# A consistent representation of cloud overlap and cloud subgrid vertical heterogeneity

Raphaël Lebrun[1], Jean-Louis Dufresne[1], and Najda Villefranque[2]

[1]LMD/IPSL
[2]Meteo-France

December 27, 2022

**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 LES results as reference, we propose a method based on exponential-random overlap (ERO) 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 an additional relative error of 20% in brightness, assuming the cloud cover is correct.

# A consistent representation of cloud overlap and cloud subgrid vertical heterogeneity

**Raphaël Lebrun[1], Jean-Louis Dufresne[1], 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 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 exponential-random overlap are highly dependent on the vertical resolutions of models and observations.

Corresponding author: Raphal Lebrun, `raphael.lebrun@lmd.ipsl.fr`

## 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 LES results as reference, we propose a method based on exponential-random overlap (ERO) 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 an additional 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, in the framework of commonly used overlap assumptions. When the cloud cover of the scene is known, we show that the exponential-random overlap allows a good representation of the vertical structure of clouds and of the cloud albedo. We find the decorrelation lengths 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. We present values that can be used to compute accurately the cloud cover and the cloud albedo of the studied scenes.

## 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 (Räisänen et al. (2004); Hogan and Shonk (2013)). The cloud cover and the mean optical depth of the cloudy region are inter-dependent when the profile of cloud fractions and water contents are known. They depend on how the cloud fractions overlap on the vertical: if they overlap maximally, the cloud cover will be minimum and the mean optical depth maximum, and if they overlap randomly, the cloud cover will be larger and the mean optical depth smaller.

How the cloud fraction ($CF$) of each atmospheric layer overlap with other layers has been widely studied (Geleyn and Hollingsworth (1979); Barker et al. (1999); Jakob and 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 (Hogan and Illingworth (2000); Bergman and Rasch (2002); Tompkins and Di Giuseppe (2007); Shonk and Hogan (2010)). The corresponding decorrelation length scale has been estimated from satellite radar observations (Jing et al. (2016)), in-situ observations (Mace and 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 (Pincus et al. (2005); Di Giuseppe and Tompkins (2015); 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 volume of the grid box that contains clouds, $CF_v$, but radiation is primarily sensitive to the surface cloud fraction $CF_s$ which is the relative surfacic fraction covered by clouds in a cell. Often implicitly, these two fractions are assumed to be equal, i.e. the clouds are assumed to be homogeneous on the vertical in each cell. This can seem logical on the first order given the area/depth ratio of the grid cells, 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 $CF_s$ is typically greater than $CF_v$ by about 30% (Neggers et al. (2011)). A direct consequence of not taking into account this difference is that, for a given cloud faction in volume, the surface 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 exponential-random overlap 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 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. Like done in the McICA method, we neglect the 3D effects and keep the classical plane parallel assumption (each vertical profile represents a stack of horizontally infinite and homogeneous slabs) in our 1D approach. Assuming that the volumic 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.

The manuscript is organised as follows: in Section 2, we consider the exponential-random overlap (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 alebdo 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 inter-layer 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 exponential-random overlap (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 combined cloud fraction of two adjacent cloudy layers of surfacic fractions $CF_1$ and $CF_2$ is:

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

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

$$CF_{1,2,max} = \max(CF_1, CF_2)$$

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

$$CF_{1,2,rand} = CF_1 + CF_2 - CF_1 CF_2$$

In this model, "exponential" refers to the fact that $\alpha$ 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 and 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 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 paper, the McICA framework is used to generate samples of vertical profiles. The main difference is that in the usual McICA algorithm, the profiles are generated on the vertical grid of the host model, while here we aim at generating profiles at any vertical resolution, including finer vertical resolutions.

Unless otherwise stated, in all this article, we consider a single vertical atmospheric column that consists of a cloudy block (with a strictly positive liquid water content at every level) of $\mathcal{N}$ vertical layers. From this column we assume the volume cloud fraction of each layer, $(CF_k)_{k=1...\mathcal{N}}$, is known. We consider the exponential-random overlap model (ERO) as a Markovian process and deduce the relationship between the overlap parameter $\alpha$ and the total cloud cover $CC$. 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).

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. If $\vec{\mathrm{C}} = (\mathrm{C}_k)_{k=1...\mathcal{N}}$ is the random variable representing the cloud fraction distribution of the subcolumn, with $\mathrm{C}_k \in \{0, 1\}$ (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 $\vec{\mathrm{C}} = (c_k)_{k=1...\mathcal{N}} \in [0, 1]^{\mathcal{N}}$ is given by:

$$P(\vec{\mathrm{C}}) = \prod_{k=1}^{\mathcal{N}} P(\mathrm{C}_k = c_k \mid \mathrm{C}_{k-1} = c_{k-1}) \tag{1}$$

where $C_0 = 0$ (i.e. there is no cloud above the cloud block considered here). We use the classic upper case notation $\mathrm{C}_k$ for the random variables and the lower case notation $c_k$ for their realizations.

For any level $k$ in the subcolumn, the probability to have $c_k = 1$ is the cloud fraction of the level, meaning $P(\mathrm{C}_k = 1) = CF_k$. We'll call $P(\mathrm{C}_k = c_k \mid \mathrm{C}_{k-1} = c_{k-1})$ a *transition probability*, it is the probability that in a subcolumn, layer $k$ is in the state $c_k$, knowing the layer $k-1$ is in the state $c_{k-1}$. Since $c_k$ is either 0 or 1, there are only four possible types of transition between two levels, and being able to compute their probabilities at every level gives the probability of any vertical cloud fraction distribution for the column. Moreover, for each level $k$, two out of the four transition probabilities are dependant, as a layer is either cloudy or clear sky:

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

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

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

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

where $(\mathrm{C}_k = 1 \cap \mathrm{C}_{k-1} = 1)$ is the event with both layers cloudy. If we assume an exponential-random overlap we have :

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

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

$$P_{rand}(\mathrm{C}_k = 1 \mid \mathrm{C}_{k-1} = 1) = P_{rand}(\mathrm{C}_k = 1 \mid \mathrm{C}_{k-1} = 0) = P_{rand}(\mathrm{C}_k = 1) = CF_k \tag{5}$$

The transition probability in a subcolumn of the maximum overlap can be obtained using Eq. (3): if $CF_{k-1} < CF_k$ : $P_{max}(\mathrm{C}_k = 1 \mid \mathrm{C}_{k-1} = 1) = 1$, and on the contrary if $CF_{k-1} \geq CF_k$: $P_{max}(\mathrm{C}_k = 1 \mid \mathrm{C}_{k-1} = 1) = \frac{CF_k}{CF_{k-1}}$

As a result,

$$P_{max}(\mathrm{C}_k = 1 \mid \mathrm{C}_{k-1} = 1) = \frac{\min(CF_{k-1}, CF_k)}{CF_{k-1}} \tag{6}$$

and (4) becomes :

$$P(\mathrm{C}_k = 1 \mid \mathrm{C}_{k-1} = 1) = \alpha \frac{\min(CF_{k-1}, CF_k)}{CF_{k-1}} + (1-\alpha) CF_k \tag{7}$$

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

$$P_{max}(\mathrm{C}_k = 0 \mid \mathrm{C}_{k-1} = 0) = \frac{1 - \max(CF_{k-1}, CF_k)}{1 - CF_{k-1}}$$

and therefore:

$$P(\mathrm{C}_k = 0 \mid \mathrm{C}_{k-1} = 0) = \alpha \times \frac{(1 - \max(CF_{k-1}, CF_k))}{1 - CF_{k-1}} + (1-\alpha)(1 - CF_k) \tag{8}$$

These equations and exponential-random overlap more generally are applicable only for non overcast cloudy layers (i.e. $CF \in ]0,1[$). 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 Eqs. (7), (8) and (2) we can now compute the different transition probabilities for each layer $k$, knowing $\alpha$. Then using Eq. (1) we can compute the probability to generate any vertical cloud fraction distribution for a subcolumn, for any exponential-random overlap parameter $\alpha \in [0,1]$.

### 2.2 The relationship between the overlap parameter $\alpha$ and the total 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 $\alpha$ and the total cloud cover $CC$, assuming ERO.

To obtain the formal expression of the total 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:

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

Using Eq. (8) we get:

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

Given this equation, if we know the overlap parameter $\alpha$, the total cloud cover is:

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

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

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

where

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

For a given $(CF)_{1\ldots\mathcal{N}}$ profile (with $CF_k \in ]0,1[$ for each layer) and knowing $CC$, the function $f_\emptyset$ is strictly increasing, so $f_\emptyset^{-1}$ exists. We compute $\alpha$ with a dichotomy method using a tolerance $\epsilon = 10^{-5}$.

Eq. (12) gives the expression of $\alpha$ for a given cloud cover $CC$ and cloud fraction profile $(CF)$. Eq. (10) allows us to compute $CC$ if we know the overlap parameter $\alpha$ and the profile $(CF)$. Therefore for any given profile $(CF)$ and given the ERO model, it is equivalent to know $CC$ or $\alpha$ (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 $\{\widehat{CF_k}\}_{k=1\ldots N}$, and we generate subcolumns with $n$ times more vertical levels, $\mathcal{N} = (N \times n)$. We introduce the hypothesis that at every coarse level of the atmospheric column, the volume cloud fraction is the same for all the $n$ sublayers :

$$\forall \quad l \in \mathcal{L}_k^n, \quad CF_l = \widehat{CF_k}$$

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 cell is uniform, the intralayer transition probability $P(C_l = 0|C_{l-1} = 0)$ (Eq. (8)) between layers inside the same coarse cell simplifies as :

$$P(C_l = 0|C_{l-1} = 0) = P_{intra,l} = \alpha + (1-\alpha)(1-CF_l) \tag{13}$$

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

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

Like done previously, we can compute the cloud cover generated by a given overlap parameter $\alpha$, or if the total cloud cover of the scene is known, we can inverse this equation using Eq. (12) to compute the overlap parameter $\alpha$ 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 $\alpha$ and the cloud generation

As done in many previous works such as Larson et al. (2002); R. A. J. Neggers et al. (2003); Neggers et al. (2011), we are using Large Eddy Simulations (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 ARMCu cloud case (Brown et al. (2002)) showing the development of shallow cumulus convection over land, as well as two marine, trade-winds cumulus cloud cases BOMEX (Siebesma et al. (2003)) and RICO (vanZanten et al. (2011)), and another case of continental cumulus SCMS (Neggers et al. (2003b)). For each case we use the corresponding LES results obtained with the atmospheric non-hydrostatic model MESO-NH (Lafore et al. (1998); Lac et al. (2018)), and all these simulations represent a 6.4 $km\times$ 6.4 $km\times$ 4 $km$ domain with a $dx$=$dy$=$dz$=25 $m$ resolution. For each LES simulation we coarsen it into a single atmospheric column with the same vertical resolution $dz$, or a lower vertical resolution $Dz$, as shown in Fig. 1. For each of these single columns we know, by means of the LES, the total cloud cover $CC$, as well as the cloud fraction and the liquid water content at each vertical level. 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.

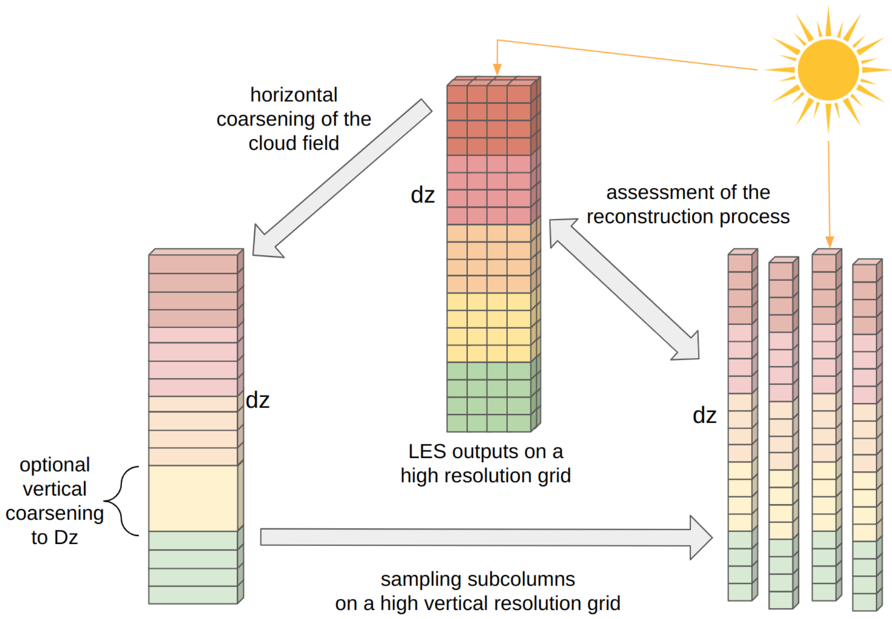


**Figure 1.** Method used to develop and assess our cloudy columns sampling. The LES cloud field of resolution $dx$=$dy$=$dz$=25 $m$ is horizontally averaged into a single column and eventually averaged vertically to a coarse resolution $Dz$>$dz$. We then sample $N_s$ subcolumns with a vertical resolution $dz$ using the ERO algorithm, and then assess the process by comparing the sample's cloud fraction profile and TOA SW cloud albedo to the ones of the original 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 $\alpha$ 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), 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 $\alpha$ used to generate this sample is computed with Eqs. (12,14) 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, $N_s \approx 6.5 \times 10^4$ subcolumns have been generated. For this number, the total 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 cloud fraction seen from above or from below at altitude $z$ varies as a function of this altitude (Fig. 2).

The blue line (Fig. 2, middle and right panels) is the cloud cover profile of the original LES, with a total cloud cover of 0.2325. The grey line is obtained using a maximum overlap assumption, and shows a total cloud cover of only $\sim$ 10%. Since the scene consists of a single cloud block, this corresponds to models using 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 total cloud cover of the LES ($\alpha = 0.921$), with a very close total 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 total cloud cover than the LES, but also a close projected cloud cover at each vertical level. The abrupt changes in the cloud cover of the sampled subcolumns are a consequence of the hyopthesis of a constant volume cloud fraction $CF_v$ in each coarse cell. For the generation without vertical subgriding of the previous section ($Dz$=$dz$=25 $m$), the vertical distribution of the cloud cover is almost indiscernible to that of the LES (not shown).

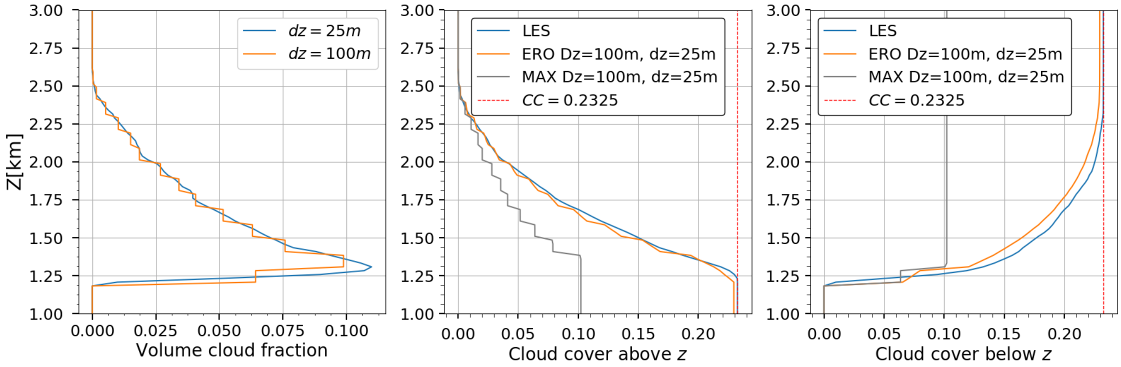


**Figure 2.** Vertical distribution of the volume cloud fraction (left), of the total cloud cover above (middle) and below (right) altitude z. The former is the projected total cloud cover of all the clouds between the top of the domain and altitude $z$, the latter is the projected cloud cover between the bottom of the domain and altitude $z$. On the middle and right panels are compared the profiles from the LES (blue) and those obtained with two overlap models : maximum overlap (grey) and ERO (orange). The red dot line shows the total 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 ARMCu cloud case (time step $h$=10).

