

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

Raphaël Lebrun<sup>1</sup>, Jean-Louis Dufresne<sup>1</sup>, Najda Villefranque<sup>1</sup>

<sup>1</sup>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: Raphael Lebrun, [raphael.lebrun@lmd.ipsl.fr](mailto:raphael.lebrun@lmd.ipsl.fr)

**Abstract**

13  
14 Many global climate models underestimate the cloud cover and overestimate the cloud  
15 albedo, especially for low-level clouds. We determine how a correct representation of the  
16 vertical structure of clouds can fix part of this bias. We use the 1D McICA framework  
17 and focus on low-level clouds. Using LES results as reference, we propose a method based  
18 on exponential-random overlap (ERO) that represents the cloud overlap between lay-  
19 ers and the subgrid cloud properties over several vertical scales, with a single value of  
20 the overlap parameter. Starting from a coarse vertical grid, representative of atmospheric  
21 models, this algorithm is used to generate the vertical profile of the cloud fraction with  
22 a finer vertical resolution, or to generate it on the coarse grid but with subgrid hetero-  
23 geneity and cloud overlap that ensures a correct cloud cover. Doing so we find decorre-  
24 lation lengths are dependent on the vertical resolution, except if the vertical subgrid het-  
25 erogeneity and interlayer overlap are taken into account coherently. We confirm that the  
26 frequently used maximum-random overlap leads to a significant error by underestimat-  
27 ing the low-level cloud cover with a relative error of about 50%, that can lead to an er-  
28 ror of SW cloud albedo as big as 70%. Not taking into account the subgrid vertical het-  
29 erogeneity of clouds can cause an additional relative error of 20% in brightness, assum-  
30 ing the cloud cover is correct.

**Plain Language Summary**

31  
32 Low-level clouds are the main source of spread in model estimates of climate sen-  
33 sitivity, but climate models resolutions do not allow them to explicitly resolve the ge-  
34 ometrical complexity of low-level clouds, which must be parametrized. Most climate mod-  
35 els low-level clouds have a cloud cover too small and a cloud albedo too high, which is  
36 known as the “too few too bright bias”. In this work we determine whether a better rep-  
37 resentation of the vertical structure of clouds can fix part of this bias. We use high-resolution  
38 simulations as references and radiative transfer algorithms to assess the performances  
39 of our cloud generation, in the framework of commonly used overlap assumptions. When  
40 the cloud cover of the scene is known, we show that the exponential-random overlap al-  
41 lows a good representation of the vertical structure of clouds and of the cloud albedo.  
42 We find the decorrelation lengths used to model the overlap are highly dependent on the  
43 model vertical resolution, and present a way to overcome this dependency when both sub-  
44 grid scale and interlayer overlap are taken into account consistently. We present values  
45 that can be used to compute accurately the cloud cover and the cloud albedo of the stud-  
46 ied 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 fraction 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 explicitly 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 par-

99 allel assumption (each vertical profile represents a stack of horizontally infinite and ho-  
 100 mogeneous slabs) in our 1D approach. Assuming that the volumic cloud fraction and wa-  
 101 ter content are known on a coarse vertical grid consisting in a single column, typical of  
 102 an atmospheric model, we developed an algorithm to generate an ensemble of subcolumns  
 103 to statistically represent the heterogeneity of clouds.

104 The manuscript is organised as follows: in Section 2, we consider the exponential-  
 105 random overlap (ERO) as a Markov process and show its ability to represent the ver-  
 106 tical distribution of the cloud fraction over a wide range of scales that includes both the  
 107 subgrid scale and the overlap between layers. In Section 3 we study cloud scenes with  
 108 known cloud covers, and compute the overlap parameters and decorrelation lengths that  
 109 should be used with ERO on finer grids to reproduce those cloud covers, and doing so  
 110 we assess the radiative impact of ERO on the SW cloud albedo of the generated subcolumns.  
 111 We also study the effects of different simplifying assumptions. Section 4 focuses on re-  
 112 producing those results directly on the coarse grid, taking into account both the inter-  
 113 layer overlap and the subgrid scale, assuming again that the cloud cover is known. The  
 114 implication for cloud parameterization in atmospheric models and for how to estimate  
 115 the decorrelation lengths are presented in Section 5.

## 116 2 Statistical representation of the cloud fraction vertical distribution

117 The model explored here is the so-called exponential-random overlap (ERO) model  
 118 of Hogan and Illingworth (2000). We will only look at single-layer cumulus cloud fields  
 119 so the “random” part of the model, which concerns cloudy layers that are separated by  
 120 clear layers, will not be studied. The “exponential” part of the model states that the com-  
 121 bined cloud fraction of two adjacent cloudy layers of surfacic fractions  $CF_1$  and  $CF_2$  is:

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

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

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

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

$$CF_{1,2,rand} = CF_1 + CF_2 - CF_1CF_2$$

126 In this model, “exponential” refers to the fact that  $\alpha$  can be parametrized with an  
 127 exponential function (see further). This model has been used in two different manners  
 128 in radiative transfer parameterizations: either in a deterministic way, to compute the over-  
 129 lap matrix that is used to distribute downwelling and upwelling fluxes from clear and  
 130 cloudy regions of a layer into clear and cloudy regions of an adjacent layer ( TripleClouds,  
 131 Shonk and Hogan (2008)), or in a probabilistic manner, to generate a sample of verti-  
 132 cal profiles that preserve, when averaged, the principal characteristics of the cloud scene  
 133 (the cloud fraction and the liquid water content in each layer), and upon which radi-  
 134 ative transfer is simulated under the plane-parallel homogeneous assumption ( McICA,  
 135 Pincus et al. (2003)). In this paper, the McICA framework is used to generate samples  
 136 of vertical profiles. The main difference is that in the usual McICA algorithm, the pro-  
 137 files are generated on the vertical grid of the host model, while here we aim at generat-  
 138 ing profiles at any vertical resolution, including finer vertical resolutions.

139 Unless otherwise stated, in all this article, we consider a single vertical atmospheric  
 140 column that consists of a cloudy block (with a strictly positive liquid water content at  
 141 every level) of  $\mathcal{N}$  vertical layers. From this column we assume the volume cloud fraction  
 142 of each layer,  $(CF_k)_{k=1\dots\mathcal{N}}$ , is known. We consider the exponential-random overlap model  
 143 (ERO) as a Markovian process and deduce the relationship between the overlap param-  
 144 eter  $\alpha$  and the total cloud cover  $CC$ . We then use the same result to deal with subgrid  
 145 vertical heterogeneity.

## 146 2.1 ERO as a Markovian process: a sequence of conditional probab- 147 ities

148 Using a certain overlap scheme in an atmospheric column to generate a cloud frac-  
 149 tion distribution from top to bottom can be interpreted as a Markovian process as it is  
 150 a sequence of overlapping or non-overlapping events. It is then possible to compute its  
 151 outcome as a sequence of conditional probabilities, as done by Bergman and Rasch (2002).

152 In a single atmospheric column of  $\mathcal{N}$  vertical layers, let us consider a 1D subcol-  
 153 umn. We want to articulate how the overlap used for the whole atmospheric column trans-  
 154 lates to a subcolumn. If  $\vec{C} = (C_k)_{k=1\dots\mathcal{N}}$  is the random variable representing the cloud  
 155 fraction distribution of the subcolumn, with  $C_k \in \{0, 1\}$  (whether the cell is cloudy or  
 156 not), and  $k$  is the vertical index, with  $k = 1$  at the top of the column, the probability  
 157 of a certain state  $\vec{C} = (c_k)_{k=1\dots\mathcal{N}} \in [0, 1]^{\mathcal{N}}$  is given by:

$$P(\vec{C}) = \prod_{k=1}^{\mathcal{N}} P(C_k = c_k \mid C_{k-1} = c_{k-1}) \quad (1)$$

158 where  $C_0 = 0$  (i.e. there is no cloud above the cloud block considered here). We  
 159 use the classic upper case notation  $C_k$  for the random variables and the lower case no-  
 160 tation  $c_k$  for their realizations.

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

$$\begin{cases} P(C_k = 0 \mid C_{k-1} = 1) = 1 - P(C_k = 1 \mid C_{k-1} = 1) \\ P(C_k = 1 \mid C_{k-1} = 0) = 1 - P(C_k = 0 \mid C_{k-1} = 0) \end{cases} \quad (2)$$

169 Therefore, it is enough to know for instance the two transition probabilities  $P(C_k =$   
 170  $1 \mid C_{k-1} = 1)$  and  $P(C_k = 0 \mid C_{k-1} = 0)$  for each level  $k$  to compute the probability of  
 171 any given state of overlap for the column, using Eq.(1).

