial cloud fraction C_k, $k \in [k_{base}, k_{top}]$.

This algorithm is mainly based on Räisänen et al. (2004). The main difference between those two algorithms is about the generation on random numbers. When generating the cloud fraction (as well as the cloud condensate amount) of a given cell k, Räisänen generator computes $x_k \in [0, 1]$ to compare it to the cloud fraction of the layer C_k and decide whether the cell is cloudy or not. The computation to get x_k is:

$$x_k = \begin{cases} x_{k-1}, & \text{for RN2}_k \leq \alpha_{k-1,k} \\ \text{RN3}_k, & \text{for RN2}_k > \alpha_{k-1,k} \end{cases} \tag{A3}$$

where $\alpha_{k-1,k}$ is the overlap parameter between layers k and $k-1$, and $RN2$ and $RN3$ are two random numbers evenly distributed between 0 and 1.

In the first case, the two cells are in maximum overlap and in the second one they are in random overlap, a new independent random number being drawn. With only two layers our method is equivalent, but for more than two layers, Räisänen's method can create correlation on the whole vertical subcolumn being generated, as the same random number can be kept for many different cells.

By computing directly the transition probabilities to generate the cloud fraction of a cell ($P_{max}(1|1)$, $P_{max}(1|0)$, $P_{rand}(1|1)$, $P_{rand}(1|0)$), and by using a different random number every time it is needed, we conserve the cloud fraction without creating this correlation between the layers.

Data Availability Statement

A repository containing the scripts for the ERO algorithm presented in this paper is pulished on GitHub at https://github.com/raphleb/ERO.git (Lebrun, 2022). The software htrdr (version 0.9.1) used in this paper for the radiative computations is available at https://gitlab.com/meso-star/htrdr (Forest et al., 2023).

Acknowledgments
Our many thanks go to Céline Cornet and Frédéric Szczap for insightful discussions about this work. We thank the two anonymous reviewers for their comments and feedback, thanks to which the originally submitted manuscript was improved. We acknowledge support from the Agence Nationale de la Recherche (ANR, grants MCG-RAD ANR-18-CE46-0012) and the Centre National d'Études Spatiales (CNES, project EMC-Sat).

References

Astin, I., & Di Girolamo, L. (2014). Technical note: The horizontal scale dependence of the cloud overlap parameter α. *Atmospheric Chemistry and Physics*, *14*(18), 9917–9922. https://doi.org/10.5194/acp-14-9917-2014

Barker, H. W. (2008a). Overlap of fractional cloud for radiation calculations in GCMs: A global analysis using CloudSat and CALIPSO data. *Journal of Geophysical Research*, *113*(D8), D00A01. https://doi.org/10.1029/2007JD009677

Barker, H. W. (2008b). Representing cloud overlap with an effective decorrelation length: An assessment using CloudSat and CALIPSO data. *Journal of Geophysical Research*, *113*(D24), D24205. https://doi.org/10.1029/2008JD010391

Barker, H. W., & Räisänen, P. (2005). Radiative sensitivities for cloud structural properties that are unresolved by conventional GCMs. *Quarterly Journal of the Royal Meteorological Society*, *131*(612), 3103–3122. https://doi.org/10.1256/qj.04.174

Barker, H. W., Stephens, G. L., & Fu, Q. (1999). The sensitivity of domain-averaged solar fluxes to assumptions about cloud geometry. *Quarterly Journal of the Royal Meteorological Society*, *125*(558), 2127–2152. https://doi.org/10.1002/qj.49712555810

Barker, H. W., Stephens, G. L., Partain, P. T., Bergman, J. W., Bonnel, B., Campana, K., et al. (2003). Assessing 1D atmospheric solar radiative transfer models: Interpretation and Handling of unresolved clouds. *Journal of Climate*, *16*(16), 2676–2699. https://doi.org/10.1175/1520-0442(2003)016⟨2676:ADASRT⟩2.0.CO;2

Bergman, J. W., & Rasch, P. J. (2002). Parameterizing vertically coherent cloud distributions. *Journal of the Atmospheric Sciences*, *59*(14), 2165–2182. https://doi.org/10.1175/1520-0469(2002)059⟨2165VCCD⟩2.0.CO;2

Bodas-Salcedo, A., Webb, M. J., Bony, S., Chepfer, H., Dufresne, J.-L., Klein, S. A., et al. (2011). COSP: Satellite simulation software for model assessment. *Bulletin of the American Meteorological Society*, *92*(8), 1023–1043. https://doi.org/10.1175/2011BAMS2856.1