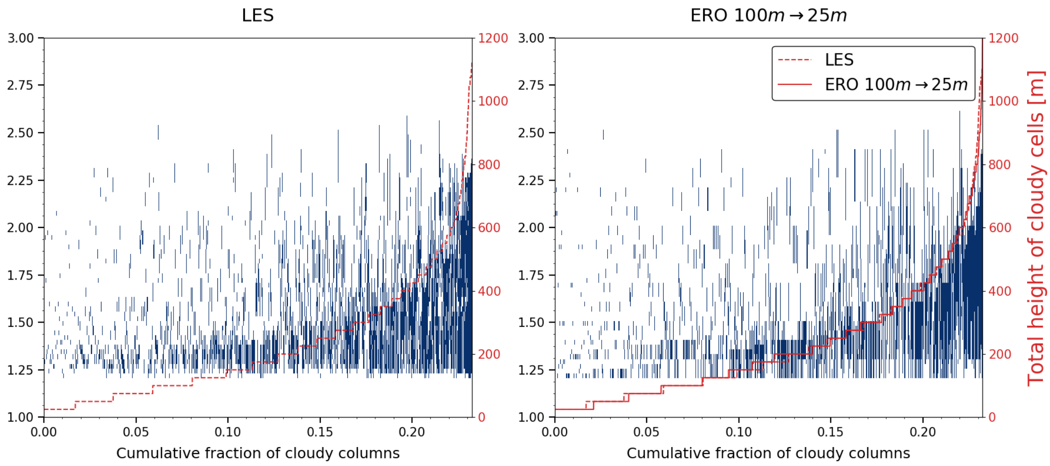


**Figure 3.** The cloudy subcolumns of the LES scene (left) are sorted along the number of cloudy cells in each subcolumns (dashed red). On the right the cloudy subcolumns out of a $N_s$≈6.5×10$^4$ sample of subcolumns generated with ERO sorted in the same way (solid red for the number of cloudy cells 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 ARMCu case.

To go further, Fig. 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. The vertical distribution of cloudy cells are very close, it shows the ERO generation not only reproduces the total cloud cover of the original scene, but also the distribution of cumulative cloud fraction.

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 Independant Column Approximation -or ICA - (Pincus et al. (2003)). Fig. 4 shows the cloud albedo of different sampling hypotheses, of the original LES scenes, as well as the total albedo of the scenes, and their total 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 total cloud covers.

The maximum overlap assumption (grey) shows a much higher cloud albedo since it produces cloud scenes with less total cloud cover and hence brighter clouds. 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 starting with a 25 $m$ vertical resolution is $\sim 1.5\%$ and only of $\sim 2.5\%$ when starting with a 200 $m$ vertical resolution, for the simulation hours $[6, 12]$.

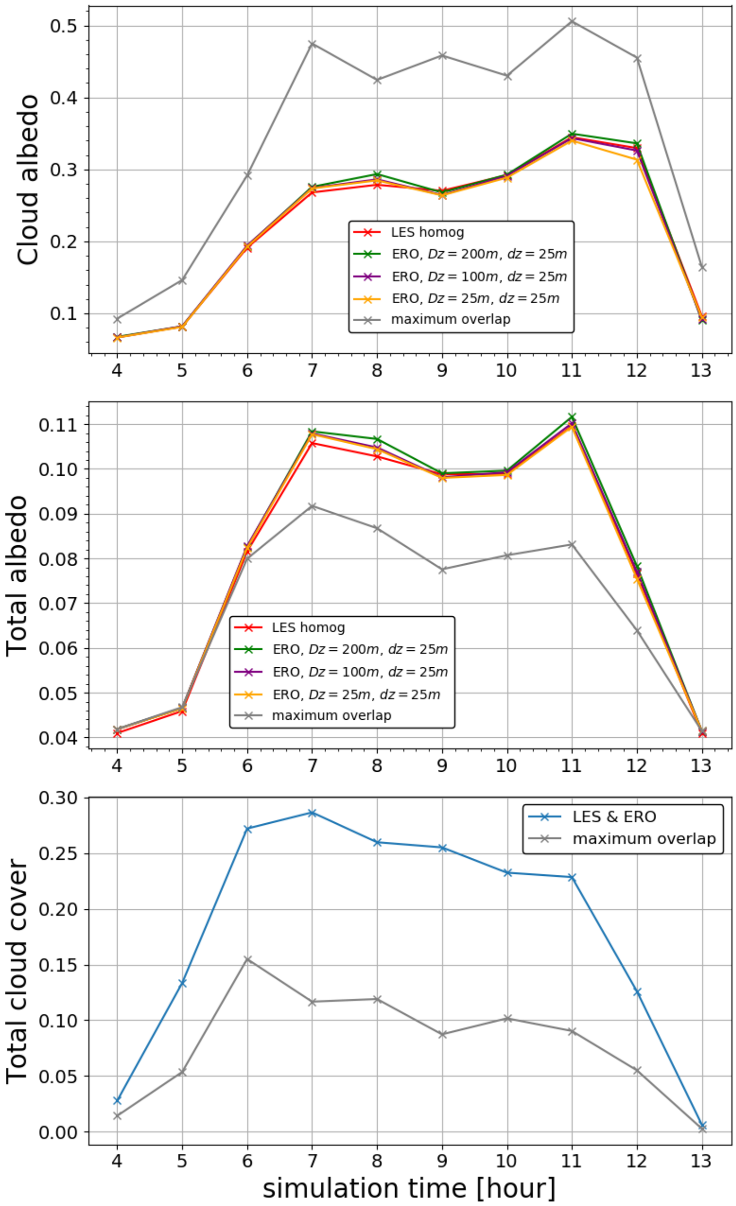


**Figure 4.** Cloud albedo (top panel), total albedo (middle panel) and total cloud cover (lower panel) for the LES (in red), for ERO with different coarse resolutions $Dz$ and for maximum overlap with the coarse resolution $Dz$=100 $m$ (in grey). The albedo of each scene is computed using a Monte-Carlo algorithm under the Indenpendant Column Approximation, for the ARMCu 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 LWC is homogeneous at each vertical level. For each computation, $10^6$ realisations were made, with a Monte-Carlo standard deviation of the cloud albedo of $10^{-6}$.

### 3.2 Analysis of the overlap parameter $\alpha$

In Section 2 we established the relationship between the overlap parameter $\alpha$ and the total cloud cover $CC$ and used it in 3.1 to determine $\alpha$ from the total cloud cover $CC$ diagnosed from LES results. In this section we analyze the overlap parameters computed this way and compare them to the values given by 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 $L_\alpha$ are usually related to each other via the following relation ( Hogan and Illingworth (2000); Bergman and Rasch (2002); Mace and Benson-Troth (2002)) :

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

If the decorrelation length $L_\alpha$ 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 and Illingworth (2000); Oreopoulos and Norris (2011)), according to Eq. (16). Fig. 5 shows the variations of the overlap parameters $\alpha$ computed at different times of the day of the ARMCu 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 Eq. (16). This profile was obtained by computing the mean overlap parameter for each possible separation distance by using $CF_s = \alpha CF_{max} + (1-\alpha) CF_{rand}$. The overlap parameter $\alpha_{25,Dz}$ corresponds to the overlap parameter computed using Eq. (14) to reproduce the total cloud cover $CC$ with vertical subgriding from a vertical resolution $Dz$=100 $m$ to $dz$=25 $m$. 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 $m$. 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 $\sim$ 300 $m$) are comparable to those computed in the litterature with similar LES simulations (Neggers et al. (2011); Sulak et al. (2020); Villefranque et al. (2021)). The difference with decorrelation lengths in the litterature that take into account the overlap of whole atmospheric columns in global model is further discussed in Section 5.

We have also computed the overlap parameter $\alpha$ 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 ARMCu($h$=10) when taking into account the 45 clouds that account for 99% of the total 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 $m$.

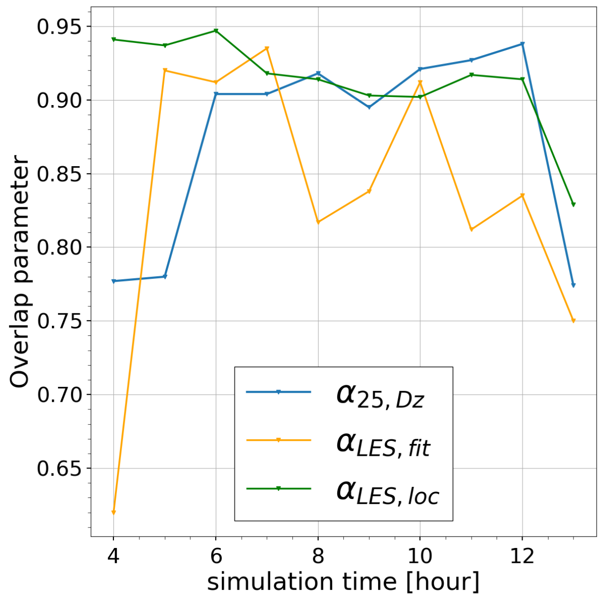


**Figure 5.** Overlap parameters computed with three different methods (see text) at each time step of the LES simulations. The data used are the ARMCu cloud fields.

## 4 Using ERO to model subgrid properties and overlap coarse vertical layers

To summarize the previous section, if we know the overlap parameter $\alpha_{25,Dz}$ or the total cloud cover of the scene, and its volume cloud fraction $CF$ for every cloudy layer of thickness $Dz$ as well as the LWC mean value, 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 ARMCu 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 total 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 surface cloud fraction $CF_s$ and the volume cloud fraction $CF_v$ (Genio et al. (1996); Jouhaud et al. (2018)). $CF_v$ represents the volume fraction of the layer that contains clouds ( i.e. where liquid or solid water particules are present), whereas $CF_s$ represents the surface fraction of the layer covered by clouds when looking from above or below. In other words, $CF_s$ is the vertical projection of $CF_v$, and it is $CF_s$ that is used by radiation codes in GCMs and teledetection.

At the LES grid scale, we have assumed that a grid cell is either clear or cloudy, and therfore $CF_v$=$CF_s$. This is no longer the case on a coarse grid, and ERO can be used to compute $CF_s$ in a coarse layer of an atmospheric column, knowing $CF_v$.

For that we consider an atmospheric cloudy column of coarse vertical resolution $Dz{=}n{\times}dz$. If $CF_v$ is known and vertically uniform within each coarse layer, we are back in the configuration we were in Section 3 when using subgriding, with $CF_{v,k}{=}\widehat{CF}_k$. We can then compute the subgrid surface cloud fraction $CF_{s,sg,k}$ as the total cloud cover of a single coarse layer, by using Eq. (14), but setting to zero the volume cloud fractions above and below the coarse layer considered ($N{=}1$) :

$$CF_{s,sg,k} = 1 - (1 - CF_{v,k})(\alpha_{sg} + (1 - \alpha_{sg})(1 - CF_{v,k}))^{n-1} \tag{17}$$

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

The top panels of Fig. 6 show the profile of $CF_s$ obtained using the LES original data, using Eq. (17), and also assuming maximum overlap within each layer, for two coarse resolutions (left panel at $Dz{=}100$ $m$ and right panel $Dz{=}200$ $m$). When using Eq. (17), two slightly different values of $\alpha$ are used for $Dz{=}100$ $m$ ($\alpha_{25,100}{=}0.921$) and $Dz{=}200$ $m$ ($\alpha_{25,200}{=}0.911$), to ensure that the total cloud cover is the same. The maximum overlap assumption (grey) does a poor job representing the surface cloud fraction profile, and leads to a relative error of 30% to 50%. It shows the error made when neglecting subgrid variability, i.e. assumming $CF_s{=}CF_v$ on the coarse grid. For this assumption, the coarser the vertical resolution, the larger the error. Using Eq. (17) allows a better representation of the surface cloud fractions, even if a substantial error remains. For all methods, the largest error corresponds to the lower layer which is the bottom of the cloud layer. On this layer the volume cloud fraction $CF_v$ decreases steeply, which makes the hypothesis of a constant $CF_v$ inaccurate.

To go further we also compare the performance of Eq. (17) with that of other references in the litterature. Neggers et al. (2011) and Jouhaud et al. (2018) have both been developed using LES data of small cumulus with $CF_v \approx 0.1$, including the ARMCu and BOMEX cases, and are therefore comparable to our method. Brooks et al. (2005) develops a lidar and radar-based parametrization of $CF_s$ using $CF_v$, with the possibility 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 $CF_s$ of the LES.

Our approach favours an accurate cloud cover on the whole vertical extent of the cloud layer. Results show that with this approach we tend to underestimate the surface cloud fraction of the coarse layers. This is because the overlap parameter $\alpha$ has been computed to match the total cloud cover of the whole scene, not the surface cloud fraction $CF_s$ of each coarse layer. When only used for the subgrid scale it creates too small a surface cloud fraction. This underestimation is still much smaller than when considering maximum overlap. The gap in surface cloud fraction caused by using our method is similar to those caused by other approximations of the litterature, but whith an opposite sign in the difference. Our underestimation of $(CF_s)_z$ was already visible in Fig. 2 on the panel showing "cloud cover above $z$". The only difference between using subgriding or not is the hypothesis $CF_{vol}{=}cst$ in each coarse layers, so we can conclude than the underestimation of our method comes from this hypothesis.

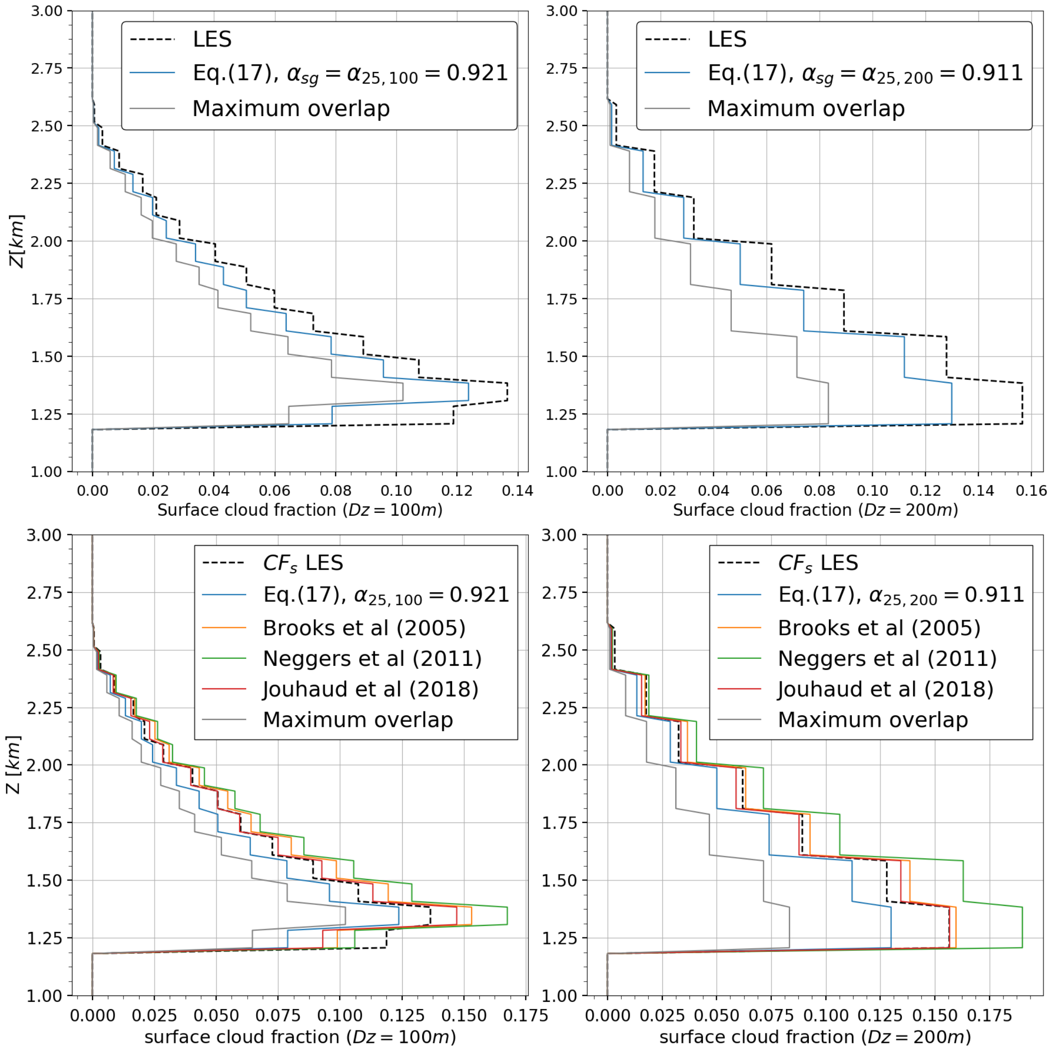


**Figure 6.** Vertical distribution of the surface cloud fraction $(CF_s)_z$ obtained with LES full resolution results or with different approximations with a coarse vertical resolution of 100 $m$ (left panels) or 200 $m$ (right panels). The top panels compare the LES (dashed black) with ERO using Eq. (17) and $\alpha_{sg}$=$\alpha_{25,Dz}$ (blue) as well as the maximum overlap sample (grey). The bottom panels also compare Eq. (17) with other parametrizations found in the litterature. The cloud case is ARMCu ($h$=10).

## 4.2 Interlayer overlap

We now consider that the vertical profile of the surface cloud fraction $(CF_{s,sg})_z$ 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 total cloud cover $CC$. To compute the subgrid surface cloud fraction profile $(CF_{s,sg})_z$ in the previous section, we were using the first part of Eq. (14), which represents the subgrid overlap. We here use the second part of the equation, which represents the interlayer overlap, using the unknown interlayer overlap $\alpha_{inter}$.