172 The transition probability  $P(C_k = 1 \mid C_{k-1} = 1)$  is the probability that both  
 173 levels of the subcolumn are cloudy, knowing that the level  $k-1$  is already cloudy. By  
 174 definition, we have:

$$P(C_k = 1 \mid C_{k-1} = 1) = \frac{P(C_k = 1 \cap C_{k-1} = 1)}{P(C_{k-1} = 1)} \quad (3)$$

175 where  $(C_k = 1 \cap C_{k-1} = 1)$  is the event with both layers cloudy. If we assume  
 176 an exponential-random overlap we have :

$$P(C_k = 1 | C_{k-1} = 1) = \alpha P_{max}(C_k = 1 | C_{k-1} = 1) + (1 - \alpha) P_{rand}(C_k = 1 | C_{k-1} = 1) \quad (4)$$

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

$$P_{rand}(C_k = 1 | C_{k-1} = 1) = P_{rand}(C_k = 1 | C_{k-1} = 0) = P_{rand}(C_k = 1) = CF_k \quad (5)$$

181 The transition probability in a subcolumn of the maximum overlap can be obtained  
 182 using Eq. (3): if  $CF_{k-1} < CF_k$ :  $P_{max}(C_k = 1 | C_{k-1} = 1) = 1$ , and on the con-  
 183 trary if  $CF_{k-1} \geq CF_k$ :  $P_{max}(C_k = 1 | C_{k-1} = 1) = \frac{CF_k}{CF_{k-1}}$

184 As a result,

$$P_{max}(C_k = 1 | C_{k-1} = 1) = \frac{\min(CF_{k-1}, CF_k)}{CF_{k-1}} \quad (6)$$

185 and (4) becomes :

$$P(C_k = 1 | C_{k-1} = 1) = \alpha \frac{\min(CF_{k-1}, CF_k)}{CF_{k-1}} + (1 - \alpha) CF_k \quad (7)$$

186 Let us compute  $P_{max}(C_k = 0 | C_{k-1} = 0)$  in the same way, and we get :

$$P_{max}(C_k = 0 | C_{k-1} = 0) = \frac{1 - \max(CF_{k-1}, CF_k)}{1 - CF_{k-1}}$$

187 and therefore:

$$P(C_k = 0 | C_{k-1} = 0) = \alpha \times \frac{(1 - \max(CF_{k-1}, CF_k))}{1 - CF_{k-1}} + (1 - \alpha)(1 - CF_k) \quad (8)$$

188 These equations and exponential-random overlap more generally are applicable only  
 189 for non overcast cloudy layers (i.e.  $CF \in ]0, 1[$ ). Having computed the transition prob-  
 190 abilities between different cloud states of the cells, we can now use them to generate sub-  
 191 columns. The details of the implementation are presented in Appendix A, along with  
 192 the main difference with the work of Räisänen et al. (2004), from which our algorithm  
 193 is very much inspired. Thanks to Eqs. (7), (8) and (2) we can now compute the differ-  
 194 ent transition probabilities for each layer  $k$ , knowing  $\alpha$ . Then using Eq. (1) we can com-  
 195 pute the probability to generate any vertical cloud fraction distribution for a subcolumn,  
 196 for any exponential-random overlap parameter  $\alpha \in [0, 1]$ .

## 197 **2.2 The relationship between the overlap parameter $\alpha$ and the total cloud** 198 **cover**

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

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

$$P_\emptyset = \prod_{k=1}^{\mathcal{N}} P(C_k = 0 \mid C_{k-1} = 0) \quad (9)$$

208 Using Eq. (8) we get:

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

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

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

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

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

212 where

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

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

216 Eq. (12) gives the expression of  $\alpha$  for a given cloud cover  $CC$  and cloud fraction  
 217 profile  $(CF)$ . Eq. (10) allows us to compute  $CC$  if we know the overlap parameter  $\alpha$  and  
 218 the profile  $(CF)$ . Therefore for any given profile  $(CF)$  and given the ERO model, it is  
 219 equivalent to know  $CC$  or  $\alpha$  (or the decorrelation length, see further).

### 220 2.3 Vertical Subgridding

221 We are now going to use the same method but to define how to generate a sam-  
 222 ple of subcolumns with a higher vertical resolution starting from an atmospheric column  
 223 with a coarse vertical resolution. We start from such a single column of  $N$  coarse lay-  
 224 ers from which we know the vertical volume cloud fraction distribution  $\{\widehat{CF}_k\}_{k=1\dots N}$ ,  
 225 and we generate subcolumns with  $n$  times more vertical levels,  $\mathcal{N} = (N \times n)$ . We in-  
 226 troduce the hypothesis that at every coarse level of the atmospheric column, the volume  
 227 cloud fraction is the same for all the  $n$  sublayers :

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

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

229 We then compute, like done previously, the probability  $P_\emptyset$  to generate a clear-sky  
 230 subcolumn. As the cloud fraction in a single coarse cell is uniform, the intralayer tran-  
 231 sition probability  $P(C_l = 0|C_{l-1} = 0)$  (Eq. (8)) between layers inside the same coarse  
 232 cell simplifies as :

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

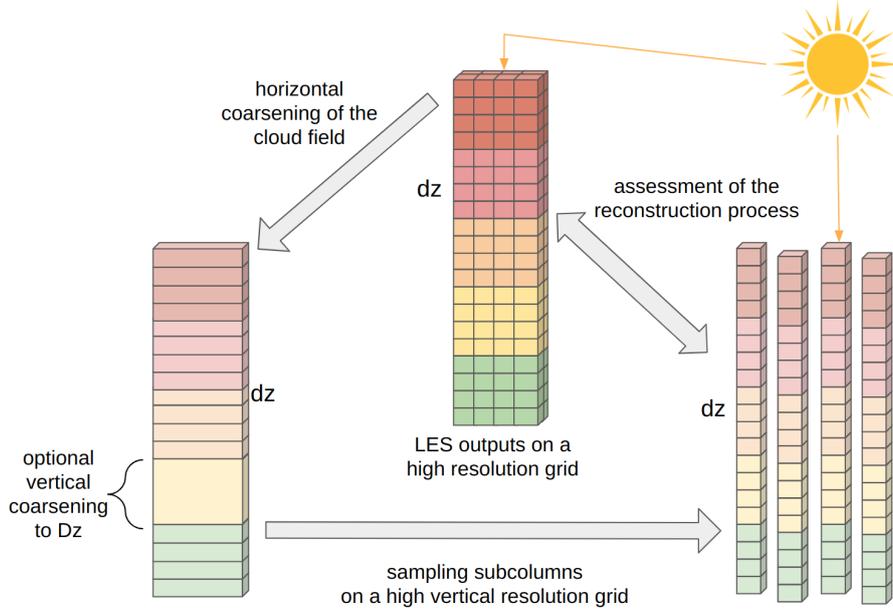
233 For two adjacent cells that belong to two adjacent coarse layers,  $CF_k$  and  $CF_{k-1}$   
 234 can be different and the interlayer overlap transition probability,  $P(C_k = 0|C_{k-1} =$   
 235  $0) = P_{inter,k}$  is given by Eq. (8). Finally,  $P_\emptyset$  is given by:

$$\begin{aligned} P_\emptyset(\alpha, N, n, CF) &= \prod_{k=1}^N \left[ P_{inter,k} \prod_1^{n-1} P_{intra,k} \right] \\ &= \prod_{k=1}^N \left[ [\alpha + (1 - \alpha)(1 - \widehat{CF}_k)]^{n-1} \right] \\ &\times \left[ \frac{\alpha * \left( 1 - \max(\widehat{CF}_{k-1}, \widehat{CF}_k) \right)}{1 - \widehat{CF}_{k-1}} + (1 - \alpha)(1 - \widehat{CF}_k) \right] \end{aligned} \quad (14)$$

236 Like done previously, we can compute the cloud cover generated by a given overlap  
 237 parameter  $\alpha$ , or if the total cloud cover of the scene is known, we can inverse this  
 238 equation using Eq. (12) to compute the overlap parameter  $\alpha$  that generates the same  
 239 cloud cover. The next section shows the results of this subgridding: both its impacts on  
 240 the cloud fraction profiles and the radiative properties of the ERO samples.