Bony, S., & Dufresne, J.-L. (2005). Marine boundary layer clouds at the heart of tropical cloud feedback uncertainties in climate models. *Geophysical Research Letters*, *32*(20), L20806. https://doi.org/10.1029/2005GL023851

Brooks, M. E., Hogan, R. J., & Illingworth, A. J. (2005). Parameterizing the difference in cloud fraction defined by area and by volume as observed with radar and lidar. *Journal of the Atmospheric Sciences*, *62*(7), 2248–2260. https://doi.org/10.1175/JAS3467.1

Brown, A. R., Cederwall, R. T., Chlond, A., Duynkerke, P. G., Golaz, J.-C., Khairoutdinov, M., et al. (2002). Large-eddy simulation of the diurnal cycle of shallow cumulus convection over land. *Quarterly Journal of the Royal Meteorological Society*, *128*(582), 1075–1093. https://doi.org/10.1256/003590002320373210

Del Genio, A. D., Yao, M.-S., Kovari, W., & Lo, K. K.-W. (1996). A prognostic cloud water parameterization for global climate models. *Journal of Climate*, *9*(2), 270–304. https://doi.org/10.1175/1520-0442(1996)009⟨0270:APCWPF⟩2.0.CO;2

Di Giuseppe, F., & Tompkins, A. M. (2015). Generalizing cloud overlap treatment to include the effect of wind shear. *Journal of the Atmospheric Sciences*, *72*(8), 2865–2876. https://doi.org/10.1175/JAS-D-14-0277.1

Forest, V., Caliot, C., Eymet, V., Sans, M., Schoetter, R., & Villefranque, N. (2023). htrdr, the Monte-Carlo radiative transfer simulator. February 27, 2023 Release (Version 0.9.1). [Software]. GPLv3+. Retrieved from https://gitlab.com/meso-star/htrdr

Geleyn, J., & Hollingsworth, A. (1979). An economical analytical method for the computation of the interaction between scattering and line absorption of radiation. *Beiträge zur Physik der Atmosphäre*.

Hogan, R. J., & Illingworth, A. J. (2000). Deriving cloud overlap statistics from radar. *Quarterly Journal of the Royal Meteorological Society*, *126*(569), 2903–2909. https://doi.org/10.1002/qj.49712656914

Hogan, R. J., & Shonk, J. K. P. (2013). Incorporating the effects of 3D radiative transfer in the presence of clouds into two-stream multilayer radiation schemes. *Journal of the Atmospheric Sciences*, *70*(2), 708–724. https://doi.org/10.1175/JAS-D-12-041.1

Jakob, C., & Klein, S. A. (1999). The role of vertically varying cloud fraction in the parametrization of microphysical processes in the ECMWF model. *Quarterly Journal of the Royal Meteorological Society*, *125*(555), 941–965. https://doi.org/10.1002/qj.49712555510

Jing, X., Zhang, H., Peng, J., Li, J., & Barker, H. W. (2016). Cloud overlapping parameter obtained from CloudSat/CALIPSO dataset and its application in AGCM with McICA scheme. *Atmospheric Research*, *170*, 52–65. https://doi.org/10.1016/j.atmosres.2015.11.007

Jouhaud, J., Dufresne, J.-L., Madeleine, J.-B., Hourdin, F., Couvreux, F., Villefranque, N., & Jam, A. (2018). Accounting for vertical subgrid-scale heterogeneity in low-level cloud fraction parameterizations. *Journal of Advances in Modeling Earth Systems*, *10*(11), 2686–2705. https://doi.org/10.1029/2018MS001379

Konsta, D., Dufresne, J.-L., Chepfer, H., Vial, J., Koshiro, T., Kawai, H., et al. (2022). Low-level marine tropical clouds in six CMIP6 models are too few, too bright but also too compact and too homogeneous. *Geophysical Research Letters*, *49*(11), e2021GL097593. https://doi.org/10.1029/2021GL097593

Koren, I., Oreopoulos, L., Feingold, G., Remer, L. A., & Altaratz, O. (2008). How small is a small cloud? *Atmospheric Chemistry and Physics*, *8*(14), 3855–3864. https://doi.org/10.5194/acp-8-3855-2008

Lac, C., Chaboureau, J.-P., Masson, V., Pinty, J.-P., Tulet, P., Escobar, J., et al. (2018). Overview of the Meso-NH model version 5.4 and its applications. *Geoscientific Model Development*, *11*(5), 1929–1969. https://doi.org/10.5194/gmd-11-1929-2018