This corresponds to using Eq. (10) on the coarse grid with $(CF_{s,sg})_z$ to produce the total cloud cover:

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

The overlap parameter $\alpha_{inter}$ can be computed as in the previous sections, by inverting Eq. (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 Eq. (12), the subgrid cloud fractions $(CF_{s,sg})_z$ using Eq. (17) with $\alpha_{sg}$=$\alpha_{25,Dz}$, and then the overlap parameter $\alpha_{inter}$ using Eqs. (18,19) in order to overlap these coarse layers to produce the total cloud cover $CC$. The corresponding decorrelation length can be computed with Eq. (16) and $Dz$ as the separation distance. However, at this stage, there is no evidence of a formal link between these two overlap parameters or decorrelation lengths, or of a dependence to the vertical resolution. In any case, we have not found one.

We find that $\alpha_{25,Dz}$ and the corresponding decorrelation length (Fig. 7, blue plots, left and middle panels) 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 Eq. (17) we then compute the subgrid profile $(CF_{s,sg})_z$, as well as the interlayer overlap parameter $\alpha_{inter}$ using Eqs. (18,19).

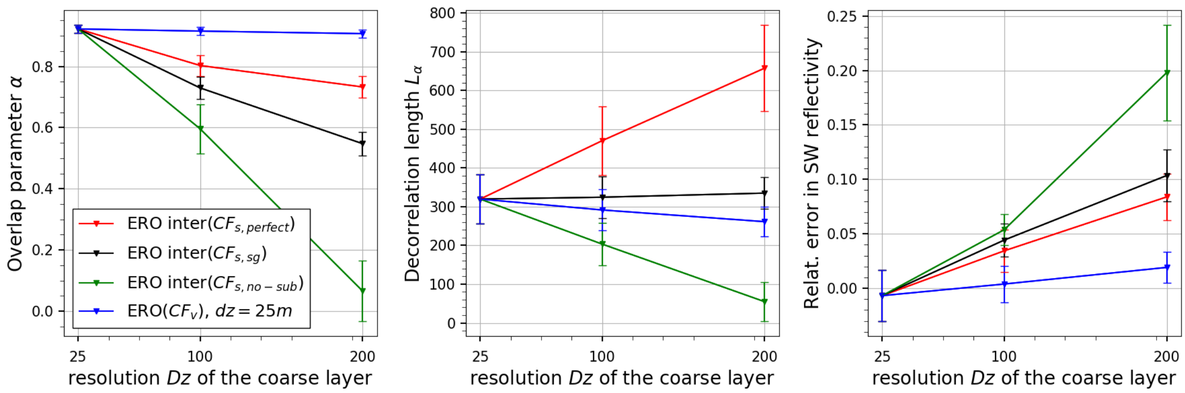


**Figure 7.** Overlap parameters (left) and decorrelation lengths (middle) for the ARMCu simulations (hours 6 to 12), for different coarse resolutions $Dz$ and for different reconstructions using ERO (see text). The daily mean value is shown. The overlap parameters are computed to match the total cloud cover of the 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.

We found that the overlap parameter $\alpha_{inter}$ varies with the resolution $Dz$ but the corresponding decorrelation length varies little from $\overline{L_{\alpha,sg}}$=326 $m$ (Fig. 7, black plots, 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 (Fig. 7, middle panel, blue and black lines). When it comes to radiative effects (Fig. 7, right panel), the error made on the SW cloud albedo is still small even when computed on the coarse grid (black plot) rather than on the finer grid (blue plot).

### 4.4 Analysis and comparisons of interlayer overlap for different estimations of the surface cloud fraction

Here we investigate, using Eqs. (18,19), how the overlap parameter $\alpha_{inter}$ and the decorrelation length should vary to keep the correct value of the total cloud cover for different estimations of the surface cloud fraction $CF_s$ in Eq. (18), instead of $CF_{s,sg}$. First we consider the extreme case where no subgrid heterogeneity is considered (Fig. 7, green plots), meaning the subgrid surface cloud fraction equals the volume cloud fraction $(CF_{s,no-sub})_z$=$(CF_v)_z$ 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 Fig. 6, for any altitude $z$ we have : $CF_{v,z} < CF_{s,z}$, so to generate the same total cloud cover, the overlap when no subgrid is taken into account has to be closer to random (i.e. $\alpha$ 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 surface cloud cover profile $(CF_{s,perfect})_z$ (Fig. 7, red plots). We then compute the interlayer overlap corresponding to this profile with Eqs. (18,19). The same reason applies to explain the difference with the interlayer overlap parameters computed for the subgrid cloud fraction profile: as shown in Fig. 6, $CF_{s,sg}$ approaches $CF_{s,perfect}$ in such a way that for any altitude $CF_{s,perfect} > CF_{s,sg} > CF_{s,no-sub}$. To conserve the same total cloud cover we then get $\alpha_{inter,perfect} > \alpha_{inter,sg} > \alpha_{inter,no-sub}$.

The middle panel of Fig. 7 shows the corresponding decorrelation lengths, computed from each overlap parameter $\alpha$ 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 plots). When doing ERO on the finer grid, the final resolution is $dz$=25 $m$ (blue plots). We see that for interlayer overlap, the decorrelation lengths have a strong dependence to the resolution when overlapping coarse layers of which the surface fraction is either perfect $(CF_{s,perfect})_z$ or determined assuming no subgrid heterogeneity $(CF_{s,no-sub})_z$, with important variations. This is not the case when the surface cloud fraction $CF_{s,sg}$ 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 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 $(CF_{s,perfect})_z$ and $(CF_{s,no-sub})_z$, but a small dependence to the resolution of the decorrelation length when overlapping $CF_{s,sg}$ . This dependence of $L_\alpha$ 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 litterature when generating cloudy subcolumns from GCMs or for observational simulators (Pincus et al. (2005); Bodas-Salcedo et al. (2011); 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 (Fig. 4 and blue plots of Fig. 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 suface cloud fraction as a bulk subgrid property using the volume cloud fraction and a subgrid overlap parameter. Overlapping this computed subgrid cloud fraction leads to a relative error in cloudy albedo of $\approx 10\%$ for coarse resolutions of 100 $m$ and 200 $m$ (Fig. 7, black plot). 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 $m$ and 200 $m$ (Fig. 7, red plot). Finally, even without taking into account any subgrid scale by overlapping $(CF_{s,no-sub})_z$ on the coarse grid, we can approach the albedo of the LES scenes within a 20% relative error (for a resolution of 200 $m$, Fig. 7, green plot) if the total cloud cover is reproduced. As all the generations shown in Fig. 7 have the same total 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 total 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 ARMCu cloud case, but similar results were also found consistently in several other cases, see Figs. S1-S3 in Supporting Information.

## 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 exponential-random overlap. 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 coarse 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 Fig. 7.

The procedure for reconstructing a cloud scene that we propose is as follow: given any volume cloud fraction profile $(CF_v)_z$ at resolution $Dz$ and the decorrelation length $L_\alpha$ for a reference resolution (here $dz$=25 $m$), the subgrid heterogeneity is taken into account by computing a profile of the surface cloud fraction $(CF_{s,sg})_z$ with Eq. (17), with $n$=$Dz/dz$ in the equation. The same decorrelation length $L_{\alpha,}$ allows to overlap these coarse layers and to compute the total cloud cover (Eq. (18)). As we can see on Fig. 7 for the case studied here, $L_{\alpha,25,Dz}$≈291 $m$ and $L_{\alpha,sg}$≈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 Fig. 7 for three other cumulus cloud cases simulated by the same LES and the same resolutions, with $L_{\alpha,25,Dz}$ and $L_{\alpha,sg}$ relatively independent of the resolution. 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 Figs. S1-S3 in Supporting Information). 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 Fig. 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 the Supporting Information on Figs. S4 to S6.

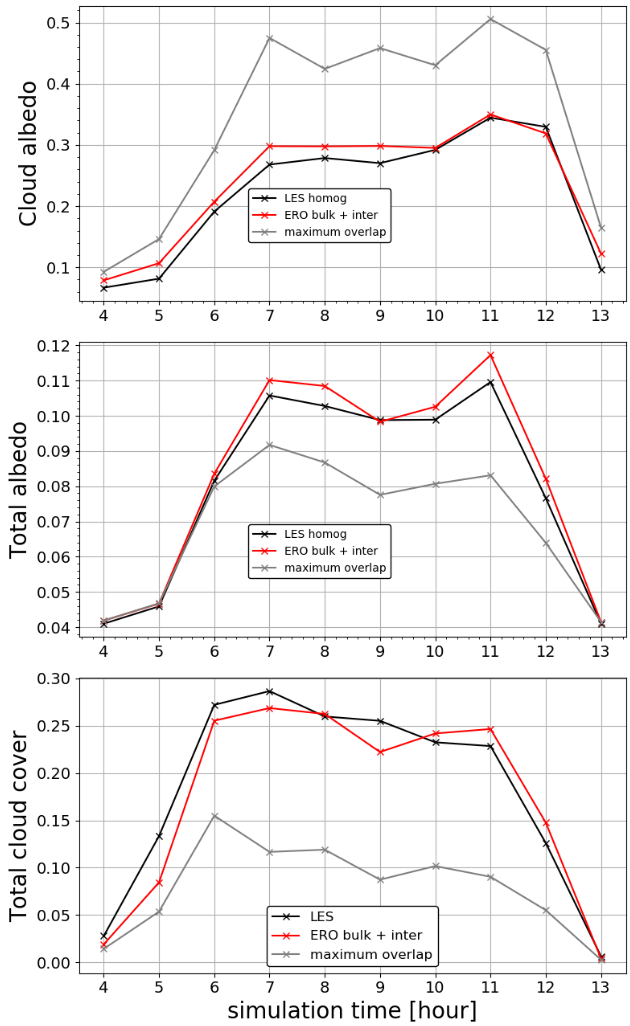


**Figure 8.** Cloud albedo (top panel), total albedo (middle panel) and total cloud cover (lower panel) for the LES (in red), our reconstruction using ERO (in black) and a maximum overlap reconstruction (grey). The constant decorrelation length used here both for the subgrid computation of the surface cloud fraction profile and its interlayer overlap is $L_\alpha$=309 $m$. The scenes are the ARMCu case (time steps $h \in [4, 13]$). In all scenes the LWC is homogeneous at each vertical level.

### 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 (Oreopoulos and Norris (2011); Jing et al. (2016)). 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 entierely cloudy even if the cloud does not fully extend on the vertical of the layer. Hence the measured profile is the surface cloud fraction $(CF_s)_z$ for a coarse layer of thickness $Dz$=480 $m$. Combining Eqs. (17,18,19), 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 : $(CF_{s,1})_z$, $L_{\alpha,1}$, and $(CF_{s,2})_z$, $L_{\alpha,2}$, where $L_{\alpha,i}$ are the decorrelation lengths corresponding to the measured surfacic cloud fraction profiles.

Using the cloud fraction profile with finer vertical resolution $CF_{s,2}$ we can use interlayer ERO with $L_{\alpha,2}$ on blocks of $n$ fine layers to compute the corresponding surface cloud fraction profile at the resolution $dz_1$, $CF'_{s,1}$. Knowing the total cloud cover $CC$, we can then compute with Eq. (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 ARMCu simulations used on Fig. 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 $L_\alpha$ with the vertical resolution for a perfect estimation of the surface cloud fraction profile. $I_2$ would measure a decorrelation length $L_{\alpha,2}$=320 $m$, while $I_1$ would measure $L_{\alpha,1}$=658 $m$ (Fig. 7 middle panel, in red). We get a factor 2 on the estimation of the decorrelation length in this case. The vertical extension of the studied clouds is too small to be able to compute the decorrelation lenght 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 $L_\alpha \sim 2\ km$ (Hogan and Illingworth (2000); Willèn et al. (2005); Barker (2008a); Oreopoulos and Norris (2011); Jing et al. (2016)). 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 litterature with LES simulations with similar vertical resolutions (Neggers et al. (2011); Sulak et al. (2020); Villefranque et al. (2021)). The difference in horizontal resolutions (Naud et al. (2008); Astin and Di Girolamo (2014); Tompkins and 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 and 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 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 $Re(k) > 0$ :

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

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 Fig. 9 the LWC of both scenes' cloudy subcolumns after they have been sorted along their vertical 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 caracteristics 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 $(CF_v)_z$ 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 ARMCu 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 ARMCu 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 $\approx$ 10% brighter, which confirms the previous findings of Barker et al. (2003), Wu and Liang (2005), and Shonk and Hogan (2010).

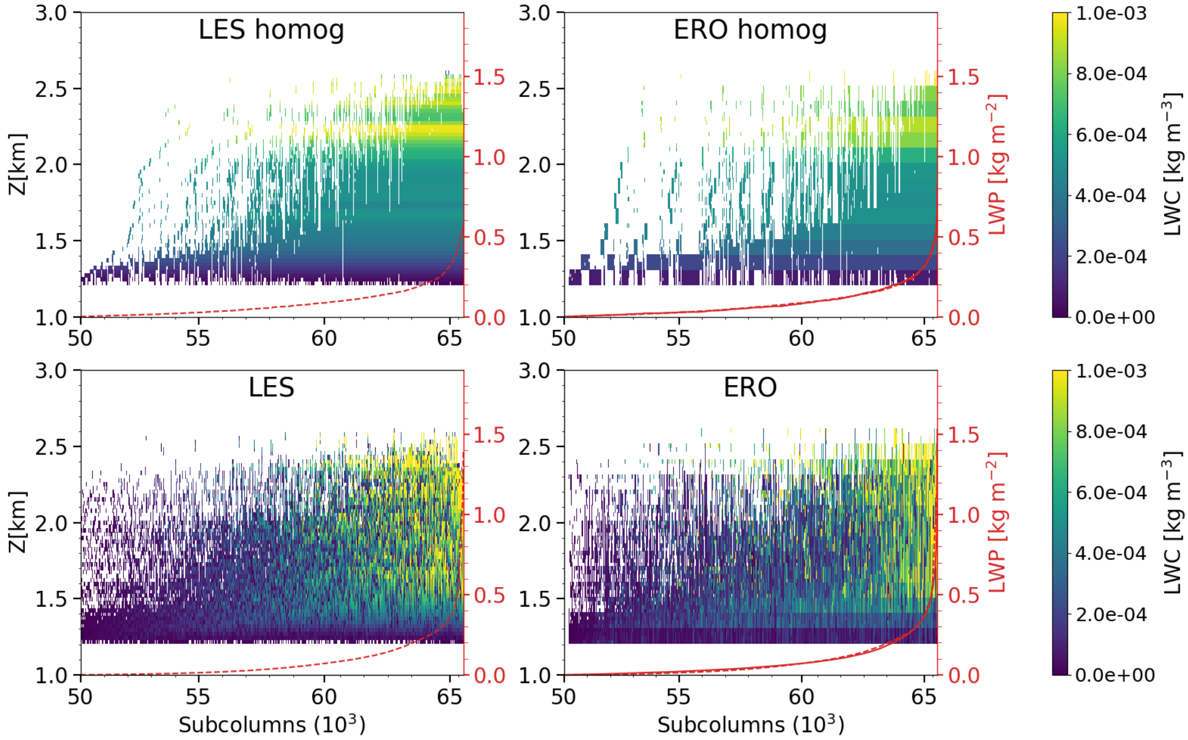


**Figure 9.** The liquid water content of each scene's cloudy subcolumns in the LES simulations (left panels) and reconstructed using ERO (right panels). The subcolumns have been sorted along their LWP (red plots). The red lines represent the LES (dashed line) and generated (solid line) LWP, the former being represented on the right panels as well to facilitate the comparison. Top panels are homogeneous LWC for each level and whereas it varies in the bottom panels.

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 exponential-random overlap (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.

Within the McICA framework, 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 total cloud cover or the overlap parameter. Compared to reference LES simulations, the generated cloud scenes show a correct representation of both the distribution of cumulative cloud fraction among cloudy subcolumns and the vertical profile of the cloud cover seen from above or below. We suggest that the later is a simple diagnostic that would usefully complement the usual cloud fraction vertical profile when comparing models with observations or when developping 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 reconstructions.

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 total 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 volumic or surfacic. Although this deserves more investigations, we provide a framework that allows to go from one vertical resolution to an other. Further work is also required to establish robust estimates of the decorrelation length for a large variety of clouds.

To our best knowledge, most current atmospheric models neglect the effect of subgrid variability on the cloud fraction and assume a maximum-random overlap of cloud layers or a ERO with a quite large decorrelation length ($\approx 2{-}3\ km$). This can lead to an underestimation of the cloud cover by a factor of 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)). 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.

In addition to the effect of the water content heterogeneity on cloud albedo, already well recognized, we show that the vertical distribution of cloud fraction also matters. Indeed, 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 how the vertical profile of the clouds fraction is represented. 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% to 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 levels $[k_{base}, k_{top}]$ (with $\#([k_{base}, k_{top}]) = \mathcal{N}$ 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 coefficient $c_{i,k} \in \{0, 1\}$, which corresponds to whether the cell is cloudy or not , as well as the liquid water content.