### 241 3 Evaluating $\alpha$ and the cloud generation

242 As done in many previous works such as Larson et al. (2002); R. A. J. Neggers et  
 243 al. (2003); Neggers et al. (2011), we are using Large Eddy Simulations (LES) as refer-  
 244 ence cases to assess our ERO algorithm. To test the algorithm presented in the previ-  
 245 ous section, different shallow cumulus cloud cases have been used. We mostly studied  
 246 the ARMCu cloud case (Brown et al. (2002)) showing the development of shallow cum-  
 247 ulus convection over land, as well as two marine, trade-winds cumulus cloud cases BOMEX  
 248 (Siebesma et al. (2003)) and RICO (vanZanten et al. (2011)), and another case of con-  
 249 tinental cumulus SCMS (Neggers et al. (2003b)). For each case we use the correspond-  
 250 ing LES results obtained with the atmospheric non-hydrostatic model MESO-NH (Lafore  
 251 et al. (1998); Lac et al. (2018)), and all these simulations represent a  $6.4 \text{ km} \times 6.4 \text{ km} \times$   
 252  $4 \text{ km}$  domain with a  $dx=dy=dz=25 \text{ m}$  resolution. For each LES simulation we coarsen  
 253 it into a single atmospheric column with the same vertical resolution  $dz$ , or a lower ver-  
 254 tical resolution  $Dz$ , as shown in Fig. 1. For each of these single columns we know, by  
 255 means of the LES, the total cloud cover  $CC$ , as well as the cloud fraction and the liq-  
 256 uid water content at each vertical level. Doing so we go from a highly detailed 3D sim-  
 257 ulation to a single column, and we lose the horizontal cloud structure. Using this sin-  
 258 gle column we then sample subcolumns with the ERO algorithm presented in the previ-  
 259 ous section. Finally, we assess this generation by comparing the statistical properties  
 260 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  $dx=dy=dz=25\text{ m}$  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.

261  
262

### 3.1 Testing ERO and subgridding assuming the overlap parameter has a vertically constant value

263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275

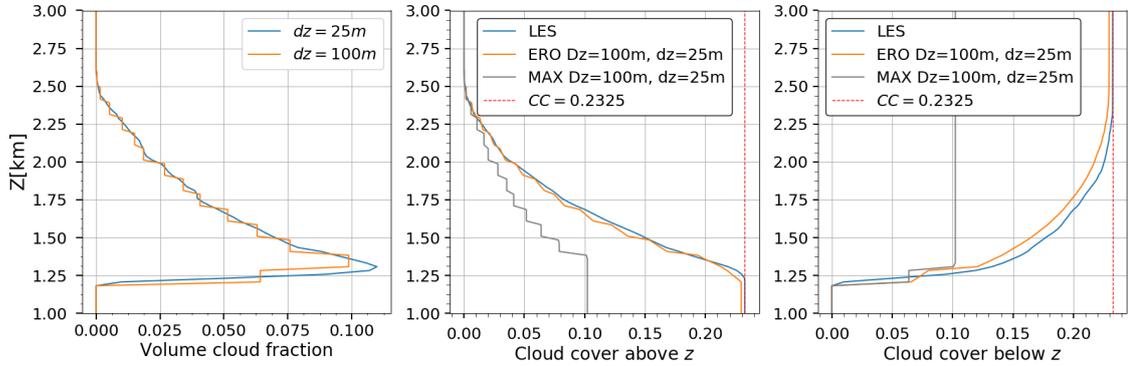
To assess the ERO generation process we first test the assumption that it is sufficient to use a single overlap parameter  $\alpha$  for the whole cloud scene. We use an atmospheric column with a coarser vertical grid than the LES ( $Dz=100\text{ m}$  for the coarse resolution,  $dz=25\text{ m}$  for the LES), and then use subgridding 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).

276  
277  
278  
279  
280  
281  
282  
283  
284  
285

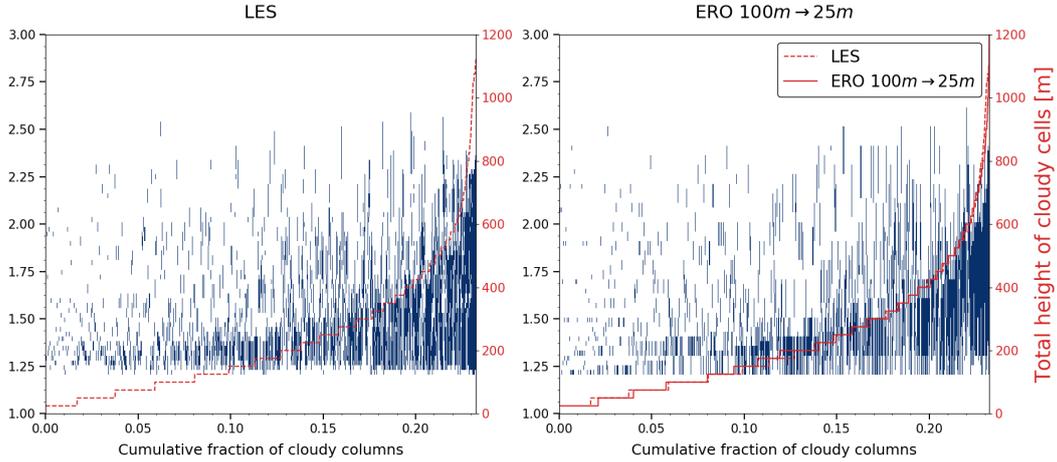
The blue line (Fig. 2, middle and right panels) is the cloud cover profile of the original LES, with a total cloud cover of 0.2325. The grey line is obtained using a maximum overlap assumption, and shows a total cloud cover of only  $\sim 10\%$ . Since the scene consists of a single cloud block, this 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 hypothe-

286  
287  
288

sis of a constant volume cloud fraction  $CF_v$  in each coarse cell. For the generation without vertical subgridding of the previous section ( $Dz=dz=25\text{ m}$ ), 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  $Dz=100\text{ m}$  and have the same final vertical resolution  $dz=25\text{ m}$  than the LES. The data presented is the ARMCu cloud case (time step  $h=10$ ).

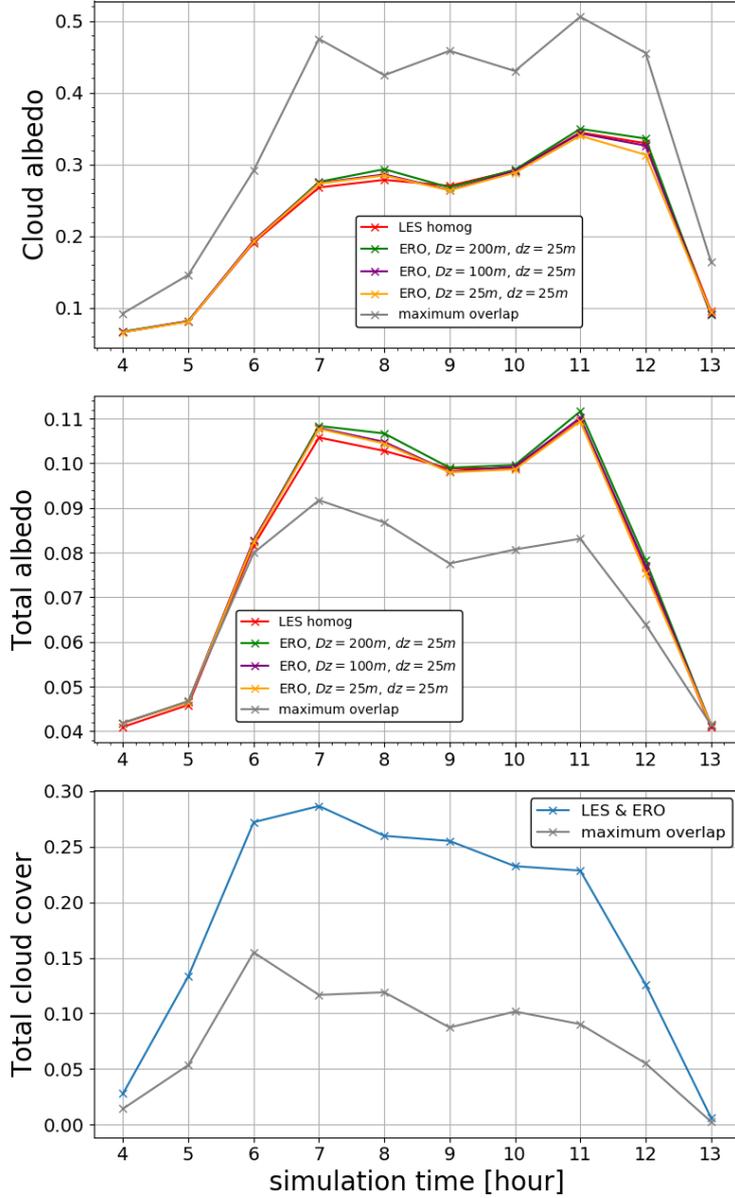


**Figure 3.** The cloudy subcolumns of the LES scene (left) are sorted along the number of cloudy cells in each subcolumns (dashed red). On the right the cloudy subcolumns out of a  $N_s \approx 6.5 \times 10^4$  sample of subcolumns generated with ERO sorted in the same way (solid red for the number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

289 To go further, Fig. 3 shows the cloudy subcolumns of the same scene (cloudy cells  
290 in blue) sorted along the number of cloudy cells in each subcolumn (red). The left panel  
291 shows the cloudy subcolumns of the original LES, and the right panel shows the same  
292 plot for the sample of subcolumns generated by ERO. The vertical distribution of cloudy  
293 cells are very close, it shows the ERO generation not only reproduces the total cloud cover  
294 of the original scene, but also the distribution of cumulative cloud fraction.