Lafore, J. P., Stein, J., Asencio, N., Bougeault, P., Ducrocq, V., Duron, J., et al. (1998). The Meso-NH atmospheric simulation system. Part I: Adiabatic formulation and control simulations. *Annales Geophysicae*, *16*(1), 90–109. https://doi.org/10.1007/s00585-997-0090-6

Larson, V. E., Golaz, J.-C., & Cotton, W. R. (2002). Small-scale and mesoscale variability in cloudy boundary layers: Joint probability density functions. *Journal of the Atmospheric Sciences*, *59*(24), 3519–3539. https://doi.org/10.1175/1520-0469(2002)059⟨3519:SSAMVI⟩2.0.CO;2

Lebrun, R. (2022). ERO, exponential-random overlap sampler. December 22, 2022. [Software]. Retrieved from https://github.com/raphleb/ERO.git

Li, J., Huang, J., Stamnes, K., Wang, T., Lv, Q., & Jin, H. (2015). A global survey of cloud overlap based on calipso and cloudsat measurements. *Atmospheric Chemistry and Physics*, *15*(1), 519–536. https://doi.org/10.5194/acp-15-519-2015

Mace, G. G., & Benson-Troth, S. (2002). Cloud-layer overlap characteristics derived from long-term cloud radar data. *Journal of Climate*, *15*(17), 2505–2515. https://doi.org/10.1175/1520-0442(2002)015⟨2505:CLOCDF⟩2.0.CO;2

Naud, C. M., Genio, A. D., Mace, G. G., Benson, S., Clothiaux, E. E., & Kollias, P. (2008). Impact of dynamics and atmospheric state on cloud vertical overlap. *Journal of Climate*, *21*(8), 1758–1770. https://doi.org/10.1175/2007JCLI1828.1

Neggers, R. A. J., Duynkerke, P. G., & Rodts, S. M. A. (2003). Shallow cumulus convection: A validation of large-eddy simulation against aircraft and landsat observations. *Quarterly Journal of the Royal Meteorological Society*, *129*(593), 2671–2696. https://doi.org/10.1256/qj.02.93

Neggers, R. A. J., Heus, T., & Siebesma, A. P. (2011). Overlap statistics of cumuliform boundary-layer cloud fields in large-eddy simulations. *Journal of Geophysical Research*, *116*(D21). https://doi.org/10.1029/2011JD015650

Neggers, R. A. J., Jonker, H. J. J., & Siebesma, A. P. (2003). Size statistics of cumulus cloud populations in large-eddy simulations. *Journal of the Atmospheric Sciences*, *60*(8), 1060–1074. https://doi.org/10.1175/1520-0469(2003)60⟨1060:SSOCCP⟩2.0.CO;2

Oreopoulos, L., Cho, N., & Lee, D. (2022). Revisiting cloud overlap with a merged dataset of liquid and ice cloud extinction from CloudSat and CALIPSO. *Frontiers in Remote Sensing*, *3*. https://doi.org/10.3389/frsen.2022.1076471

Oreopoulos, L., Cho, N., Lee, D., Lebsock, M., & Zhang, Z. (2022). Assessment of two stochastic cloud subcolumn generators using observed fields of vertically resolved cloud extinction. *Journal of Atmospheric and Oceanic Technology*, *39*(8), 1229–1244. https://doi.org/10.1175/JTECH-D-21-0166.1

Oreopoulos, L., Lee, D., Sud, Y. C., & Suarez, M. J. (2012). Radiative impacts of cloud heterogeneity and overlap in an atmospheric general circulation model. *Atmospheric Chemistry and Physics*, *12*(19), 9097–9111. https://doi.org/10.5194/acp-12-9097-2012

Oreopoulos, L., & Norris, P. M. (2011). An analysis of cloud overlap at a midlatitude atmospheric observation facility. *Atmospheric Chemistry and Physics*, *11*(12), 5557–5567. https://doi.org/10.5194/acp-11-5557-2011

Pincus, R., Barker, H. W., & Morcrette, J.-J. (2003). A fast, flexible, approximate technique for computing radiative transfer in inhomogeneous cloud fields. *Journal of Geophysical Research*, *108*(D13). https://doi.org/10.1029/2002JD003322