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

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

where $RN1$ are random numbers evenly distributed on $[0, 1]$. Working its way down, the algorithm computes the next coefficients, 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 cloudy state $c_{i,k}$ is computed as :

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

  where $(c_k|c_{k-1})_{max}$ are booleans computed according to the transition probabilities $P_{max}(C_k = c_k \mid C_{k-1} = c_{k-1})$ which is defined by Eq. 6 when $C_k{=}C_{k-1}$. To complete this implementation, according to Eq. (2), we also have:

$$P_{max}(\mathrm{C}_k = 1 \mid \mathrm{C}_{k-1} = 0) = 1 - P_{max}(\mathrm{C}_k = 0 \mid \mathrm{C}_{k-1} = 0) = \frac{\max(CF_{k-1}, CF_k)}{1 - CF_{k-1}} \tag{A2}$$

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

$$c_{i,k} = (1 \mid 1)_{rand} = (1 \mid 0)_{rand}$$

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

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

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 amout) of a given cell $k$, Räisänen generator computes $x_k \in [0, 1]$ to compare it to the cloud fraction of the cell $CF_k$ and decide wether 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 levels $k$ 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 levels our method is equivalent, but for more than two levels, 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 ($P_{max}(1 \mid 1), P_{max}(1 \mid 0), P_{rand}(1 \mid 1), P_{rand}(1 \mid 0)$), and by using a different random number every time it is needed, we conserve the cloud fraction without creating this correlation between the layers.

**Acknowledgments**
Our many thanks go to Céline Cornet and Frédéric Szczap for insightful discussions about this work. 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). A repository containing the scripts for the ERO algorithm presented in this paper is available at https://github.com/raphleb/ERO.git. The sources described in this paper for the radiative computations are available at the websites (https://www.meso-star.com/projects/htrdr/htrdr.html and https://www.meso-star.com/projects/star-engine/star-engine.html).

## References

Astin, I., & Di Girolamo, L. (2014). Technical note: The horizontal scale dependence of the cloud overlap parameter . *Atmos. Chem. Phys.*, *14*(18), 9917–9922. doi: 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. *J. Geophys. Res.-Atm.*, *113*(D8). doi: 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. *J. Geophys. Res.-Atm.*, *113*(D24). doi: https://doi.org/10.1029/2008JD010391

Barker, H. W., & Räisänen, P. (2005). Radiative sensitivities for cloud structural properties that are unresolved by conventional GCMs. *Q. J. R. Meteorol. Soc.*, *131*(612), 3103-3122. doi: 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. *Q. J. R. Meteorol. Soc.*, *125*(558), 2127-2152. doi: https://doi.org/10.1002/qj.49712555810

Barker, H. W., Stephens, G. L., Partain, P. T., Bergman, J. W., Bonnel, B., Campana, K., … Yang, F. (2003). Assessing 1D Atmospheric Solar Radiative Transfer Models: Interpretation and Handling of Unresolved Clouds. *J. Climate*, *16*(16), 2676 - 2699. doi: 10.1175/1520-0442(2003)016⟨2676:ADASRT⟩2.0.CO;2

Bergman, J. W., & Rasch, P. J. (2002). Parameterizing vertically coherent cloud distributions. *J. Atmos. Sci.*, *59*(14), 2165-2182. doi: 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., … John, V. O. (2011). COSP: Satellite simulation software for model assessment. *Bull. Am. Meteorol. Soc.*, *92*(8), 1023 - 1043. doi: 10.1175/2011BAMS2856.1

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. *J. Atmos. Sci.*, *62*(7), 2248-2260. doi: 10.1175/JAS3467.1

Brown, A. R., Cederwall, R. T., Chlond, A., Duynkerke, P. G., Golaz, J.-C., Khairoutdinov, M., … Stevens, B. (2002). Large-eddy simulation of the diurnal cycle of shallow cumulus convection over land. *Q. J. R. Meteorol. Soc.*, *128*(582), 1075-1093. doi: https://doi.org/10.1256/003590002320373210

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

Geleyn, J., & Hollingsworth, A. (1979). An economical analytical method for the computation of the interaction between scattering and line absorption of radiation. *Beitr. Phys. Atmosph.*.

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

Hogan, R. J., & Illingworth, A. J. (2000). Deriving cloud overlap statistics from radar. *Q. J. R. Meteorol. Soc.*, *126*(569), 2903-2909. doi: 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. *J. Atmos. Sci.*, *70*(2), 708 - 724. doi: 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. *Q. J. R. Meteorol. Soc.*, *125*(555), 941-965. doi: 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. doi: 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. *J. Adv. Model. Earth Syst.*, *10*(11), 2686-2705. doi: https://doi.org/10.1029/2018MS001379

Konsta, D., Dufresne, J.-L., Chepfer, H., Vial, J., Koshiro, T., Kawai, H., … Ogura, T. (2022). Low-level marine tropical clouds in six cmip6 models are too few, too bright but also too compact and too homogeneous. *Geophys. Res. Lett.*, *49*(11), e2021GL097593. doi: https://doi.org/10.1029/2021GL097593

Koren, I., Oreopoulos, L., Feingold, G., Remer, L. A., & Altaratz, O. (2008). How small is a small cloud? *Atmos. Chem. Phys.*, *8*(14), 3855–3864. doi: 10.5194/acp-8-3855-2008

Lac, C., Chaboureau, J.-P., Masson, V., Pinty, J.-P., Tulet, P., Escobar, J., … Wautelet, P. (2018). Overview of the Meso-NH model version 5.4 and its applications. *Geosci. Model Dev.*, *11*(5), 1929–1969. doi: 10.5194/gmd-11-1929-2018

Lafore, J. P., Stein, J., Asencio, N., Bougeault, P., Ducrocq, V., Duron, J., … Vilà-Guerau de Arellano, J. (1998). The Meso-NH Atmospheric Simulation System. Part I: adiabatic formulation and control simulations. *Annales Geophysicae*, *16*(1), 90–109. doi: 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. *J. Atmos. Sci.*, *59*(24), 3519 - 3539. doi: 10.1175/1520-0469(2002)059⟨3519:SSAMVI⟩2.0.CO;2

Mace, G. G., & Benson-Troth, S. (2002). Cloud-layer overlap characteristics derived from long-term cloud radar data. *J. Climate*, *15*(17), 2505 - 2515. doi: 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. *J. Climate*, *21*(8), 1758 - 1770. doi: 10.1175/2007JCLI1828.1

Neggers, Duynkerke, P. G., & Rodts, S. M. A. (2003b). Shallow cumulus convection: A validation of large-eddy simulation against aircraft and landsat observations. *Q. J. R. Meteorol. Soc.*, *129*(593), 2671-2696. doi: https://doi.org/10.1256/qj.02.93

Neggers, Heus, T., & Siebesma, A. P. (2011). Overlap statistics of cumuliform boundary-layer cloud fields in large-eddy simulations. *J. Geophys. Res.-Atm.*, *116*(D21). doi: 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. *J. Atmos. Sci.*, *60*(8), 1060 - 1074. doi: 10.1175/1520-0469(2003)60⟨1060:SSOCCP⟩2.0.CO;2

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

Oreopoulos, L., & Norris, P. M. (2011). An analysis of cloud overlap at a midlatitude atmospheric observation facility. *Atmos. Chem. Phys.*, *11*(12), 5557–5567. doi: 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. *J. Geophys. Res.-Atm.*, *108*(D13). doi: 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. *J. Geophys. Res.-Atm.*, *110*(D15). doi: 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. *Q. J. R. Meteorol. Soc.*, *130*(601), 2047-2067. doi: 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. *J. Climate*, *21*(11), 2352 - 2370. doi: 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. *Q. J. R. Meteorol. Soc.*, *136*(650), 1205-1215. doi: https://doi.org/10.1002/qj.646

Siebesma, A. P., Bretherton, C. S., Brown, A., Chlond, A., Cuxart, J., Duynkerke, P. G., … Stevens, D. E. (2003). A large eddy simulation intercomparison study of shallow cumulus convection. *J. Atmos. Sci.*, *60*(10), 1201 - 1219. doi: 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. *J. Geophys. Res.-Atm.*, *125*(10), e2019JD032017. doi: 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. *Geosci. Model Dev.*, *11*(1), 77–81. doi: 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. *J. Atmos. Sci.*, *64*(6), 2116-2125. doi: 10.1175/JAS3925.1

Tompkins, A. M., & Di Giuseppe, F. (2015). An interpretation of cloud overlap statistics. *J. Atmos. Sci.*, *72*(8), 2877 - 2889. doi: 10.1175/JAS-D-14-0278.1

vanZanten, M. C., Stevens, B., Nuijens, L., Siebesma, A. P., Ackerman, A. S., Burnet, F., … Wyszogrodzki, A. (2011). Controls on precipitation and cloudiness in simulations of trade-wind cumulus as observed during RICO. *J. Adv. Model. Earth Syst.*, *3*(2). doi: https://doi.org/10.1029/2011MS000056

Villefranque, N., Blanco, S., Couvreux, F., Fournier, R., Gautrais, J., Hogan, R. J., … Williamson, D. (2021). Process-based climate model development harnessing machine learning: III. the representation of cumulus geometry and their 3D radiative effects. *J. Adv. Model. Earth Syst.*, *13*(4), e2020MS002423. doi: https://doi.org/10.1029/2020MS002423

Villefranque, N., Fournier, R., Couvreux, F., Blanco, S., Cornet, C., Eymet, V., ... Tregan, J.-M. (2019). A Path-Tracing Monte Carlo Library for 3-D Radiative Transfer in Highly Resolved Cloudy Atmospheres. *J. Adv. Model. Earth Syst.*, *11*(8), 2449-2473. doi: 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) doi: 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. *J. Atmos. Sci.*, *62*(11), 4105 - 4112. doi: 10.1175/JAS3565.1

# A consistent representation of cloud overlap and cloud subgrid vertical heterogeneity

**Raphaël Lebrun**[1], **Jean-Louis Dufresne**[1], **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 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 exponential-random overlap are highly dependent on the vertical resolutions of models and observations.

Corresponding author: Raphal Lebrun, `raphael.lebrun@lmd.ipsl.fr`

**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 LES results as reference, we propose a method based on exponential-random overlap (ERO) 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 an additional 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, in the framework of commonly used overlap assumptions. When the cloud cover of the scene is known, we show that the exponential-random overlap allows a good representation of the vertical structure of clouds and of the cloud albedo. We find the decorrelation lengths 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. We present values that can be used to compute accurately the cloud cover and the cloud albedo of the studied scenes.

# 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 (Räisänen et al. (2004); Hogan and Shonk (2013)). The cloud cover and the mean optical depth of the cloudy region are inter-dependent when the profile of cloud fractions and water contents are known. They depend on how the cloud fractions overlap on the vertical: if they overlap maximally, the cloud cover will be minimum and the mean optical depth maximum, and if they overlap randomly, the cloud cover will be larger and the mean optical depth smaller.

How the cloud fraction ($CF$) of each atmospheric layer overlap with other layers has been widely studied (Geleyn and Hollingsworth (1979); Barker et al. (1999); Jakob and 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 (Hogan and Illingworth (2000); Bergman and Rasch (2002); Tompkins and Di Giuseppe (2007); Shonk and Hogan (2010)). The corresponding decorrelation length scale has been estimated from satellite radar observations (Jing et al. (2016)), in-situ observations (Mace and 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 (Pincus et al. (2005); Di Giuseppe and Tompkins (2015); 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 volume of the grid box that contains clouds, $CF_v$, but radiation is primarily sensitive to the surface cloud fraction $CF_s$ which is the relative surfacic fraction covered by clouds in a cell. Often implicitly, these two fractions are assumed to be equal, i.e. the clouds are assumed to be homogeneous on the vertical in each cell. This can seem logical on the first order given the area/depth ratio of the grid cells, 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 $CF_s$ is typically greater than $CF_v$ by about 30% (Neggers et al. (2011)). A direct consequence of not taking into account this difference is that, for a given cloud faction in volume, the surface 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 exponential-random overlap 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 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. Like done in the McICA method, we neglect the 3D effects and keep the classical plane parallel assumption (each vertical profile represents a stack of horizontally infinite and homogeneous slabs) in our 1D approach. Assuming that the volumic 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.

The manuscript is organised as follows: in Section 2, we consider the exponential-random overlap (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 alebdo 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 inter-layer 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 exponential-random overlap (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 combined cloud fraction of two adjacent cloudy layers of surfacic fractions $CF_1$ and $CF_2$ is:

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

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

$$CF_{1,2,max} = \max(CF_1, CF_2)$$

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

$$CF_{1,2,rand} = CF_1 + CF_2 - CF_1 CF_2$$

In this model, “exponential” refers to the fact that $\alpha$ 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 and 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 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 paper, the McICA framework is used to generate samples of vertical profiles. The main difference is that in the usual McICA algorithm, the profiles are generated on the vertical grid of the host model, while here we aim at generating profiles at any vertical resolution, including finer vertical resolutions.

Unless otherwise stated, in all this article, we consider a single vertical atmospheric column that consists of a cloudy block (with a strictly positive liquid water content at every level) of $\mathcal{N}$ vertical layers. From this column we assume the volume cloud fraction of each layer, $(CF_k)_{k=1\ldots\mathcal{N}}$, is known. We consider the exponential-random overlap model (ERO) as a Markovian process and deduce the relationship between the overlap parameter $\alpha$ and the total cloud cover $CC$. 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).

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. If $\vec{\mathrm{C}} = (\mathrm{C}_k)_{k=1\ldots\mathcal{N}}$ is the random variable representing the cloud fraction distribution of the subcolumn, with $\mathrm{C}_k \in \{0,1\}$ (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 $\vec{\mathrm{C}} = (c_k)_{k=1\ldots\mathcal{N}} \in [0,1]^{\mathcal{N}}$ is given by:

$$P(\vec{\mathrm{C}}) = \prod_{k=1}^{\mathcal{N}} P(\mathrm{C}_k = c_k \mid \mathrm{C}_{k-1} = c_{k-1}) \tag{1}$$

where $C_0 = 0$ (i.e. there is no cloud above the cloud block considered here). We use the classic upper case notation $\mathrm{C}_k$ for the random variables and the lower case notation $c_k$ for their realizations.

For any level $k$ in the subcolumn, the probability to have $c_k = 1$ is the cloud fraction of the level, meaning $P(\mathrm{C}_k = 1) = CF_k$. We'll call $P(\mathrm{C}_k = c_k \mid \mathrm{C}_{k-1} = c_{k-1})$ a *transition probability*, it is the probability that in a subcolumn, layer $k$ is in the state $c_k$, knowing the layer $k-1$ is in the state $c_{k-1}$. Since $c_k$ is either 0 or 1, there are only four possible types of transition between two levels, and being able to compute their probabilities at every level gives the probability of any vertical cloud fraction distribution for the column. Moreover, for each level $k$, two out of the four transition probabilities are dependant, as a layer is either cloudy or clear sky:

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

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

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

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

where $(\mathrm{C}_k = 1 \cap \mathrm{C}_{k-1} = 1)$ is the event with both layers cloudy. If we assume an exponential-random overlap we have :

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

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

$$P_{rand}(\mathrm{C}_k = 1 \mid \mathrm{C}_{k-1} = 1) = P_{rand}(\mathrm{C}_k = 1 \mid \mathrm{C}_{k-1} = 0) = P_{rand}(\mathrm{C}_k = 1) = CF_k \tag{5}$$

The transition probability in a subcolumn of the maximum overlap can be obtained using Eq. (3): if $CF_{k-1} < CF_k$ : $P_{max}(\mathrm{C}_k = 1 \mid \mathrm{C}_{k-1} = 1) = 1$, and on the contrary if $CF_{k-1} \geq CF_k$: $P_{max}(\mathrm{C}_k = 1 \mid \mathrm{C}_{k-1} = 1) = \frac{CF_k}{CF_{k-1}}$

As a result,

$$P_{max}(\mathrm{C}_k = 1 \mid \mathrm{C}_{k-1} = 1) = \frac{\min(CF_{k-1}, CF_k)}{CF_{k-1}} \tag{6}$$

and (4) becomes :

$$P(\mathrm{C}_k = 1 \mid \mathrm{C}_{k-1} = 1) = \alpha \frac{\min(CF_{k-1}, CF_k)}{CF_{k-1}} + (1-\alpha) CF_k \tag{7}$$

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

$$P_{max}(\mathrm{C}_k = 0 \mid \mathrm{C}_{k-1} = 0) = \frac{1 - \max(CF_{k-1}, CF_k)}{1 - CF_{k-1}}$$

and therefore:

$$P(\mathrm{C}_k = 0 \mid \mathrm{C}_{k-1} = 0) = \alpha \times \frac{(1 - \max(CF_{k-1}, CF_k))}{1 - CF_{k-1}} + (1-\alpha)(1 - CF_k) \tag{8}$$

These equations and exponential-random overlap more generally are applicable only for non overcast cloudy layers (i.e. $CF \in ]0,1[$). 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 Eqs. (7), (8) and (2) we can now compute the different transition probabilities for each layer $k$, knowing $\alpha$. Then using Eq. (1) we can compute the probability to generate any vertical cloud fraction distribution for a subcolumn, for any exponential-random overlap parameter $\alpha \in [0,1]$.

### 2.2 The relationship between the overlap parameter $\alpha$ and the total 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 $\alpha$ and the total cloud cover $CC$, assuming ERO.