295 We then assess the radiative characteristics of the sample by comparing the short-  
296 wave (SW) radiative properties of the LES and that of the ERO sample. We compute  
297 the mean albedo of the cloudy subcolumns (i.e we do not consider any clear sky subcolumns)  
298 for different cloud scenes using a path-tracing Monte Carlo code from Villefranque et al.  
299 (2019). It tracks photon paths throughout a virtual atmosphere, explicitly simulating  
300 the radiative processes such as scattering, absorption, and surface albedo. When a pho-  
301 ton hits the top of the atmosphere (TOA), the algorithm adds its weight to a TOA counter  
302 (for reflection toward space), to a ground counter when it touches the ground (for ground  
303 absorption, here we put the ground albedo at zero), or to an atmospheric counter when  
304 it is absorbed (by liquid water or a gas). As the generated sample has no horizontal struc-  
305 ture, we use the Independant Column Approximation -or ICA - (Pincus et al. (2003)).  
306 Fig. 4 shows the cloud albedo of different sampling hypotheses, of the original LES scenes,  
307 as well as the total albedo of the scenes, and their total cloud cover. For each value of  
308 the coarse resolution  $Dz$ , a new overlap parameter has been computed : the different ERO  
309 scenes hence have the same total cloud covers.

310 The maximum overlap assumption (grey) shows a much higher cloud albedo since  
311 it produces cloud scenes with less total cloud cover and hence brighter clouds. Using ERO  
312 produces a much closer cloud albedo, and the coarse resolution of the initial atmospheric  
313 single column has little impact : the relative difference with the cloud albedo of the ho-  
314 mogeneous LES starting with a 25  $m$  vertical resolution is  $\sim 1.5\%$  and only of  $\sim 2.5\%$   
315 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  $Dz=100\text{ m}$  (in grey). The albedo of each scene is computed using a Monte-Carlo algorithm under the Independent 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  $10^{-6}$ .

316

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

317

318

319

320

321

322

323

324

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)) :

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

325

326

If the decorrelation length  $L_\alpha$  is constant on the vertical (which is generally assumed), it becomes :

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

327

328

329

330

331

332

333

334

335

336

337

338

339

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 dependence 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 subgridding 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.

340

341

342

343

344

345

346

347

348

349

350

351

352

353

354

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 developing 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 literature with similar LES simulations (Neggers et al. (2011); Sulak et al. (2020); Villefranche et al. (2021)). The difference with decorrelation lengths in the literature that take into account the overlap of whole atmospheric columns in global model is further discussed in Section 5.

355

356

357

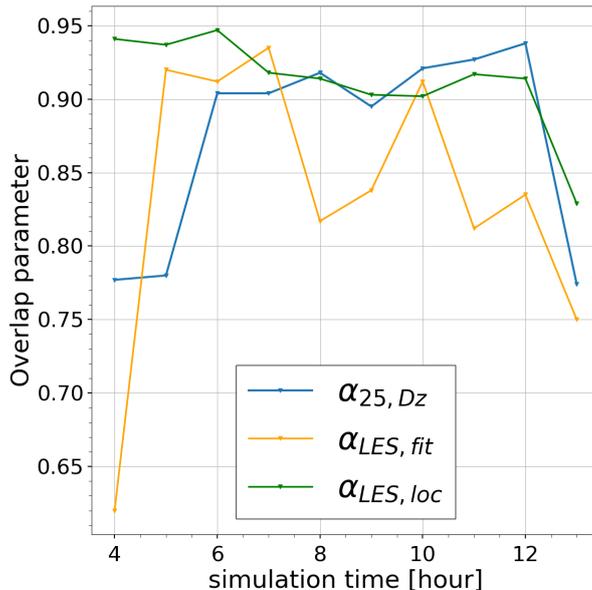
358

359

360

361

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 sub-columns 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 therefore  $CF_v = CF_s$ . 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  $CF_v$ .

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

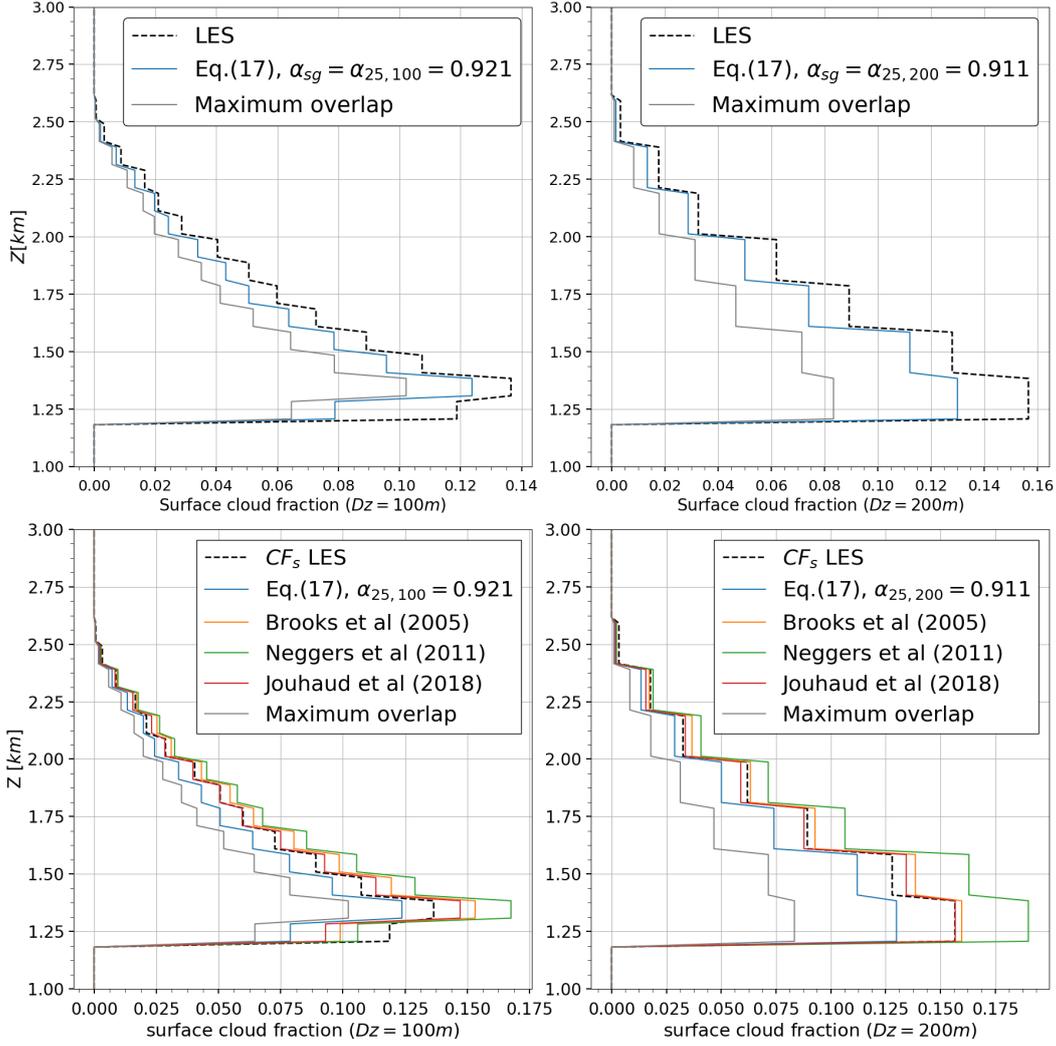
$$CF_{s,sg,k} = 1 - (1 - CF_{v,k})(\alpha_{sg} + (1 - \alpha_{sg})(1 - CF_{v,k}))^{n-1} \quad (17)$$

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

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

411 To go further we also compare the performance of Eq. (17) with that of other ref-  
 412 erences in the litterature. Neggers et al. (2011) and Jouhaud et al. (2018) have both been  
 413 developed using LES data of small cumulus with  $CF_v \approx 0.1$ , including the ARMCu and  
 414 BOMEX cases, and are therefore comparable to our method. Brooks et al. (2005) de-  
 415 velops a lidar and radar-based parametrization of  $CF_s$  using  $CF_v$ , with the possibility  
 416 to take into account wind shear (not used here), and is valid on a wider range of cloud  
 417 covers and situations. Brooks et al. (2005) and Jouhaud et al. (2018) show the small-  
 418 est errors with  $CF_s$  of the LES.