Pincus, R., Hannay, C., Klein, S. A., Xu, K.-M., & Hemler, R. (2005). Overlap assumptions for assumed probability distribution function cloud schemes in large-scale models. *Journal of Geophysical Research*, *110*(D15), D15S09. https://doi.org/10.1029/2004JD005100

Räisänen, P., Barker, H. W., Khairoutdinov, M. F., Li, J., & Randall, D. A. (2004). Stochastic generation of subgrid-scale cloudy columns for large-scale models. *Quarterly Journal of the Royal Meteorological Society*, *130*(601), 2047–2067. https://doi.org/10.1256/qj.03.99

Shonk, J. K. P., & Hogan, R. J. (2008). Tripleclouds: An efficient method for representing horizontal cloud inhomogeneity in 1D radiation schemes by using three regions at each height. *Journal of Climate*, *21*(11), 2352–2370. https://doi.org/10.1175/2007JCLI1940.1

Shonk, J. K. P., & Hogan, R. J. (2010). Effect of improving representation of horizontal and vertical cloud structure on the Earth's global radiation budget. Part II: The global effects. *Quarterly Journal of the Royal Meteorological Society*, *136*(650), 1205–1215. https://doi.org/10.1002/qj.646

Siebesma, A. P., Bretherton, C. S., Brown, A., Chlond, A., Cuxart, J., Duynkerke, P. G., et al. (2003). A large eddy simulation intercomparison study of shallow cumulus convection. *Journal of the Atmospheric Sciences*, *60*(10), 1201–1219. https://doi.org/10.1175/1520-0469(2003)60⟨1201:ALESIS⟩2.0.CO;2

Sulak, A. M., Calabrase, W. J., Ryan, S. D., & Heus, T. (2020). The contributions of shear and turbulence to cloud overlap for cumulus clouds. *Journal of Geophysical Research*, *125*(10), e2019JD032017. https://doi.org/10.1029/2019JD032017

Swales, D. J., Pincus, R., & Bodas-Salcedo, A. (2018). The cloud feedback model intercomparison project observational simulator package: Version 2. *Geoscientific Model Development*, *11*(1), 77–81. https://doi.org/10.5194/gmd-11-77-2018

Tompkins, A. M., & Di Giuseppe, F. (2007). Generalizing cloud overlap treatment to include solar Zenith angle effects on cloud geometry. *Journal of the Atmospheric Sciences*, *64*(6), 2116–2125. https://doi.org/10.1175/JAS3925.1

Tompkins, A. M., & Di Giuseppe, F. (2015). An interpretation of cloud overlap statistics. *Journal of the Atmospheric Sciences*, *72*(8), 2877–2889. https://doi.org/10.1175/JAS-D-14-0278.1

van Zanten, M. C., Stevens, B., Nuijens, L., Siebesma, A. P., Ackerman, A. S., Burnet, F., et al. (2011). Controls on precipitation and cloudiness in simulations of trade-wind cumulus as observed during RICO. *Journal of Advances in Modeling Earth Systems*, *3*(2). https://doi.org/10.1029/2011MS000056

Villefranque, N., Blanco, S., Couvreux, F., Fournier, R., Gautrais, J., Hogan, R. J., et al. (2021). Process-based climate model development harnessing machine learning: III. The representation of cumulus geometry and their 3D radiative effects. *Journal of Advances in Modeling Earth Systems*, *13*(4), e2020MS002423. https://doi.org/10.1029/2020MS002423

Villefranque, N., Fournier, R., Couvreux, F., Blanco, S., Cornet, C., Eymet, V., et al. (2019). A path-tracing Monte Carlo library for 3-D radiative transfer in highly resolved cloudy atmospheres. *Journal of Advances in Modeling Earth Systems*, *11*(8), 2449–2473. https://doi.org/10.1029/2018MS001602

Willèn, U., Crewell, S., Baltink, H. K., & Sievers, O. (2005). Assessing model predicted vertical cloud structure and cloud overlap with radar and lidar ceilometer observations for the baltex bridge campaign of CLIWA-NET. *Atmospheric Research*, *75*(3), 227–255. (CLIWA-NET: Observation and Modelling of Liquid Water Clouds). https://doi.org/10.1016/j.atmosres.2004.12.008

Wu, X., & Liang, X.-Z. (2005). Radiative effects of cloud horizontal inhomogeneity and vertical overlap identified from a monthlong cloud-resolving model simulation. *Journal of the Atmospheric Sciences*, *62*(11), 4105–4112. https://doi.org/10.1175/JAS3565.1