To obtain the formal expression of the total 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:

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

Using Eq. (8) we get:

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

Given this equation, if we know the overlap parameter $\alpha$, the total cloud cover is:

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

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

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

where

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

For a given $(CF)_{1\ldots\mathcal{N}}$ profile (with $CF_k \in ]0,1[$ for each layer) and knowing $CC$, the function $f_\emptyset$ is strictly increasing, so $f_\emptyset^{-1}$ exists. We compute $\alpha$ with a dichotomy method using a tolerance $\epsilon = 10^{-5}$.

Eq. (12) gives the expression of $\alpha$ for a given cloud cover $CC$ and cloud fraction profile $(CF)$. Eq. (10) allows us to compute $CC$ if we know the overlap parameter $\alpha$ and the profile $(CF)$. Therefore for any given profile $(CF)$ and given the ERO model, it is equivalent to know $CC$ or $\alpha$ (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 $\{\widehat{CF_k}\}_{k=1\ldots N}$, and we generate subcolumns with $n$ times more vertical levels, $\mathcal{N} = (N \times n)$. We introduce the hypothesis that at every coarse level of the atmospheric column, the volume cloud fraction is the same for all the $n$ sublayers :

$$\forall \quad l \in \mathcal{L}_k^n, \quad CF_l = \widehat{CF_k}$$

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 cell is uniform, the intralayer transition probability $P(C_l = 0|C_{l-1} = 0)$ (Eq. (8)) between layers inside the same coarse cell simplifies as :

$$P(C_l = 0|C_{l-1} = 0) = P_{intra,l} = \alpha + (1-\alpha)(1 - CF_l) \tag{13}$$

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

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

Like done previously, we can compute the cloud cover generated by a given overlap parameter $\alpha$, or if the total cloud cover of the scene is known, we can inverse this equation using Eq. (12) to compute the overlap parameter $\alpha$ 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 $\alpha$ and the cloud generation

As done in many previous works such as Larson et al. (2002); R. A. J. Neggers et al. (2003); Neggers et al. (2011), we are using Large Eddy Simulations (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 ARMCu cloud case (Brown et al. (2002)) showing the development of shallow cumulus convection over land, as well as two marine, trade-winds cumulus cloud cases BOMEX (Siebesma et al. (2003)) and RICO (vanZanten et al. (2011)), and another case of continental cumulus SCMS (Neggers et al. (2003b)). For each case we use the corresponding LES results obtained with the atmospheric non-hydrostatic model MESO-NH (Lafore et al. (1998); Lac et al. (2018)), and all these simulations represent a 6.4 $km\times$ 6.4 $km\times$ 4 $km$ domain with a $dx$=$dy$=$dz$=25 $m$ resolution. For each LES simulation we coarsen it into a single atmospheric column with the same vertical resolution $dz$, or a lower vertical resolution $Dz$, as shown in Fig. 1. For each of these single columns we know, by means of the LES, the total cloud cover $CC$, as well as the cloud fraction and the liquid water content at each vertical level. 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.

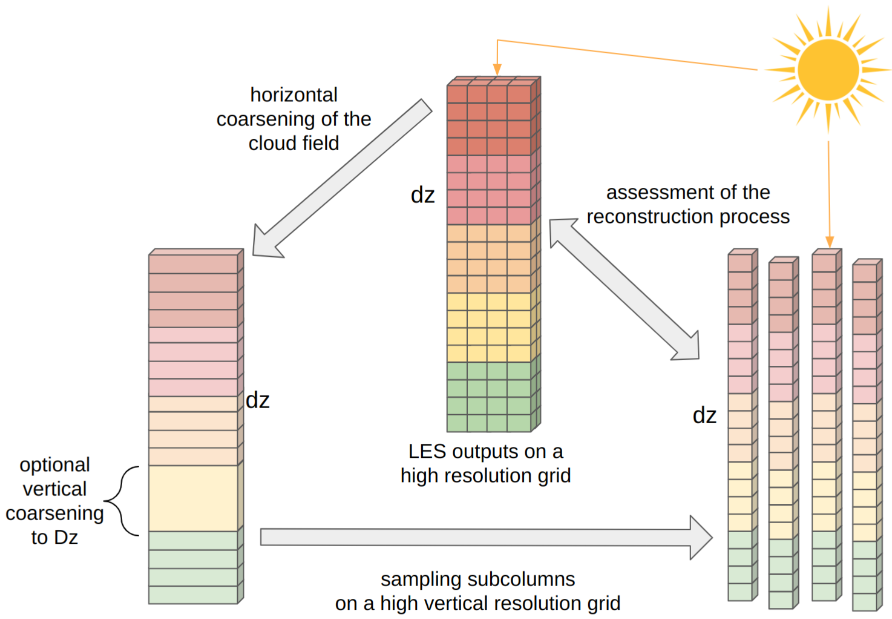


**Figure 1.** Method used to develop and assess our cloudy columns sampling. The LES cloud field of resolution $dx$=$dy$=$dz$=25 $m$ is horizontally averaged into a single column and eventually averaged vertically to a coarse resolution $Dz$>$dz$. We then sample $N_s$ subcolumns with a vertical resolution $dz$ using the ERO algorithm, and then assess the process by comparing the sample's cloud fraction profile and TOA SW cloud albedo to the ones of the original 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 $\alpha$ 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), 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 $\alpha$ used to generate this sample is computed with Eqs. (12,14) 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, $N_s \approx 6.5 \times 10^4$ subcolumns have been generated. For this number, the total 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 cloud fraction seen from above or from below at altitude $z$ varies as a function of this altitude (Fig. 2).

The blue line (Fig. 2, middle and right panels) is the cloud cover profile of the original LES, with a total cloud cover of 0.2325. The grey line is obtained using a maximum overlap assumption, and shows a total cloud cover of only $\sim 10\%$. Since the scene consists of a single cloud block, this corresponds to models using 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 total cloud cover of the LES ($\alpha = 0.921$), with a very close total 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 total cloud cover than the LES, but also a close projected cloud cover at each vertical level. The abrupt changes in the cloud cover of the sampled subcolumns are a consequence of the hyopthesis of a constant volume cloud fraction $CF_v$ in each coarse cell. For the generation without vertical subgriding of the previous section ($Dz$=$dz$=25 $m$), the vertical distribution of the cloud cover is almost indiscernible to that of the LES (not shown).

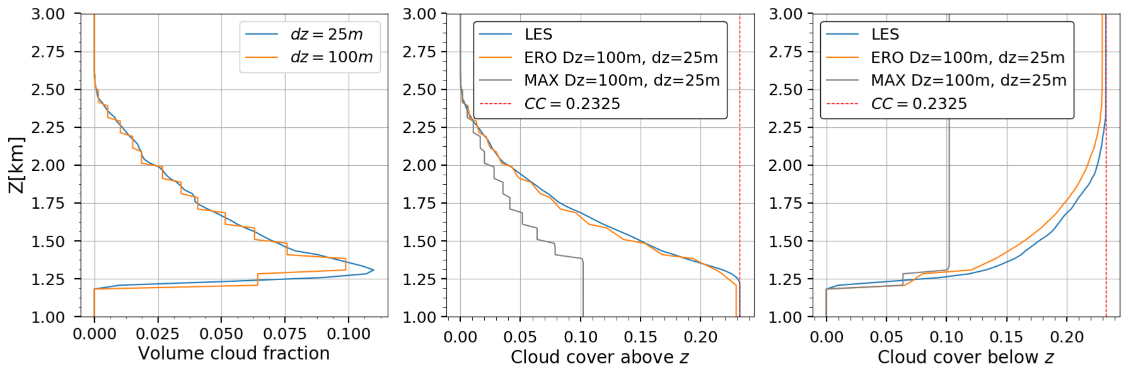


**Figure 2.** Vertical distribution of the volume cloud fraction (left), of the total cloud cover above (middle) and below (right) altitude z. The former is the projected total cloud cover of all the clouds between the top of the domain and altitude $z$, the latter is the projected cloud cover between the bottom of the domain and altitude $z$. On the middle and right panels are compared the profiles from the LES (blue) and those obtained with two overlap models : maximum overlap (grey) and ERO (orange). The red dot line shows the total 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 ARMCu cloud case (time step $h$=10).

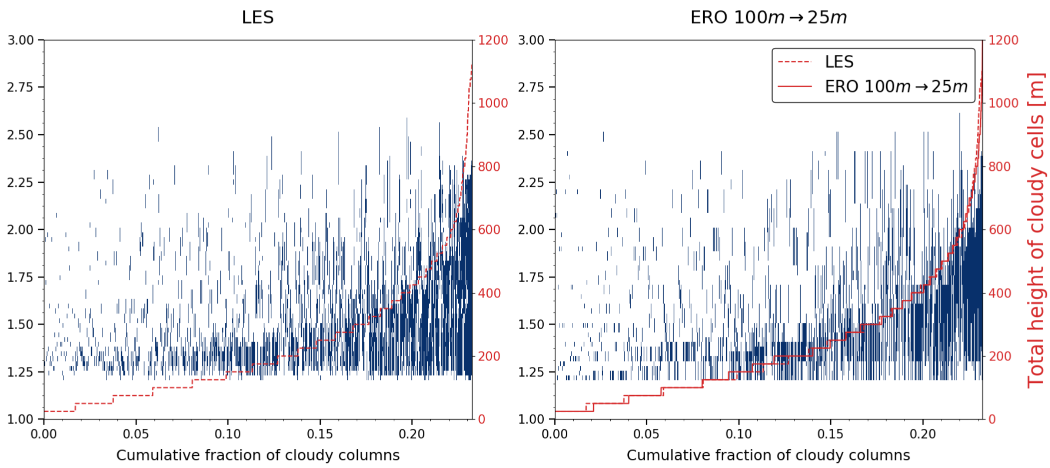


**Figure 3.** The cloudy subcolumns of the LES scene (left) are sorted along the number of cloudy cells in each subcolumns (dashed red). On the right the cloudy subcolumns out of a $N_s \approx 6.5 \times 10^4$ sample of subcolumns generated with ERO sorted in the same way (solid red for the number of 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 ARMCu case.

To go further, Fig. 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. The vertical distribution of cloudy cells are very close, it shows the ERO generation not only reproduces the total cloud cover of the original scene, but also the distribution of cumulative cloud fraction.

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 Independant Column Approximation -or ICA - (Pincus et al. (2003)). Fig. 4 shows the cloud albedo of different sampling hypotheses, of the original LES scenes, as well as the total albedo of the scenes, and their total 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 total cloud covers.

The maximum overlap assumption (grey) shows a much higher cloud albedo since it produces cloud scenes with less total cloud cover and hence brighter clouds. 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 starting with a 25 $m$ vertical resolution is $\sim 1.5\%$ and only of $\sim 2.5\%$ when starting with a 200 $m$ vertical resolution, for the simulation hours $[6, 12]$.

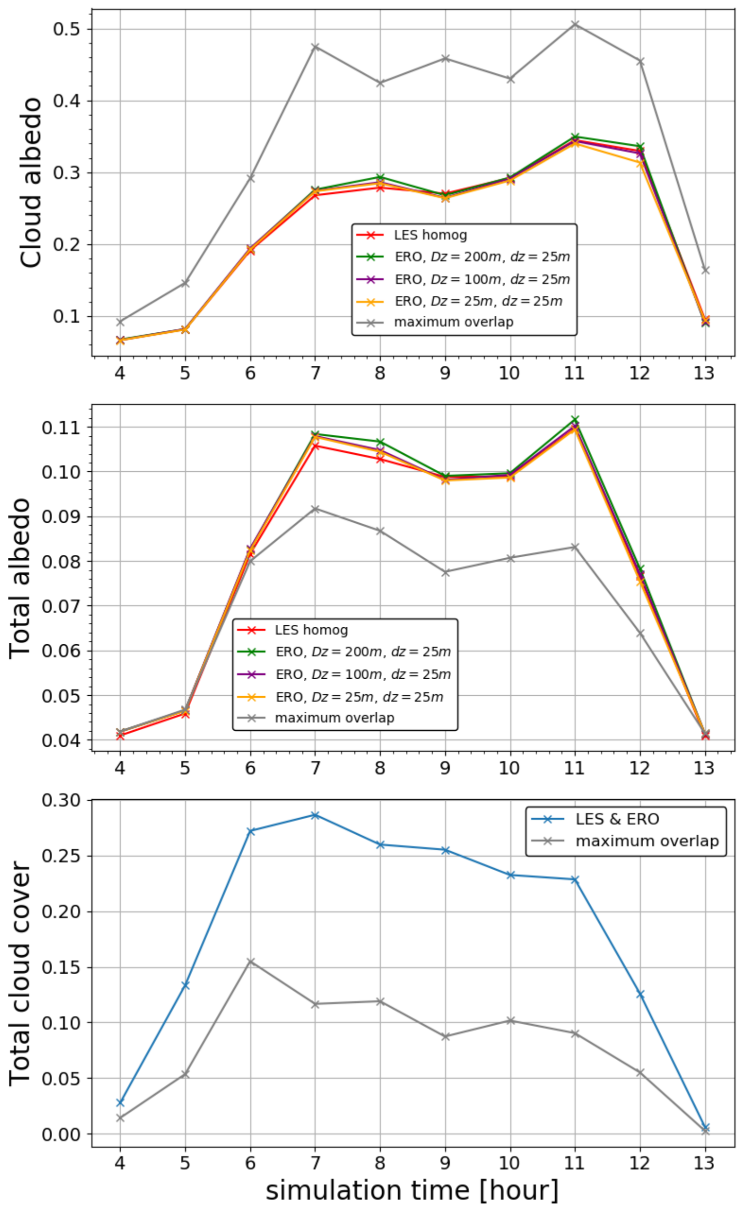


**Figure 4.** Cloud albedo (top panel), total albedo (middle panel) and total cloud cover (lower panel) for the LES (in red), for ERO with different coarse resolutions $Dz$ and for maximum overlap with the coarse resolution $Dz$=100 $m$ (in grey). The albedo of each scene is computed using a Monte-Carlo algorithm under the Indenpendant Column Approximation, for the ARMCu 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 LWC is homogeneous at each vertical level. For each computation, $10^6$ realisations were made, with a Monte-Carlo standard deviation of the cloud albedo of $10^{-6}$.

### 3.2 Analysis of the overlap parameter $\alpha$

In Section 2 we established the relationship between the overlap parameter $\alpha$ and the total cloud cover $CC$ and used it in 3.1 to determine $\alpha$ from the total cloud cover $CC$ diagnosed from LES results. In this section we analyze the overlap parameters computed this way and compare them to the values given by 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 $L_\alpha$ are usually related to each other via the following relation ( Hogan and Illingworth (2000); Bergman and Rasch (2002); Mace and Benson-Troth (2002)) :

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

If the decorrelation length $L_\alpha$ 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 and Illingworth (2000); Oreopoulos and Norris (2011)), according to Eq. (16). Fig. 5 shows the variations of the overlap parameters $\alpha$ computed at different times of the day of the ARMCu 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 Eq. (16). This profile was obtained by computing the mean overlap parameter for each possible separation distance by using $CF_s = \alpha CF_{max} + (1-\alpha)CF_{rand}$. The overlap parameter $\alpha_{25,Dz}$ corresponds to the overlap parameter computed using Eq. (14) to reproduce the total cloud cover $CC$ with vertical subgriding from a vertical resolution $Dz$=100 $m$ to $dz$=25 $m$. 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 $m$. 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 $\sim$ 300 $m$) are comparable to those computed in the litterature with similar LES simulations (Neggers et al. (2011); Sulak et al. (2020); Villefranque et al. (2021)). The difference with decorrelation lengths in the litterature that take into account the overlap of whole atmospheric columns in global model is further discussed in Section 5.

We have also computed the overlap parameter $\alpha$ 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 ARMCu($h$=10) when taking into account the 45 clouds that account for 99% of the total 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 $m$.

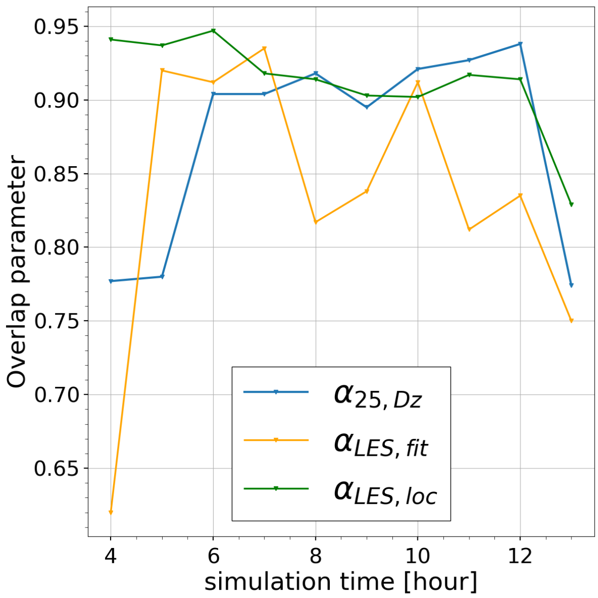


**Figure 5.** Overlap parameters computed with three different methods (see text) at each time step of the LES simulations. The data used are the ARMCu cloud fields.