419 Our approach favours an accurate cloud cover on the whole vertical extent of the  
 420 cloud layer. Results show that with this approach we tend to underestimate the surface  
 421 cloud fraction of the coarse layers. This is because the overlap parameter  $\alpha$  has been com-  
 422 puted to match the total cloud cover of the whole scene, not the surface cloud fraction  
 423  $CF_s$  of each coarse layer. When only used for the subgrid scale it creates too small a sur-  
 424 face cloud fraction. This underestimation is still much smaller than when considering  
 425 maximum overlap. The gap in surface cloud fraction caused by using our method is sim-  
 426 ilar to those caused by other approximations of the litterature, but with an opposite  
 427 sign in the difference. Our underestimation of  $(CF_s)_z$  was already visible in Fig. 2 on  
 428 the panel showing “cloud cover above  $z$ ”. The only difference between using subgridding  
 429 or not is the hypothesis  $CF_{vol} = cst$  in each coarse layers, so we can conclude than the  
 430 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 ( $h=10$ ).

431

## 4.2 Interlayer overlap

432

433

434

435

436

437

438

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  $\alpha_{inter}$ .

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

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

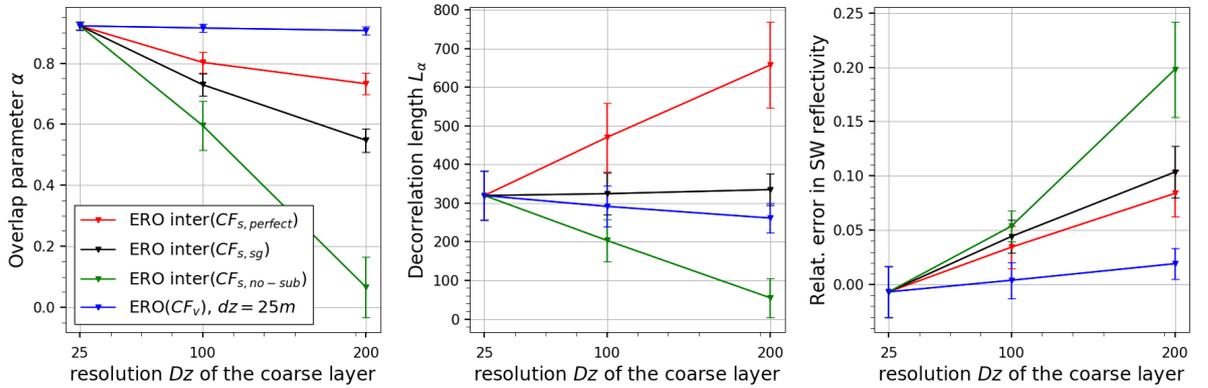
441 The overlap parameter  $\alpha_{inter}$  can be computed as in the previous sections, by in-  
 442 verting Eq. (18) to constrain the cloud cover  $CC$ :

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

### 443 4.3 Generating subcolumns on the coarse grid

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

452 We find that  $\alpha_{25,Dz}$  and the corresponding decorrelation length (Fig. 7, blue plots,  
 453 left and middle panels) depend little on the starting coarse resolution  $Dz$  on this 25–200  
 454  $m$  range, with mean values  $\bar{\alpha}_{25,Dz}=0.915$  and  $\bar{L}_{\alpha,25,Dz}=291 m$ . Using this overlap and  
 455 Eq. (17) we then compute the subgrid profile  $(CF_{s,sg})_z$ , as well as the interlayer over-  
 456 lap 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.

457 We found that the overlap parameter  $\alpha_{inter}$  varies with the resolution  $Dz$  but the  
 458 corresponding decorrelation length varies little from  $\bar{L}_{\alpha,sg}=326 m$  (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  $Dz=25\text{ m}$ , 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  $Dz=200\text{ m}$ , the interlayer overlap without subgridding 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  $dz=25\text{ m}$  (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  $400\text{ m}$  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 literature 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

510 to compute the surface cloud fraction as a bulk subgrid property using the volume cloud  
 511 fraction and a subgrid overlap parameter. Overlapping this computed subgrid cloud frac-  
 512 tion leads to a relative error in cloudy albedo of  $\approx 10\%$  for coarse resolutions of  $100\ m$   
 513 and  $200\ m$  (Fig. 7, black plot). If this subgrid computation were perfect to take into ac-  
 514 count the subgrid scale, it would lead to a slightly improved  $5\text{--}8\%$  relative error in cloud  
 515 albedo for coarse resolutions of  $100\ m$  and  $200\ m$  (Fig. 7, red plot). Finally, even with-  
 516 out taking into account any subgrid scale by overlapping  $(CF_{s,no-sub})_z$  on the coarse  
 517 grid, we can approach the albedo of the LES scenes within a  $20\%$  relative error (for a  
 518 resolution of  $200\ m$ , Fig. 7, green plot) if the total cloud cover is reproduced. As all the  
 519 generations shown in Fig. 7 have the same total cloud cover and mean liquid water path  
 520 as the LES simulations, the difference in cloud albedo are all due to vertical subgrid het-  
 521 erogeneity. If the conservation of the total cloud cover is of first order importance for the  
 522 cloud albedo, the subgrid scale information contained in the cloud fraction profile can  
 523 have a significant impact on the cloud albedo as well, up to  $20\%$ . Numbers in this sec-  
 524 tion are computed on 7 scenes from the ARMCu cloud case, but similar results were also  
 525 found consistently in several other cases, see Figs. S1-S3 in Supporting Information.

## 526 5 Implications

527 In this last section we address some more global implications of our method, es-  
 528 pecially on the use and estimate of the decorrelation lengths, as well as the radiative im-  
 529 pact of LWC horizontal heterogeneity, which had not been taken into account in this pa-  
 530 per until now.

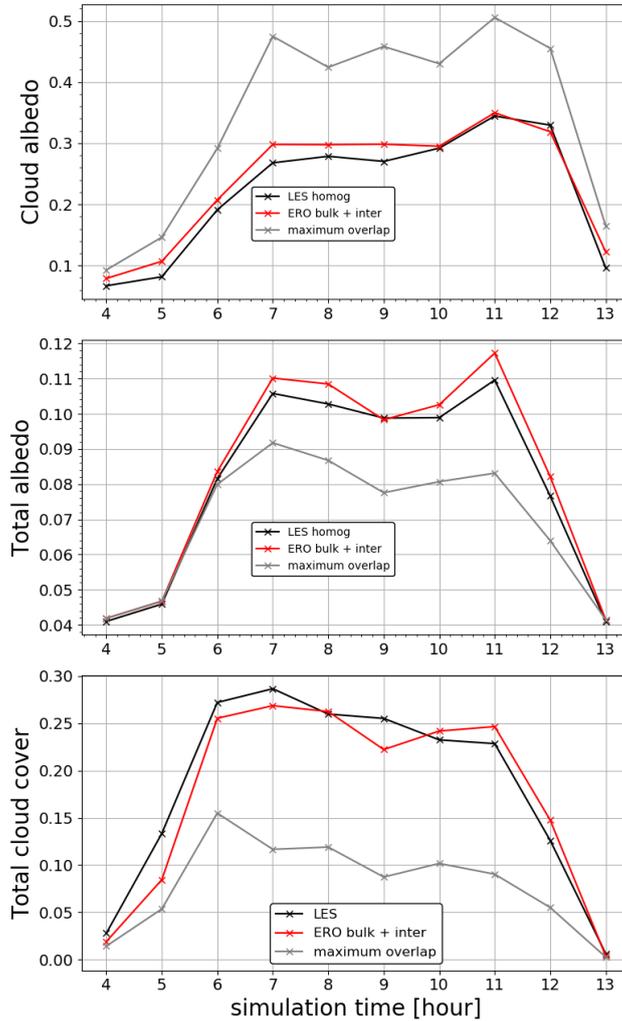
### 531 5.1 How to generate the cloud vertical profile

532 The starting point of the developments in Section 3 and 4 was to determine how  
 533 to correctly represent the cloud cover and the SW cloud albedo of a cloud scene in the  
 534 context of exponential-random overlap. We have shown in Section 3 that by defining the  
 535 appropriate decorrelation length  $L_{\alpha,25,Dz}$  we can generate a cloud scene with the cor-  
 536 rect cloud cover and a close SW cloud albedo. This can be done on a new grid with higher  
 537 vertical resolution ( $25\ m$  here) as long as the initial coarse resolution and the final res-  
 538 olution are both taken into account in the computation of the overlap. This can also be  
 539 done directly on the coarse grid without losing much accuracy on the cloud albedo by  
 540 taking into account both the subgrid scale and the interlayer overlap (section 4.3).