# 4 Using ERO to model subgrid properties and overlap coarse vertical layers

To summarize the previous section, if we know the overlap parameter $\alpha_{25,Dz}$ or the total cloud cover of the scene, and its volume cloud fraction $CF$ for every cloudy layer of thickness $Dz$ as well as the LWC mean value, 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 ARMCu 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 total 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 surface cloud fraction $CF_s$ and the volume cloud fraction $CF_v$ (Genio et al. (1996); Jouhaud et al. (2018)). $CF_v$ represents the volume fraction of the layer that contains clouds ( i.e. where liquid or solid water particules are present), whereas $CF_s$ represents the surface fraction of the layer covered by clouds when looking from above or below. In other words, $CF_s$ is the vertical projection of $CF_v$, and it is $CF_s$ that is used by radiation codes in GCMs and teledetection.

At the LES grid scale, we have assumed that a grid cell is either clear or cloudy, and therfore $CF_v$=$CF_s$. This is no longer the case on a coarse grid, and ERO can be used to compute $CF_s$ in a coarse layer of an atmospheric column, knowing $CF_v$.

For that we consider an atmospheric cloudy column of coarse vertical resolution $Dz = n \times dz$. If $CF_v$ is known and vertically uniform within each coarse layer, we are back in the configuration we were in Section 3 when using subgriding, with $CF_{v,k} = \widehat{CF}_k$. We can then compute the subgrid surface cloud fraction $CF_{s,sg,k}$ as the total cloud cover of a single coarse layer, by using Eq. (14), but setting to zero the volume cloud fractions above and below the coarse layer considered ($N$=1) :

$$CF_{s,sg,k} = 1 - (1 - CF_{v,k})(\alpha_{sg} + (1 - \alpha_{sg})(1 - CF_{v,k}))^{n-1} \tag{17}$$

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

The top panels of Fig. 6 show the profile of $CF_s$ obtained using the LES original data, using Eq. (17), and also assuming maximum overlap within each layer, for two coarse resolutions (left panel at $Dz$=100 $m$ and right panel $Dz$=200 $m$). When using Eq. (17), two slightly different values of $\alpha$ are used for $Dz$=100 $m$ ($\alpha_{25,100}$=0.921) and $Dz$=200 $m$ ($\alpha_{25,200}$=0.911), to ensure that the total cloud cover is the same. The maximum overlap assumption (grey) does a poor job representing the surface cloud fraction profile, and leads to a relative error of 30% to 50%. It shows the error made when neglecting subgrid variability, i.e. assumming $CF_s = CF_v$ on the coarse grid. For this assumption, the coarser the vertical resolution, the larger the error. Using Eq. (17) allows a better representation of the surface cloud fractions, even if a substantial error remains. For all methods, the largest error corresponds to the lower layer which is the bottom of the cloud layer. On this layer the volume cloud fraction $CF_v$ decreases steeply, which makes the hypothesis of a constant $CF_v$ inaccurate.

To go further we also compare the performance of Eq. (17) with that of other references in the litterature. Neggers et al. (2011) and Jouhaud et al. (2018) have both been developed using LES data of small cumulus with $CF_v \approx 0.1$, including the ARMCu and BOMEX cases, and are therefore comparable to our method. Brooks et al. (2005) develops a lidar and radar-based parametrization of $CF_s$ using $CF_v$, with the possibility 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 $CF_s$ of the LES.

Our approach favours an accurate cloud cover on the whole vertical extent of the cloud layer. Results show that with this approach we tend to underestimate the surface cloud fraction of the coarse layers. This is because the overlap parameter $\alpha$ has been computed to match the total cloud cover of the whole scene, not the surface cloud fraction $CF_s$ of each coarse layer. When only used for the subgrid scale it creates too small a surface cloud fraction. This underestimation is still much smaller than when considering maximum overlap. The gap in surface cloud fraction caused by using our method is similar to those caused by other approximations of the litterature, but whith an opposite sign in the difference. Our underestimation of $(CF_s)_z$ was already visible in Fig. 2 on the panel showing "cloud cover above $z$". The only difference between using subgriding or not is the hypothesis $CF_{vol} = cst$ in each coarse layers, so we can conclude than the underestimation of our method comes from this hypothesis.

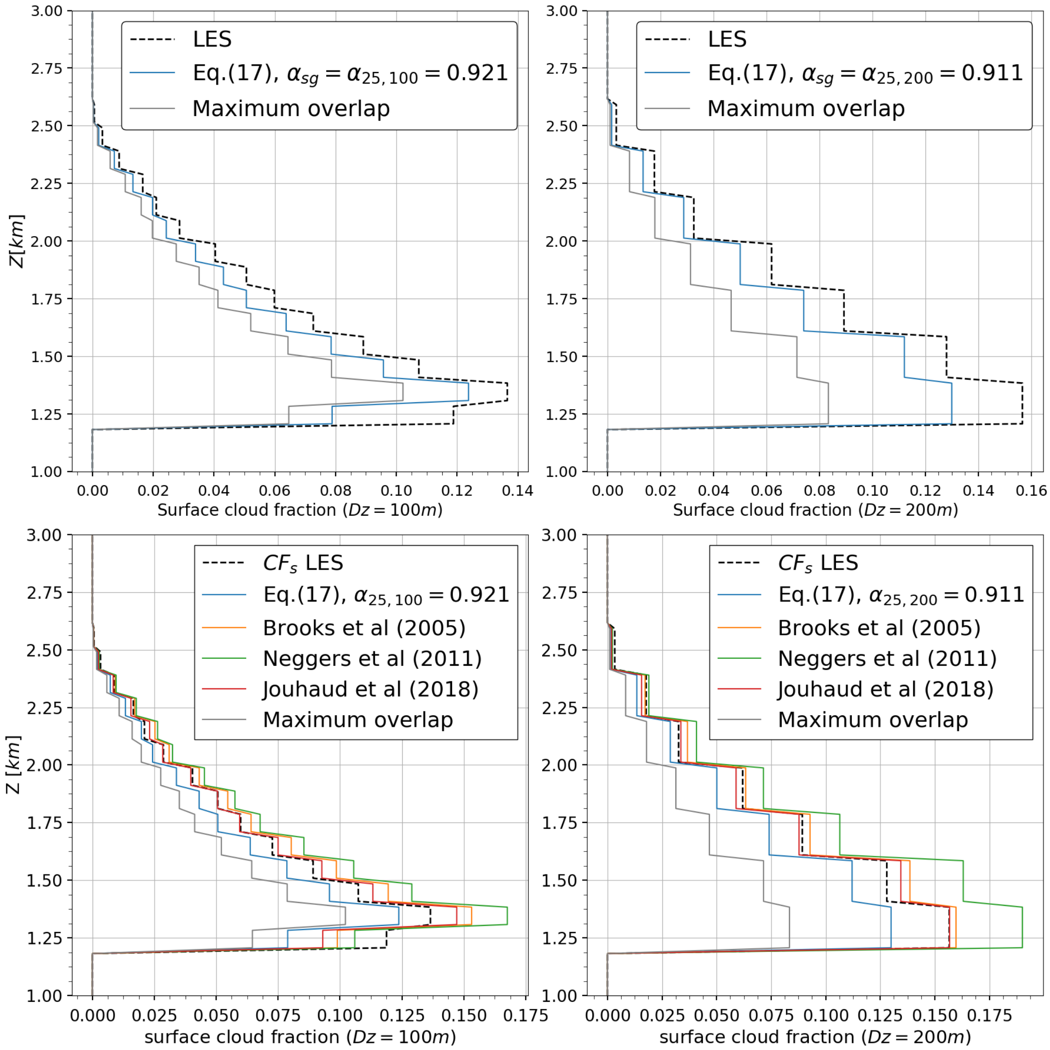


**Figure 6.** Vertical distribution of the surface cloud fraction $(CF_s)_z$ obtained with LES full resolution results or with different approximations with a coarse vertical resolution of 100 $m$ (left panels) or 200 $m$ (right panels). The top panels compare the LES (dashed black) with ERO using Eq. (17) and $\alpha_{sg}=\alpha_{25,Dz}$ (blue) as well as the maximum overlap sample (grey). The bottom panels also compare Eq. (17) with other parametrizations found in the litterature. The cloud case is ARMCu ($h$=10).

### 4.2 Interlayer overlap

We now consider that the vertical profile of the surface cloud fraction $(CF_{s,sg})_z$ 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 total cloud cover $CC$. To compute the subgrid surface cloud fraction profile $(CF_{s,sg})_z$ in the previous section, we were using the first part of Eq. (14), which represents the subgrid overlap. We here use the second part of the equation, which represents the interlayer overlap, using the unknown interlayer overlap $\alpha_{inter}$.

This corresponds to using Eq. (10) on the coarse grid with $(CF_{s,sg})_z$ to produce the total cloud cover:

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

The overlap parameter $\alpha_{inter}$ can be computed as in the previous sections, by inverting Eq. (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 Eq. (12), the subgrid cloud fractions $(CF_{s,sg})_z$ using Eq. (17) with $\alpha_{sg}$=$\alpha_{25,Dz}$, and then the overlap parameter $\alpha_{inter}$ using Eqs. (18,19) in order to overlap these coarse layers to produce the total cloud cover $CC$. The corresponding decorrelation length can be computed with Eq. (16) and $Dz$ as the separation distance. However, at this stage, there is no evidence of a formal link between these two overlap parameters or decorrelation lengths, or of a dependence to the vertical resolution. In any case, we have not found one.

We find that $\alpha_{25,Dz}$ and the corresponding decorrelation length (Fig. 7, blue plots, left and middle panels) 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 Eq. (17) we then compute the subgrid profile $(CF_{s,sg})_z$, as well as the interlayer overlap parameter $\alpha_{inter}$ using Eqs. (18,19).

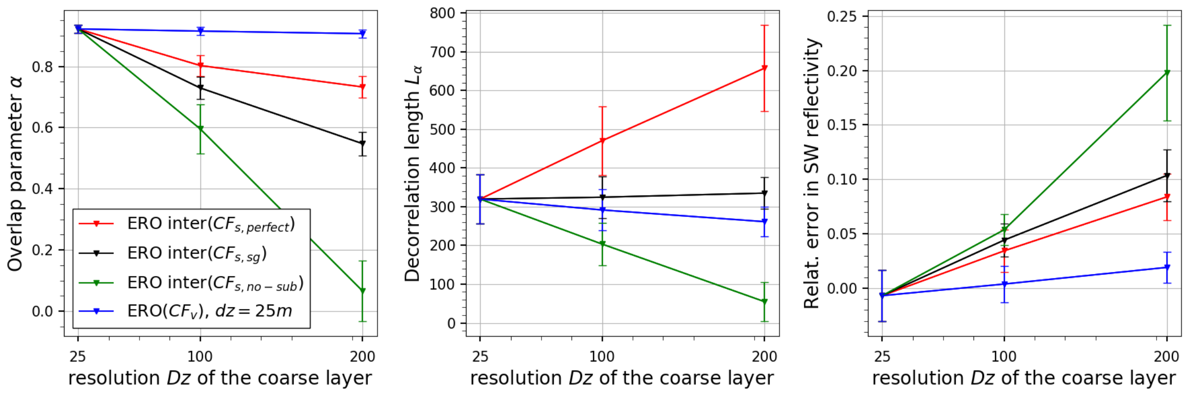


**Figure 7.** Overlap parameters (left) and decorrelation lengths (middle) for the ARMCu simulations (hours 6 to 12), for different coarse resolutions $Dz$ and for different reconstructions using ERO (see text). The daily mean value is shown. The overlap parameters are computed to match the total cloud cover of the 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.

We found that the overlap parameter $\alpha_{inter}$ varies with the resolution $Dz$ but the corresponding decorrelation length varies little from $\overline{L_{\alpha,sg}}$=326 $m$ (Fig. 7, black plots, 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 (Fig. 7, middle panel, blue and black lines). When it comes to radiative effects (Fig. 7, right panel), the error made on the SW cloud albedo is still small even when computed on the coarse grid (black plot) rather than on the finer grid (blue plot).

### 4.4 Analysis and comparisons of interlayer overlap for different estimations of the surface cloud fraction

Here we investigate, using Eqs. (18,19), how the overlap parameter $\alpha_{inter}$ and the decorrelation length should vary to keep the correct value of the total cloud cover for different estimations of the surface cloud fraction $CF_s$ in Eq. (18), instead of $CF_{s,sg}$. First we consider the extreme case where no subgrid heterogeneity is considered (Fig. 7, green plots), meaning the subgrid surface cloud fraction equals the volume cloud fraction $(CF_{s,no-sub})_z$=$(CF_v)_z$ 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 Fig. 6, for any altitude $z$ we have : $CF_{v,z} < CF_{s,z}$, so to generate the same total cloud cover, the overlap when no subgrid is taken into account has to be closer to random (i.e. $\alpha$ 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 surface cloud cover profile $(CF_{s,perfect})_z$ (Fig. 7, red plots). We then compute the interlayer overlap corresponding to this profile with Eqs. (18,19). The same reason applies to explain the difference with the interlayer overlap parameters computed for the subgrid cloud fraction profile: as shown in Fig. 6, $CF_{s,sg}$ approaches $CF_{s,perfect}$ in such a way that for any altitude $CF_{s,perfect} > CF_{s,sg} > CF_{s,no-sub}$. To conserve the same total cloud cover we then get $\alpha_{inter,perfect} > \alpha_{inter,sg} > \alpha_{inter,no-sub}$.

The middle panel of Fig. 7 shows the corresponding decorrelation lengths, computed from each overlap parameter $\alpha$ 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 plots). When doing ERO on the finer grid, the final resolution is $dz$=25 $m$ (blue plots). We see that for interlayer overlap, the decorrelation lengths have a strong dependence to the resolution when overlapping coarse layers of which the surface fraction is either perfect $(CF_{s,perfect})_z$ or determined assuming no subgrid heterogeneity $(CF_{s,no-sub})_z$, with important variations. This is not the case when the surface cloud fraction $CF_{s,sg}$ 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 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 $(CF_{s,perfect})_z$ and $(CF_{s,no-sub})_z$, but a small dependence to the resolution of the decorrelation length when overlapping $CF_{s,sg}$ . This dependence of $L_\alpha$ 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 litterature when generating cloudy subcolumns from GCMs or for observational simulators (Pincus et al. (2005); Bodas-Salcedo et al. (2011); 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 (Fig. 4 and blue plots of Fig. 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 suface cloud fraction as a bulk subgrid property using the volume cloud fraction and a subgrid overlap parameter. Overlapping this computed subgrid cloud fraction leads to a relative error in cloudy albedo of $\approx 10\%$ for coarse resolutions of 100 $m$ and 200 $m$ (Fig. 7, black plot). 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 $m$ and 200 $m$ (Fig. 7, red plot). Finally, even without taking into account any subgrid scale by overlapping $(CF_{s,no-sub})_z$ on the coarse grid, we can approach the albedo of the LES scenes within a 20% relative error (for a resolution of 200 $m$, Fig. 7, green plot) if the total cloud cover is reproduced. As all the generations shown in Fig. 7 have the same total 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 total 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 ARMCu cloud case, but similar results were also found consistently in several other cases, see Figs. S1-S3 in Supporting Information.

## 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 exponential-random overlap. 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 coarse 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 Fig. 7.

The procedure for reconstructing a cloud scene that we propose is as follow: given any volume cloud fraction profile $(CF_v)_z$ at resolution $Dz$ and the decorrelation length $L_\alpha$ for a reference resolution (here $dz$=25 $m$), the subgrid heterogeneity is taken into account by computing a profile of the surface cloud fraction $(CF_{s,sg})_z$ with Eq. (17), with $n$=$Dz/dz$ in the equation. The same decorrelation length $L_{\alpha,}$ allows to overlap these coarse layers and to compute the total cloud cover (Eq. (18)). As we can see on Fig. 7 for the case studied here, $L_{\alpha,25,Dz}$≈291 $m$ and $L_{\alpha,sg}$≈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 Fig. 7 for three other cumulus cloud cases simulated by the same LES and the same resolutions, with $L_{\alpha,25,Dz}$ and $L_{\alpha,sg}$ relatively independent of the resolution. 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 Figs. S1-S3 in Supporting Information). 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 Fig. 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 the Supporting Information on Figs. S4 to S6.

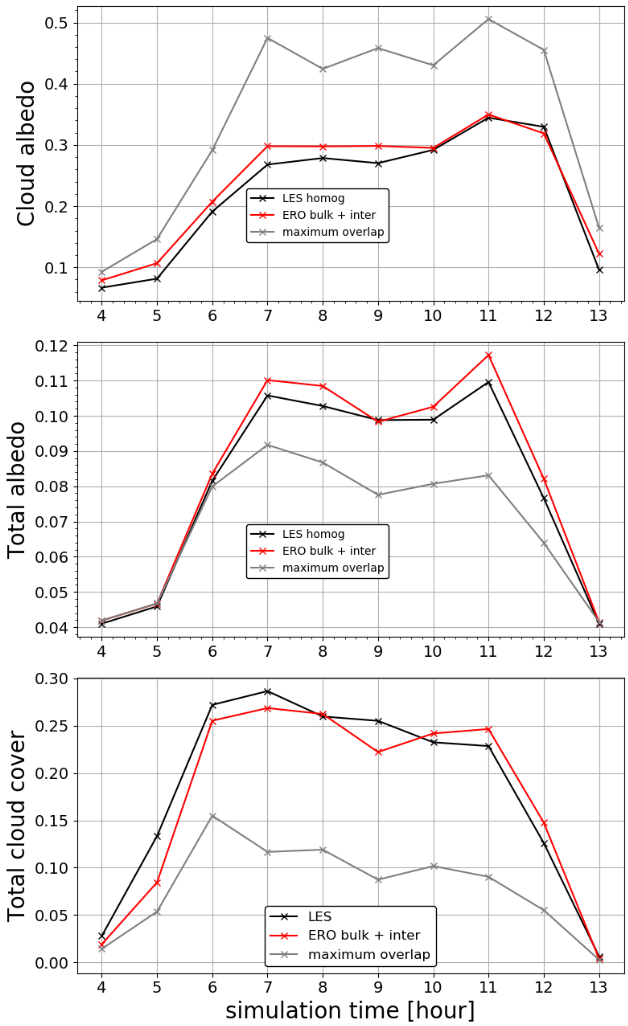


**Figure 8.** Cloud albedo (top panel), total albedo (middle panel) and total cloud cover (lower panel) for the LES (in red), our reconstruction using ERO (in black) and a maximum overlap reconstruction (grey). The constant decorrelation length used here both for the subgrid computation of the surface cloud fraction profile and its interlayer overlap is $L_\alpha$=309 $m$. The scenes are the ARMCu case (time steps $h$∈[4, 13]). In all scenes the LWC is homogeneous at each vertical level.

### 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 (Oreopoulos and Norris (2011); Jing et al. (2016)). 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 entierely cloudy even if the cloud does not fully extend on the vertical of the layer. Hence the measured profile is the surface cloud fraction $(CF_s)_z$ for a coarse layer of thickness $Dz$=480 $m$. Combining Eqs. (17,18,19), 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 $dz_1$=$n$×$dz_2$. 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 : $(CF_{s,1})_z$, $L_{\alpha,1}$, and $(CF_{s,2})_z$, $L_{\alpha,2}$, where $L_{\alpha,i}$ are the decorrelation lengths corresponding to the measured surfacic cloud fraction profiles.

Using the cloud fraction profile with finer vertical resolution $CF_{s,2}$ we can use interlayer ERO with $L_{\alpha,2}$ on blocks of $n$ fine layers to compute the corresponding surface cloud fraction profile at the resolution $dz_1$, $CF'_{s,1}$. Knowing the total cloud cover $CC$, we can then compute with Eq. (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 ARMCu simulations used on Fig. 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 $L_\alpha$ with the vertical resolution for a perfect estimation of the surface cloud fraction profile. $I_2$ would measure a decorrelation length $L_{\alpha,2}$=320 $m$, while $I_1$ would measure $L_{\alpha,1}$=658 $m$ (Fig. 7 middle panel, in red). We get a factor 2 on the estimation of the decorrelation length in this case. The vertical extension of the studied clouds is too small to be able to compute the decorrelation lenght 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 $L_\alpha \sim 2\ km$ (Hogan and Illingworth (2000); Willèn et al. (2005); Barker (2008a); Oreopoulos and Norris (2011); Jing et al. (2016)). 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 litterature with LES simulations with similar vertical resolutions (Neggers et al. (2011); Sulak et al. (2020); Villefranque et al. (2021)). The difference in horizontal resolutions (Naud et al. (2008); Astin and Di Girolamo (2014); Tompkins and 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 and 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 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) :

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

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 $Re(k) > 0$ :

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

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 Fig. 9 the LWC of both scenes' cloudy subcolumns after they have been sorted along their vertical 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 caracteristics 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 $(CF_v)_z$ 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 ARMCu 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 ARMCu 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 $\approx 10\%$ brighter, which confirms the previous findings of Barker et al. (2003), Wu and Liang (2005), and Shonk and Hogan (2010).

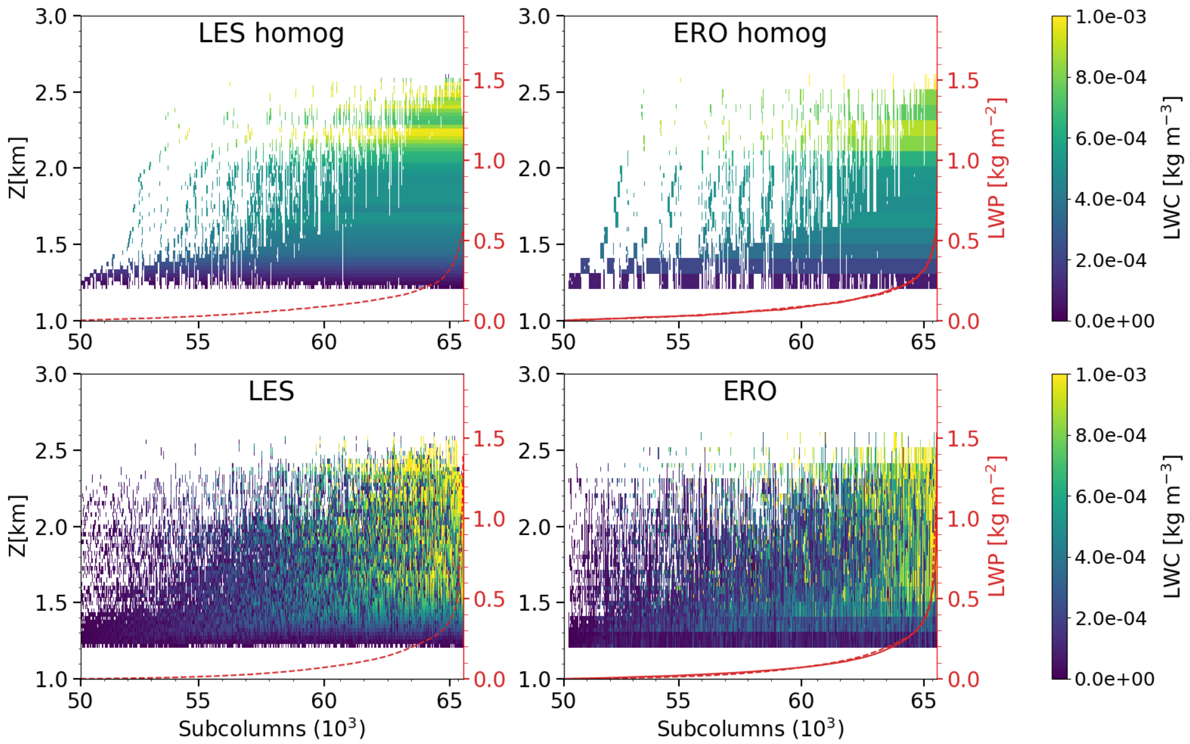


**Figure 9.** The liquid water content of each scene's cloudy subcolumns in the LES simulations (left panels) and reconstructed using ERO (right panels). The subcolumns have been sorted along their LWP (red plots). The red lines represent the LES (dashed line) and generated (solid line) LWP, the former being represented on the right panels as well to facilitate the comparison. Top panels are homogeneous LWC for each level and whereas it varies in the bottom panels.

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 exponential-random overlap (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.

Within the McICA framework, 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 total cloud cover or the overlap parameter. Compared to reference LES simulations, the generated cloud scenes show a correct representation of both the distribution of cumulative cloud fraction among cloudy subcolumns and the vertical profile of the cloud cover seen from above or below. We suggest that the later is a simple diagnostic that would usefully complement the usual cloud fraction vertical profile when comparing models with observations or when developping 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 reconstructions.

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 total 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 volumic or surfacic. Although this deserves more investigations, we provide a framework that allows to go from one vertical resolution to an other. Further work is also required to establish robust estimates of the decorrelation length for a large variety of clouds.

To our best knowledge, most current atmospheric models neglect the effect of subgrid variability on the cloud fraction and assume a maximum-random overlap of cloud layers or a ERO with a quite large decorrelation length ($\approx 2{-}3$ $km$). This can lead to an underestimation of the cloud cover by a factor of 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)). 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.

In addition to the effect of the water content heterogeneity on cloud albedo, already well recognized, we show that the vertical distribution of cloud fraction also matters. Indeed, 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 how the vertical profile of the clouds fraction is represented. 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% to 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 levels $[k_{base}, k_{top}]$ (with $\#([k_{base}, k_{top}]) = \mathcal{N}$ 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 coefficient $c_{i,k} \in \{0, 1\}$, which corresponds to whether the cell is cloudy or not , as well as the liquid water content.

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

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

where $RN1$ are random numbers evenly distributed on $[0, 1]$. Working its way down, the algorithm computes the next coefficients, 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 cloudy state $c_{i,k}$ is computed as :

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

  where $(c_k|c_{k-1})_{max}$ are booleans computed according to the transition probabilities $P_{max}(C_k = c_k \mid C_{k-1} = c_{k-1})$ which is defined by Eq. 6 when $C_k$=$C_{k-1}$. To complete this implementation, according to Eq. (2), we also have:

  $$P_{max}(\mathrm{C}_k = 1 \mid \mathrm{C}_{k-1} = 0) = 1 - P_{max}(\mathrm{C}_k = 0 \mid \mathrm{C}_{k-1} = 0) = \frac{\max(CF_{k-1}, CF_k)}{1 - CF_{k-1}} \tag{A2}$$

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

  $$c_{i,k} = (1 \mid 1)_{rand} = (1 \mid 0)_{rand}$$

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

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

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 amout) of a given cell $k$, Räisänen generator computes $x_k \in [0, 1]$ to compare it to the cloud fraction of the cell $CF_k$ and decide wether 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 levels $k$ 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 levels our method is equivalent, but for more than two levels, 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 ($P_{max}(1 \mid 1), P_{max}(1 \mid 0), P_{rand}(1 \mid 1), P_{rand}(1 \mid 0)$), and by using a different random number every time it is needed, we conserve the cloud fraction without creating this correlation between the layers.

**Acknowledgments**

Our many thanks go to Céline Cornet and Frédéric Szczap for insightful discussions about this work. 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). A repository containing the scripts for the ERO algorithm presented in this paper is available at https://github.com/raphleb/ERO.git. The sources described in this paper for the radiative computations are available at the websites (https://www.meso-star.com/projects/htrdr/htrdr.html and https://www.meso-star.com/projects/star-engine/star-engine.html).

## References

Astin, I., & Di Girolamo, L. (2014). Technical note: The horizontal scale dependence of the cloud overlap parameter . *Atmos. Chem. Phys.*, *14*(18), 9917–9922. doi: 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. *J. Geophys. Res.-Atm.*, *113*(D8). doi: 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. *J. Geophys. Res.-Atm.*, *113*(D24). doi: https://doi.org/10.1029/2008JD010391

Barker, H. W., & Räisänen, P. (2005). Radiative sensitivities for cloud structural properties that are unresolved by conventional GCMs. *Q. J. R. Meteorol. Soc.*, *131*(612), 3103-3122. doi: 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. *Q. J. R. Meteorol. Soc.*, *125*(558), 2127-2152. doi: https://doi.org/10.1002/qj.49712555810

Barker, H. W., Stephens, G. L., Partain, P. T., Bergman, J. W., Bonnel, B., Campana, K., ... Yang, F. (2003). Assessing 1D Atmospheric Solar Radiative Transfer Models: Interpretation and Handling of Unresolved Clouds. *J. Climate*, *16*(16), 2676 - 2699. doi: 10.1175/1520-0442(2003)016⟨2676:ADASRT⟩2.0.CO;2

Bergman, J. W., & Rasch, P. J. (2002). Parameterizing vertically coherent cloud distributions. *J. Atmos. Sci.*, *59*(14), 2165-2182. doi: 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., ... John, V. O. (2011). COSP: Satellite simulation software for model assessment. *Bull. Am. Meteorol. Soc.*, *92*(8), 1023 - 1043. doi: 10.1175/2011BAMS2856.1

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. *J. Atmos. Sci.*, *62*(7), 2248-2260. doi: 10.1175/JAS3467.1

Brown, A. R., Cederwall, R. T., Chlond, A., Duynkerke, P. G., Golaz, J.-C., Khairoutdinov, M., ... Stevens, B. (2002). Large-eddy simulation of the diurnal cycle of shallow cumulus convection over land. *Q. J. R. Meteorol. Soc.*, *128*(582), 1075-1093. doi: https://doi.org/10.1256/003590002320373210

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

Geleyn, J., & Hollingsworth, A. (1979). An economical analytical method for the computation of the interaction between scattering and line absorption of radiation. *Beitr. Phys. Atmosph.*.

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

Hogan, R. J., & Illingworth, A. J. (2000). Deriving cloud overlap statistics from radar. *Q. J. R. Meteorol. Soc.*, *126*(569), 2903-2909. doi: 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. *J. Atmos. Sci.*, *70*(2), 708 - 724. doi: 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. *Q. J. R. Meteorol. Soc.*, *125*(555), 941-965. doi: 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. doi: 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. *J. Adv. Model. Earth Syst.*, *10*(11), 2686-2705. doi: https://doi.org/10.1029/2018MS001379

Konsta, D., Dufresne, J.-L., Chepfer, H., Vial, J., Koshiro, T., Kawai, H., … Ogura, T. (2022). Low-level marine tropical clouds in six cmip6 models are too few, too bright but also too compact and too homogeneous. *Geophys. Res. Lett.*, *49*(11), e2021GL097593. doi: https://doi.org/10.1029/2021GL097593

Koren, I., Oreopoulos, L., Feingold, G., Remer, L. A., & Altaratz, O. (2008). How small is a small cloud? *Atmos. Chem. Phys.*, *8*(14), 3855–3864. doi: 10.5194/acp-8-3855-2008

Lac, C., Chaboureau, J.-P., Masson, V., Pinty, J.-P., Tulet, P., Escobar, J., … Wautelet, P. (2018). Overview of the Meso-NH model version 5.4 and its applications. *Geosci. Model Dev.*, *11*(5), 1929–1969. doi: 10.5194/gmd-11-1929-2018

Lafore, J. P., Stein, J., Asencio, N., Bougeault, P., Ducrocq, V., Duron, J., … Vilà-Guerau de Arellano, J. (1998). The Meso-NH Atmospheric Simulation System. Part I: adiabatic formulation and control simulations. *Annales Geophysicae*, *16*(1), 90–109. doi: 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. *J. Atmos. Sci.*, *59*(24), 3519 - 3539. doi: 10.1175/1520-0469(2002)059⟨3519:SSAMVI⟩2.0.CO;2

Mace, G. G., & Benson-Troth, S. (2002). Cloud-layer overlap characteristics derived from long-term cloud radar data. *J. Climate*, *15*(17), 2505 - 2515. doi: 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. *J. Climate*, *21*(8), 1758 - 1770. doi: 10.1175/2007JCLI1828.1

Neggers, Duynkerke, P. G., & Rodts, S. M. A. (2003b). Shallow cumulus convection: A validation of large-eddy simulation against aircraft and landsat observations. *Q. J. R. Meteorol. Soc.*, *129*(593), 2671-2696. doi: https://doi.org/10.1256/qj.02.93

Neggers, Heus, T., & Siebesma, A. P. (2011). Overlap statistics of cumuliform boundary-layer cloud fields in large-eddy simulations. *J. Geophys. Res.-Atm.*, *116*(D21). doi: 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. *J. Atmos. Sci.*, *60*(8), 1060 - 1074. doi: 10.1175/1520-0469(2003)60⟨1060:SSOCCP⟩2.0.CO;2

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

Oreopoulos, L., & Norris, P. M. (2011). An analysis of cloud overlap at a midlatitude atmospheric observation facility. *Atmos. Chem. Phys.*, *11*(12), 5557–5567. doi: 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. *J. Geophys. Res.-Atm.*, *108*(D13). doi: 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. *J. Geophys. Res.-Atm.*, *110*(D15). doi: 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. *Q. J. R. Meteorol. Soc.*, *130*(601), 2047-2067. doi: 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. *J. Climate*, *21*(11), 2352 - 2370. doi: 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. *Q. J. R. Meteorol. Soc.*, *136*(650), 1205-1215. doi: https://doi.org/10.1002/qj.646

Siebesma, A. P., Bretherton, C. S., Brown, A., Chlond, A., Cuxart, J., Duynkerke, P. G., ... Stevens, D. E. (2003). A large eddy simulation intercomparison study of shallow cumulus convection. *J. Atmos. Sci.*, *60*(10), 1201 - 1219. doi: 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. *J. Geophys. Res.-Atm.*, *125*(10), e2019JD032017. doi: 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. *Geosci. Model Dev.*, *11*(1), 77–81. doi: 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. *J. Atmos. Sci.*, *64*(6), 2116-2125. doi: 10.1175/JAS3925.1