541 So far we have assumed that the cloud cover is known, whereas in general we are  
 542 trying to determine the cloud cover. So we have to reverse the previous problem and ad-  
 543 dress the following question : how to create the right cloud cover and the right cloud albedo  
 544 from the information given by a coarse grid? In this context, an important result of sec-  
 545 tion 4.3 is that if we consistently account for subgrid heterogeneity and coarse layer over-  
 546 lap, then the decorrelation lengths used for the subgrid and the overlap are almost the  
 547 same and they depend weakly on the vertical resolution, as we can see on Fig. 7.

548 The procedure for reconstructing a cloud scene that we propose is as follow: given  
 549 any volume cloud fraction profile  $(CF_v)_z$  at resolution  $Dz$  and the decorrelation length  
 550  $L_\alpha$  for a reference resolution (here  $dz=25\ m$ ), the subgrid heterogeneity is taken into ac-  
 551 count by computing a profile of the surface cloud fraction  $(CF_{s,sg})_z$  with Eq. (17), with  
 552  $n=Dz/dz$  in the equation. The same decorrelation length  $L_\alpha$ , allows to overlap these coarse  
 553 layers and to compute the total cloud cover (Eq. (18)). As we can see on Fig. 7 for the  
 554 case studied here,  $L_{\alpha,25,Dz}\approx 291\ m$  and  $L_{\alpha,sg}\approx 326\ m$ , so for both steps of this reconstruc-  
 555 tion we choose to use the unique decorrelation length that is the mean of the two:  $\bar{L}_\alpha=309$   
 556  $m$ . We find similar results than those shown on Fig. 7 for three other cumulus cloud cases  
 557 simulated by the same LES and the same resolutions, with  $L_{\alpha,25,Dz}$  and  $L_{\alpha,sg}$  relatively  
 558 independent of the resolution. For the RICO case we have  $\bar{L}_\alpha=217\ m$ , for BOMEX  $\bar{L}_\alpha=202$   
 559  $m$  and for SCMS  $\bar{L}_\alpha=273\ m$  (see Figs. S1-S3 in Supporting Information). Here a dif-

560 ferent decorrelation length has been computed for each cloud case. The determination  
 561 of this decorrelation length in a more general case is beyond the scope of this study. As  
 562 it can be seen on Fig. 8, the scenes generated with this method show a good reproduc-  
 563 tion of the cloud cover, cloud albedo and total albedo, with relative errors compared to  
 564 the LES of only  $-10\%$ ,  $11\%$ , and  $-3\%$  respectively, which is significantly better than  
 565 the errors caused by the maximum-random assumption. We also see from this figure that  
 566 the maximum overlap causes a “too few too bright” bias here, with a cloud cover too small  
 567 and a cloud albedo too large. But the two errors do not compensate and the total albedo  
 568 of the scenes is underestimated. Increasing the liquid water content seen in the radiative  
 569 computations to balance the mean radiative flux at TOA could correct the value of  
 570 total albedo but in the same time would also worsen the “too bright” part of the bias.  
 571 Similar results are found for the three other cloud cases and can be found in the Sup-  
 572 porting 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  $L_\alpha=309\text{ m}$ . The scenes are the ARMCu case (time steps  $h\in[4, 13]$ ). In all scenes the LWC is homogeneous at each vertical level.

573  
574

## 5.2 Variations of the decorrelation length with the measurement resolution

575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587

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 entirely 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.

588  
589  
590  
591  
592  
593

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.

594  
595  
596  
597  
598  
599

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.

600  
601  
602  
603  
604  
605  
606  
607

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 length in the case of the vertical resolution of CloudSat at 480 *m*, but an even larger effect is expected.

608  
609  
610  
611  
612  
613  
614  
615  
616

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); Willen 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.

617

## 5.3 Considering LWC distributions

618  
619  
620  
621  
622  
623  
624

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

625 to reproduce those results, and compare the second order impacts of the LWC horizon-  
 626 tal heterogeneity to those of the cloud fraction subgrid vertical heterogeneity shown in  
 627 Section 4.5. To do so we use ERO with vertical subgridding, assuming that the horizon-  
 628 tal distribution of the LWC in each horizontal layer follows the following gamma distri-  
 629 bution, as done in Räisänen et al. (2004) :

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

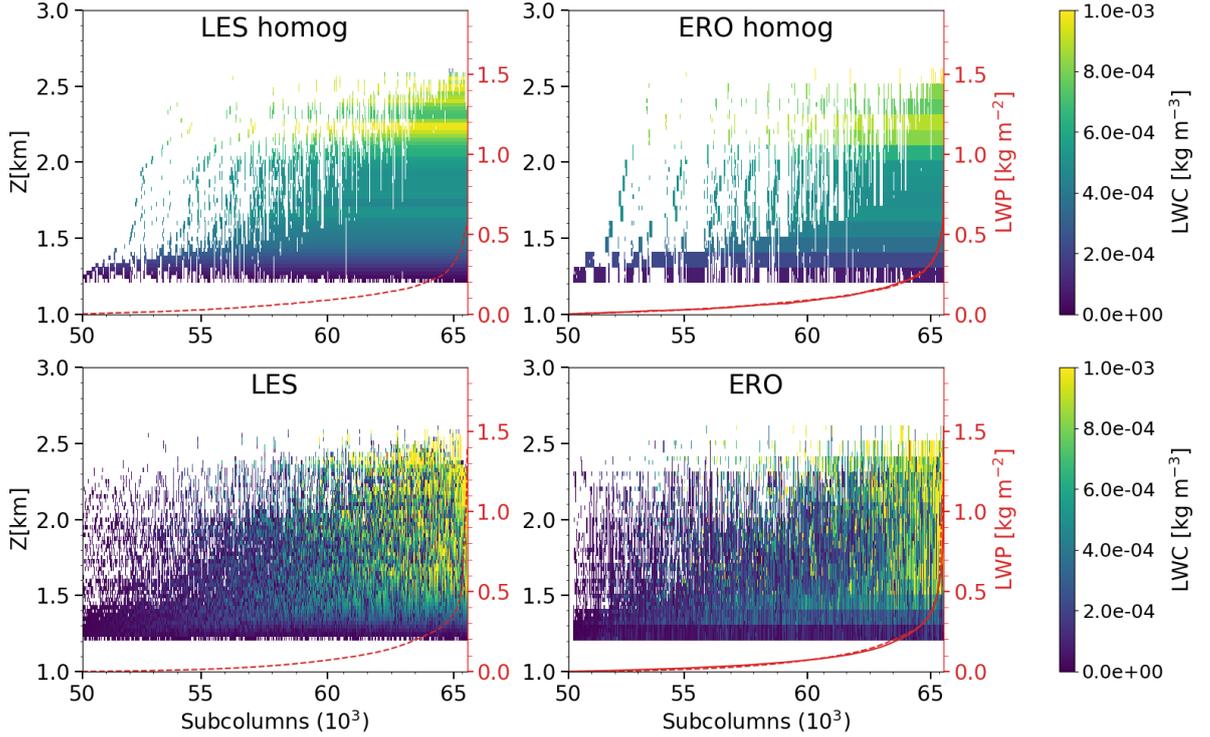
630 where  $x$  is the liquid water content in  $kg/kg$ ,  $k\theta$  is the mean of the distribution and  
 631  $k\theta^2$  its variance,  $\Gamma(k)$  is the gamma function, with  $Re(k) > 0$  :

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

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

637 We generate the cloud field with LWC distributions from an atmospheric column  
 638 ( $Dz=100 m$ ) to a sample of subcolumns with the same vertical resolution as the LES  
 639 ( $dz=25 m$ ), and display on Fig. 9 the LWC of both scenes' cloudy subcolumns after they  
 640 have been sorted along their vertical LWP (bottom panels). The equivalent generation  
 641 with no horizontal heterogeneity of the LWC is shown as a comparison in the top pan-  
 642 els. When using LWC distributions, the generated subcolumns shows the same charac-  
 643 teristics than the LES : a lot of subcolumns with a small LWP, as well as a LWP increas-  
 644 ing with the altitude, and a small number of subcolumns with a high amount of LWP.  
 645 The generated subcolumns shows demarcations every  $100 m$  that are coming from the  
 646 coarse vertical resolution of the atmospheric column because the profile  $(CF_v)_z$  and the  
 647 LWC properties are assumed to be constant in each coarse horizontal layer. The LWC  
 648 heterogeneity also causes more disparity in the LWC values, especially high values, which  
 649 are smoothed out in the homogeneous plots.

650 We then quantify the impact of the LWC horizontal distribution on radiative prop-  
 651 erties. To do so we look at the relative difference of cloud albedo between LES simula-  
 652 tions with the exact LWC heterogeneity and their ERO generations with and without  
 653 LWC heterogeneity. They were generated from the coarse resolution  $Dz=100 m$  to the  
 654 LES vertical resolution  $dz=25 m$  like done in Section 3, for the two cases ARMCu and  
 655 BOMEX. Introducing LWC horizontal distributions significantly improves the cloudy albedo  
 656 : the mean relative difference with that of the LES with exact LWC goes from 8.5% to  
 657 2.4% for ARMCu and from 12.7% to 2% for BOMEX. Comparing the LES with exact  
 658 LWC and their homogeneous versions we find the scenes without LWC horizontal het-  
 659 erogeneity are  $\approx 10\%$  brighter, which confirms the previous findings of Barker et al. (2003),  
 660 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.

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

664 **6 Summary and conclusion**

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

670 Within the McICA framework, we propose an algorithm to generate the cloud fraction  
 671 on a high resolution vertical grid for an ensemble of subcolumns using a single low  
 672 resolution atmospheric column and either the total cloud cover or the overlap param-  
 673 eter. Compared to reference LES simulations, the generated cloud scenes show a correct  
 674 representation of both the distribution of cumulative cloud fraction among cloudy sub-  
 675 columns and the vertical profile of the cloud cover seen from above or below. We sug-  
 676 gest that the later is a simple diagnostic that would usefully complement the usual cloud  
 677 fraction vertical profile when comparing models with observations or when developing  
 678 models. The generated cloudy albedos are very close to the ones of the original LES cloud  
 679 scenes, with only a 2% relative error for the best reconstructions.

680 To avoid having to generate the cloud fraction profile on a high resolution verti-  
681 cal grid, we investigate how to represent both the subgrid variability within coarse lay-  
682 ers and the overlap of these coarse layers to ensure correct values of total cloud cover and  
683 cloud albedo. We demonstrate that, depending on how the subgrid variability is repre-  
684 sented, the decorrelation length used to overlap the coarse layers may be highly depen-  
685 dent on their vertical resolution. However, we show that the subgrid variability and the  
686 interlayer overlap can be defined in such a way to define a decorrelation length almost  
687 independent of the resolution.

688 We also demonstrate that the decorrelation lengths obtained from remote sensing  
689 depend on the vertical resolution of the instruments. For a same cloud scene, the decor-  
690 relation length obtained from an instrument with a vertical resolution of 200 *m* can be  
691 two times larger than the one obtained with an instrument with a vertical resolution of  
692 25 *m*. This may partly explain why the decorrelation lengths obtained by the studies  
693 using CloudSat observations are about 7 times larger than those obtained from high res-  
694 olution models. If the decorrelation length can take into account the distance between  
695 cloudy layers to compute the overlap parameters, the thickness of the layers also has to  
696 be taken into account when estimating decorrelation lengths, as well as whether the cloud  
697 fractions are volumic or surfacic. Although this deserves more investigations, we provide  
698 a framework that allows to go from one vertical resolution to another. Further work is  
699 also required to establish robust estimates of the decorrelation length for a large vari-  
700 ety of clouds.

701 To our best knowledge, most current atmospheric models neglect the effect of sub-  
702 grid variability on the cloud fraction and assume a maximum-random overlap of cloud  
703 layers or a ERO with a quite large decorrelation length ( $\approx 2-3$  *km*). This can lead to  
704 an underestimation of the cloud cover by a factor of two, at least for low-level clouds,  
705 and therefore explain a significant part of the underestimation of these clouds that is iden-  
706 tified in current climate models (Konsta et al. (2022)). A better consideration of sub-  
707 grid heterogeneity and cloud overlap in the models should allow this bias to be reduced,  
708 but would also require a significant revision of the amount of condensed water so that  
709 the global albedo does not change too much. This would contribute to reduce the cur-  
710 rent too few too bright bias.

711 In addition to the effect of the water content heterogeneity on cloud albedo, already  
712 well recognized, we show that the vertical distribution of cloud fraction also matters. In-  
713 deed, for a low-level cloud scene with a given cloud cover and cloud water path, the cloud  
714 albedo can change by about 20% according to how the vertical profile of the clouds frac-  
715 tion is represented. As we focused on the vertical structure of clouds within the plan par-  
716 allel approximation, we have not taken into account the solar angle or 3D radiative ef-  
717 fects. We computed that averaged over a whole day, the relative 3D effects on the SW  
718 cloud albedo are about 7% to 18% for the cases used in this study. Further work would  
719 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:

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

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 :

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

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

$$P_{max}(C_k = 1 | C_{k-1} = 0) = 1 - P_{max}(C_k = 0 | C_{k-1} = 0) = \frac{\max(CF_{k-1}, CF_k)}{1 - CF_{k-1}} \quad (\text{A2})$$

- **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 :

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

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 amount) 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 whether the cell is cloudy or not. The computation to get  $x_k$  is :

$$x_k = \begin{cases} x_{k-1}, & \text{for } RN2_k \leq \alpha_{k-1,k} \\ RN3_k, & \text{for } RN2_k > \alpha_{k-1,k} \end{cases} \quad (\text{A3})$$

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.

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

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

## 757 Acknowledgments

758 Our many thanks go to Céline Cornet and Frédéric Szczap for insightful discussions about  
 759 this work. We acknowledge support from the Agence Nationale de la Recherche (ANR,  
 760 grants MCG-RAD ANR-18-CE46-0012) and the Centre National d'Études Spatiales (CNES,  
 761 project EMC-Sat). A repository containing the scripts for the ERO algorithm presented  
 762 in this paper is available at <https://github.com/raphleb/ERO.git>. The sources described  
 763 in this paper for the radiative computations are available at the websites ([https://www.meso-  
 764 star.com/projects/htrdr/htrdr.html](https://www.meso-star.com/projects/htrdr/htrdr.html) and [https://www.meso-star.com/projects/star-engine/star-  
 765 engine.html](https://www.meso-star.com/projects/star-engine/star-engine.html)).

## 766 References

- 767 Astin, I., & Di Girolamo, L. (2014). Technical note: The horizontal scale depen-  
 768 dence of the cloud overlap parameter. *Atmos. Chem. Phys.*, *14*(18), 9917–9922.  
 769 doi: 10.5194/acp-14-9917-2014
- 770 Barker, H. W. (2008a). Overlap of fractional cloud for radiation calculations in  
 771 GCMs: A global analysis using CloudSat and CALIPSO data. *J. Geophys. Res.-  
 772 Atm.*, *113*(D8). doi: <https://doi.org/10.1029/2007JD009677>
- 773 Barker, H. W. (2008b). Representing cloud overlap with an effective decorrela-  
 774 tion length: An assessment using cloudsat and calipso data. *J. Geophys. Res.-Atm.*,  
 775 *113*(D24). doi: <https://doi.org/10.1029/2008JD010391>
- 776 Barker, H. W., & Räisänen, P. (2005). Radiative sensitivities for cloud structural  
 777 properties that are unresolved by conventional GCMs. *Q. J. R. Meteorol. Soc.*,  
 778 *131*(612), 3103–3122. doi: <https://doi.org/10.1256/qj.04.174>
- 779 Barker, H. W., Stephens, G. L., & Fu, Q. (1999). The sensitivity of domain-  
 780 averaged solar fluxes to assumptions about cloud geometry. *Q. J. R. Meteo-  
 781 rol. Soc.*, *125*(558), 2127–2152. doi: <https://doi.org/10.1002/qj.49712555810>
- 782 Barker, H. W., Stephens, G. L., Partain, P. T., Bergman, J. W., Bonnel, B., Cam-  
 783 pana, K., ... Yang, F. (2003). Assessing 1D Atmospheric Solar Radiative Transfer  
 784 Models: Interpretation and Handling of Unresolved Clouds. *J. Climate*, *16*(16),  
 785 2676 - 2699. doi: 10.1175/1520-0442(2003)016<2676:ADASRT>2.0.CO;2
- 786 Bergman, J. W., & Rasch, P. J. (2002). Parameterizing vertically coherent cloud  
 787 distributions. *J. Atmos. Sci.*, *59*(14), 2165–2182. doi: 10.1175/1520-0469(2002)  
 788 059<2165VCCD>2.0.CO;2
- 789 Bodas-Salcedo, A., Webb, M. J., Bony, S., Chepfer, H., Dufresne, J.-L., Klein,  
 790 S. A., ... John, V. O. (2011). COSP: Satellite simulation software for model  
 791 assessment. *Bull. Am. Meteorol. Soc.*, *92*(8), 1023 - 1043. doi: 10.1175/  
 792 2011BAMS2856.1
- 793 Brooks, M. E., Hogan, R. J., & Illingworth, A. J. (2005). Parameterizing the dif-  
 794 ference in cloud fraction defined by area and by volume as observed with radar and  
 795 lidar. *J. Atmos. Sci.*, *62*(7), 2248–2260. doi: 10.1175/JAS3467.1
- 796 Brown, A. R., Cederwall, R. T., Chlond, A., Duynkerke, P. G., Golaz, J.-C.,  
 797 Khairoutdinov, M., ... Stevens, B. (2002). Large-eddy simulation of the diurnal