Tompkins, A. M., & Di Giuseppe, F. (2015). An interpretation of cloud overlap statistics. *J. Atmos. Sci.*, *72*(8), 2877 - 2889. doi: 10.1175/JAS-D-14-0278.1

vanZanten, M. C., Stevens, B., Nuijens, L., Siebesma, A. P., Ackerman, A. S., Burnet, F., ... Wyszogrodzki, A. (2011). Controls on precipitation and cloudiness in simulations of trade-wind cumulus as observed during RICO. *J. Adv. Model. Earth Syst.*, *3*(2). doi: https://doi.org/10.1029/2011MS000056

Villefranque, N., Blanco, S., Couvreux, F., Fournier, R., Gautrais, J., Hogan, R. J., ... Williamson, D. (2021). Process-based climate model development harnessing machine learning: III. the representation of cumulus geometry and their 3D radiative effects. *J. Adv. Model. Earth Syst.*, *13*(4), e2020MS002423. doi: https://doi.org/10.1029/2020MS002423

Villefranque, N., Fournier, R., Couvreux, F., Blanco, S., Cornet, C., Eymet, V., ... Tregan, J.-M. (2019). A Path-Tracing Monte Carlo Library for 3-D Radiative Transfer in Highly Resolved Cloudy Atmospheres. *J. Adv. Model. Earth Syst.*, *11*(8), 2449-2473. doi: 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) doi: 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. *J. Atmos. Sci.*, *62*(11), 4105 - 4112. doi: 10.1175/JAS3565.1

# Supporting Information for "A consistent representation of cloud overlap and cloud subgrid vertical heterogeneity"

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

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

## Contents of this file

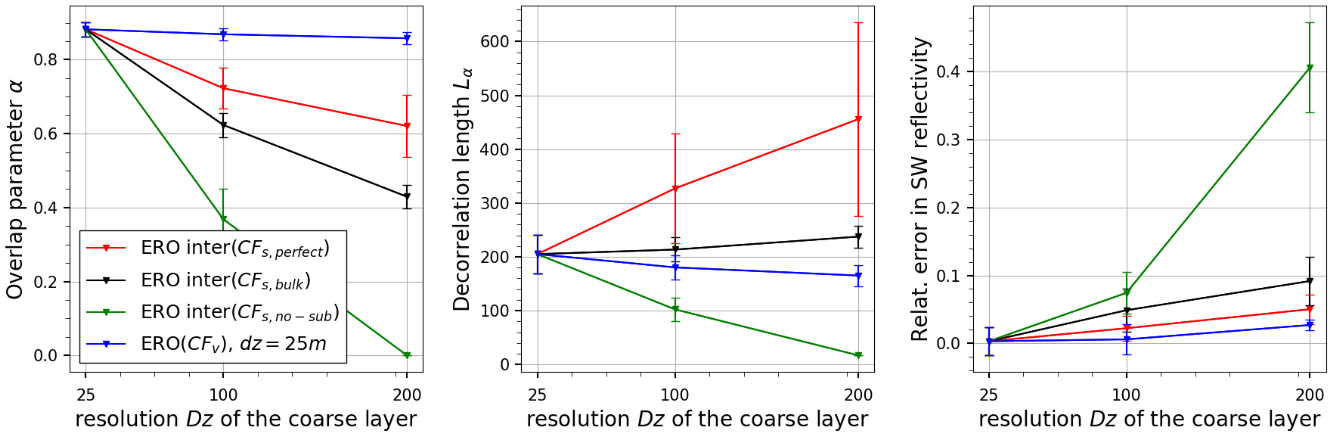


**Figure S1.** Overlap parameters (left) and decorrelation lengths (middle) for the BOMEX simulations (hours 1 to 15), for different coarse resolutions Dz and for different reconstructions using ERO. The daily mean value is shown. The overlap parameters are computed to match the total cloud cover of the 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.

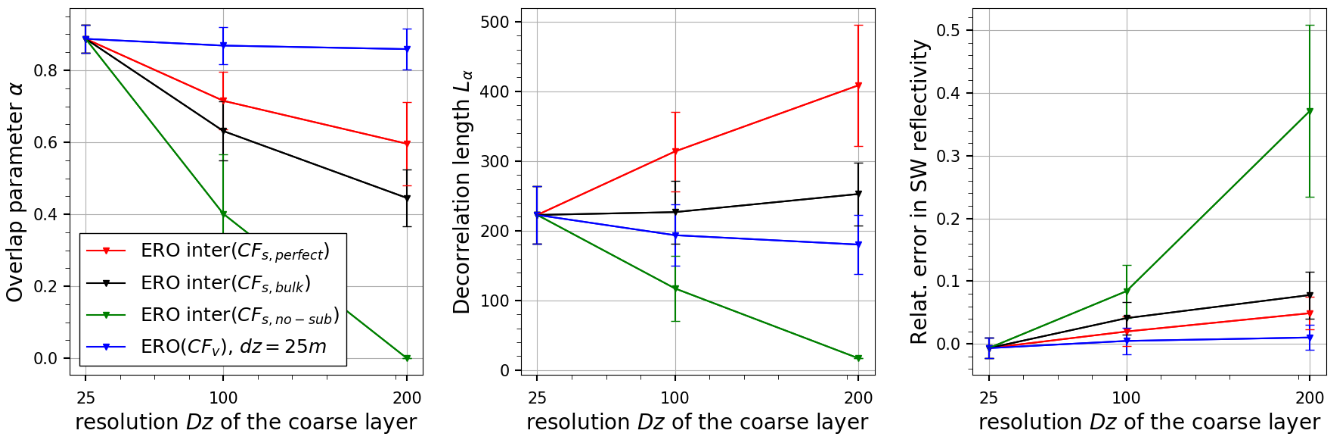


**Figure S2.** Same plots for the RICO simulations (hours 1 to 15)

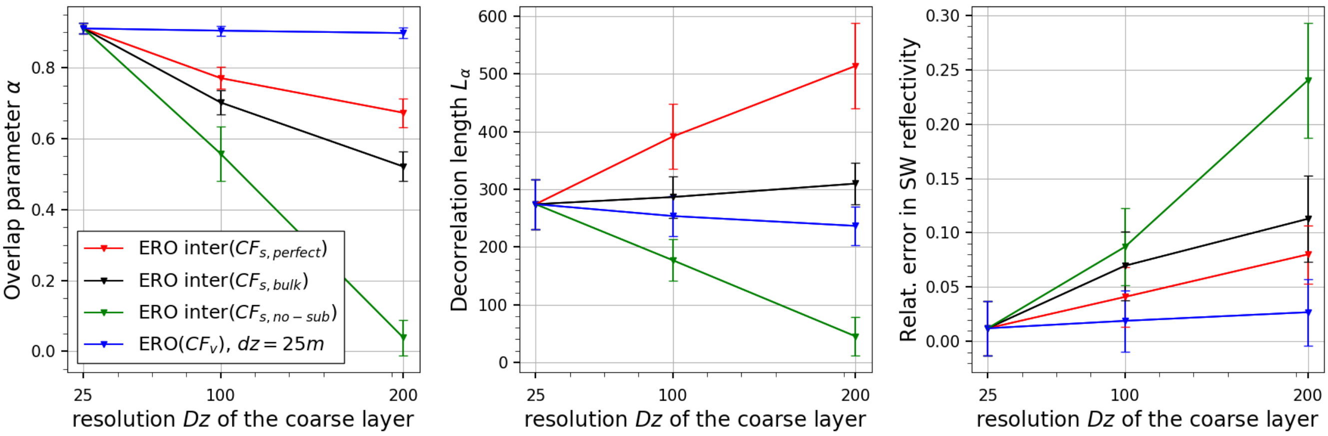


**Figure S3.** Same plots for the SCMS simulations (hours 2 to 12)

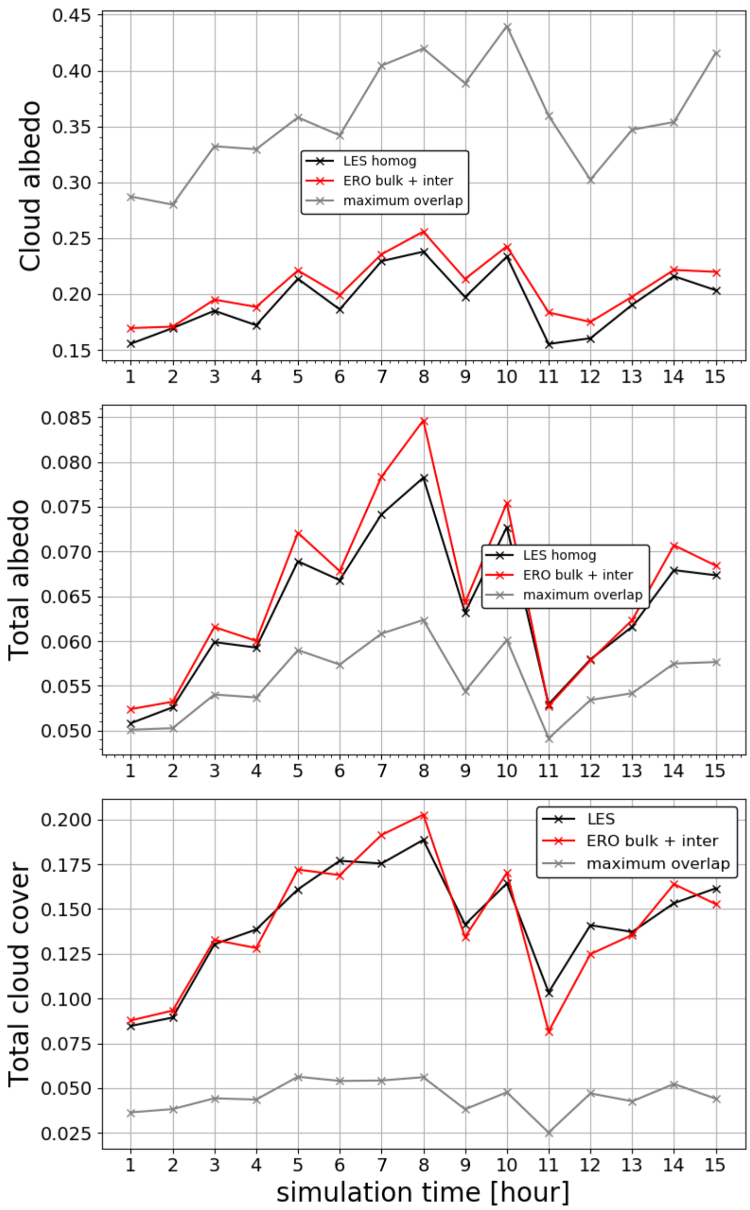


**Figure S4.** Cloud albedo (top panel), total albedo (middle panel) and total cloud cover (lower panel) for the LES (in red), our reconstruction using ERO (in black) and a maximum overlap reconstruction (grey). The constant decorrelation length used here both for the subgrid computation of the surface cloud fraction profile and its interlayer overlap is $L_\alpha$=202 $m$. The cloud albedo of the ERO reconstruction shows a relative error of 7% on the whole day compared to the LES cloud albedo. The scenes used are the BOMEX case (simulation hours $h \in [1, 15]$). In all scenes the LWC is homogeneous at each vertical level.

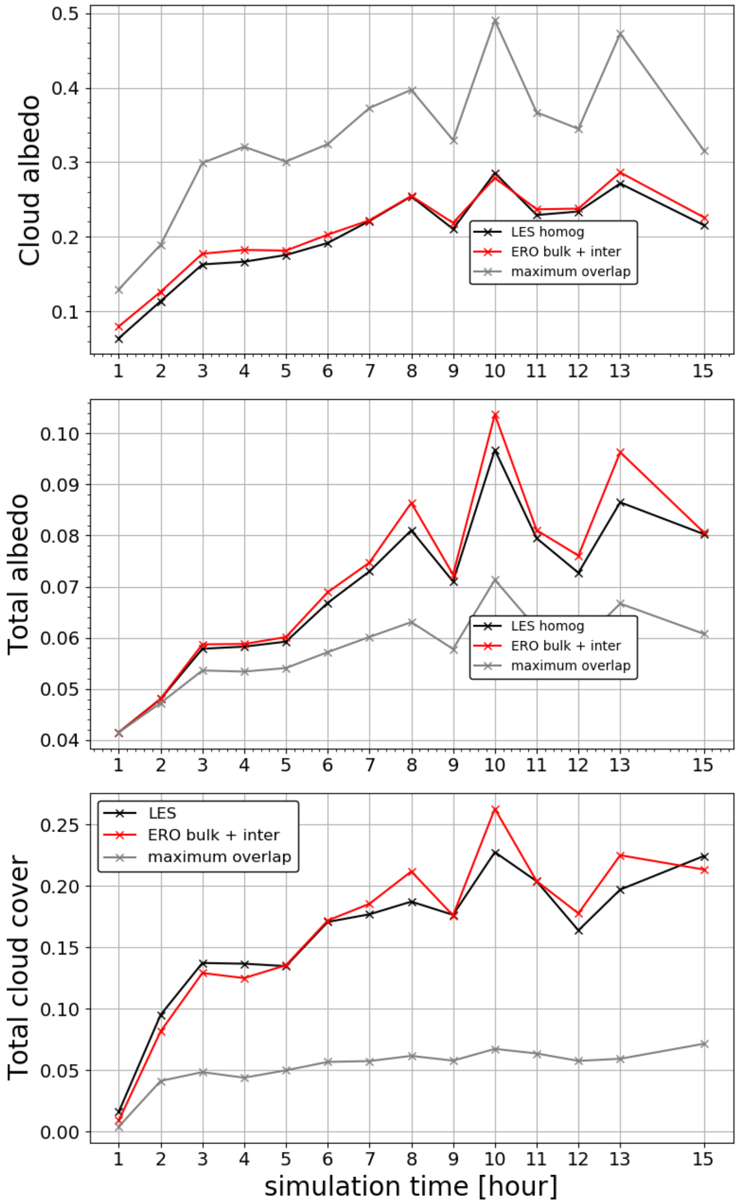


**Figure S5.** Same plots for the RICO case (simulation hours $h \in [1, 15]$) with $L_\alpha$=217 $m$. The cloud albedo of the ERO reconstruction shows a relative error of 6% on the whole day compared to the LES cloud albedo.

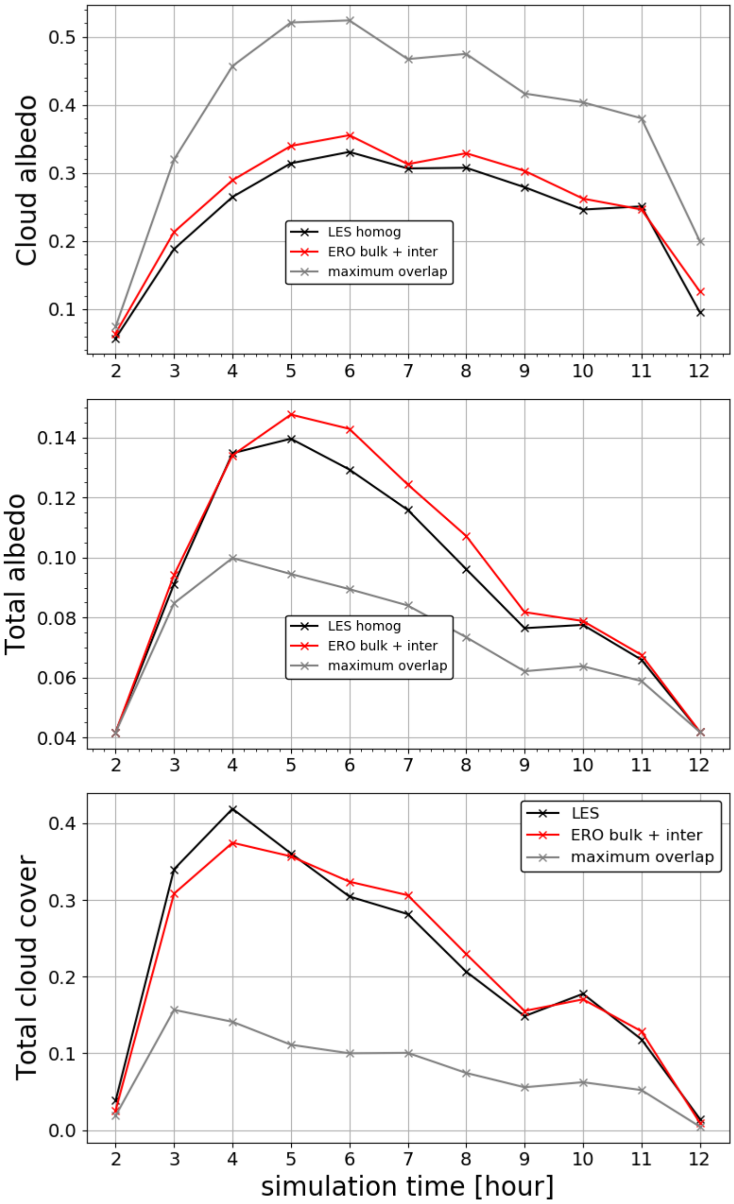


**Figure S6.** Same plots for the SCMS case (simulation hours $h \in [2, 12]$) with $L_\alpha$=273 $m$. The cloud albedo of the ERO reconstruction shows a relative error of 10% on the whole day compared to the LES cloud albedo.