- 798 cycle of shallow cumulus convection over land. *Q. J. R. Meteorol. Soc.*, 128(582),  
799 1075-1093. doi: <https://doi.org/10.1256/003590002320373210>
- 800 Di Giuseppe, F., & Tompkins, A. M. (2015). Generalizing cloud overlap treat-  
801 ment to include the effect of wind shear. *J. Atmos. Sci.*, 72(8), 2865 - 2876. doi: 10  
802 .1175/JAS-D-14-0277.1
- 803 Geleyn, J., & Hollingsworth, A. (1979). An economical analytical method for the  
804 computation of the interaction between scattering and line absorption of radiation.  
805 *Beitr. Phys. Atmosph.*
- 806 Genio, A. D. D., Yao, M.-S., Kovari, W., & Lo, K. K.-W. (1996). A prognos-  
807 tic cloud water parameterization for global climate models. *J. Climate*, 9(2), 270 -  
808 304. doi: 10.1175/1520-0442(1996)009<0270:APCWPF>2.0.CO;2
- 809 Hogan, R. J., & Illingworth, A. J. (2000). Deriving cloud overlap statistics from  
810 radar. *Q. J. R. Meteorol. Soc.*, 126(569), 2903-2909. doi: [https://doi.org/10.1002/](https://doi.org/10.1002/qj.49712656914)  
811 [qj.49712656914](https://doi.org/10.1002/qj.49712656914)
- 812 Hogan, R. J., & Shonk, J. K. P. (2013). Incorporating the effects of 3d radiative  
813 transfer in the presence of clouds into two-stream multilayer radiation schemes.  
814 *J. Atmos. Sci.*, 70(2), 708 - 724. doi: 10.1175/JAS-D-12-041.1
- 815 Jakob, C., & Klein, S. A. (1999). The role of vertically varying cloud fraction in  
816 the parametrization of microphysical processes in the ecmwf model. *Q. J. R. Mete-*  
817 *orol. Soc.*, 125(555), 941-965. doi: 10.1002/qj.49712555510
- 818 Jing, X., Zhang, H., Peng, J., Li, J., & Barker, H. W. (2016). Cloud overlap-  
819 ping parameter obtained from CloudSat/CALIPSO dataset and its applica-  
820 tion in AGCM with McICA scheme. *Atmospheric Research*, 170, 52-65. doi:  
821 <https://doi.org/10.1016/j.atmosres.2015.11.007>
- 822 Jouhaud, J., Dufresne, J.-L., Madeleine, J.-B., Hourdin, F., Couvreux, F., Ville-  
823 franque, N., & Jam, A. (2018). Accounting for vertical subgrid-scale heterogeneity  
824 in low-level cloud fraction parameterizations. *J. Adv. Model. Earth Syst.*, 10(11),  
825 2686-2705. doi: <https://doi.org/10.1029/2018MS001379>
- 826 Konsta, D., Dufresne, J.-L., Chepfer, H., Vial, J., Koshiro, T., Kawai, H., ...  
827 Ogura, T. (2022). Low-level marine tropical clouds in six cmip6 models are too  
828 few, too bright but also too compact and too homogeneous. *Geophys. Res. Lett.*,  
829 49(11), e2021GL097593. doi: <https://doi.org/10.1029/2021GL097593>
- 830 Koren, I., Oreopoulos, L., Feingold, G., Remer, L. A., & Altaratz, O. (2008). How  
831 small is a small cloud? *Atmos. Chem. Phys.*, 8(14), 3855-3864. doi: 10.5194/acp-8  
832 -3855-2008
- 833 Lac, C., Chaboureaud, J.-P., Masson, V., Pinty, J.-P., Tulet, P., Escobar, J., ...  
834 Wautelet, P. (2018). Overview of the Meso-NH model version 5.4 and its applica-  
835 tions. *Geosci. Model Dev.*, 11(5), 1929-1969. doi: 10.5194/gmd-11-1929-2018
- 836 Lafore, J. P., Stein, J., Asencio, N., Bougeault, P., Ducrocq, V., Duron, J., ...  
837 Vilà-Guerau de Arellano, J. (1998). The Meso-NH Atmospheric Simulation Sys-  
838 tem. Part I: adiabatic formulation and control simulations. *Annales Geophysicae*,  
839 16(1), 90-109. doi: 10.1007/s00585-997-0090-6
- 840 Larson, V. E., Golaz, J.-C., & Cotton, W. R. (2002). Small-scale and mesoscale  
841 variability in cloudy boundary layers: Joint probability density functions.  
842 *J. Atmos. Sci.*, 59(24), 3519 - 3539. doi: 10.1175/1520-0469(2002)059<3519:  
843 SSAMVI>2.0.CO;2
- 844 Mace, G. G., & Benson-Troth, S. (2002). Cloud-layer overlap characteristics de-  
845 rived from long-term cloud radar data. *J. Climate*, 15(17), 2505 - 2515. doi: 10  
846 .1175/1520-0442(2002)015<2505:CLOCDF>2.0.CO;2
- 847 Naud, C. M., Genio, A. D., Mace, G. G., Benson, S., Clothiaux, E. E., & Kollias,  
848 P. (2008). Impact of dynamics and atmospheric state on cloud vertical overlap.  
849 *J. Climate*, 21(8), 1758 - 1770. doi: 10.1175/2007JCLI1828.1
- 850 Neggers, Duynkerke, P. G., & Rodts, S. M. A. (2003b). Shallow cumulus convec-  
851 tion: A validation of large-eddy simulation against aircraft and landsat observa-

- 852 tions. *Q. J. R. Meteorol. Soc.*, *129*(593), 2671-2696. doi: [https://doi.org/10.1256/](https://doi.org/10.1256/qj.02.93)  
853 [qj.02.93](https://doi.org/10.1256/qj.02.93)
- 854 Neggers, Heus, T., & Siebesma, A. P. (2011). Overlap statistics of cumuliform  
855 boundary-layer cloud fields in large-eddy simulations. *J. Geophys. Res.-Atm.*,  
856 *116*(D21). doi: <https://doi.org/10.1029/2011JD015650>
- 857 Neggers, R. A. J., Jonker, H. J. J., & Siebesma, A. P. (2003). Size statistics of  
858 cumulus cloud populations in large-eddy simulations. *J. Atmos. Sci.*, *60*(8), 1060 -  
859 1074. doi: [10.1175/1520-0469\(2003\)60<1060:SSOCCP>2.0.CO;2](https://doi.org/10.1175/1520-0469(2003)60<1060:SSOCCP>2.0.CO;2)
- 860 Oreopoulos, L., Lee, D., Sud, Y. C., & Suarez, M. J. (2012). Radiative impacts  
861 of cloud heterogeneity and overlap in an atmospheric general circulation model. *At-*  
862 *mos. Chem. Phys.*, *12*(19), 9097–9111. doi: [10.5194/acp-12-9097-2012](https://doi.org/10.5194/acp-12-9097-2012)
- 863 Oreopoulos, L., & Norris, P. M. (2011). An analysis of cloud overlap at a midlati-  
864 tude atmospheric observation facility. *Atmos. Chem. Phys.*, *11*(12), 5557–5567. doi:  
865 [10.5194/acp-11-5557-2011](https://doi.org/10.5194/acp-11-5557-2011)
- 866 Pincus, R., Barker, H. W., & Morcrette, J.-J. (2003). A fast, flexible, approximate  
867 technique for computing radiative transfer in inhomogeneous cloud fields. *J. Geo-*  
868 *phys. Res.-Atm.*, *108*(D13). doi: <https://doi.org/10.1029/2002JD003322>
- 869 Pincus, R., Hannay, C., Klein, S. A., Xu, K.-M., & Hemler, R. (2005). Overlap  
870 assumptions for assumed probability distribution function cloud schemes in large-  
871 scale models. *J. Geophys. Res.-Atm.*, *110*(D15). doi: [https://doi.org/10.1029/](https://doi.org/10.1029/2004JD005100)  
872 [2004JD005100](https://doi.org/10.1029/2004JD005100)
- 873 Räisänen, P., Barker, H. W., Khairoutdinov, M. F., Li, J., & Randall, D. A.  
874 (2004). Stochastic generation of subgrid-scale cloudy columns for large-scale mod-  
875 els. *Q. J. R. Meteorol. Soc.*, *130*(601), 2047-2067. doi: [https://doi.org/10.1256/](https://doi.org/10.1256/qj.03.99)  
876 [qj.03.99](https://doi.org/10.1256/qj.03.99)
- 877 Shonk, J. K. P., & Hogan, R. J. (2008). Tripleclouds: An efficient method  
878 for representing horizontal cloud inhomogeneity in 1d radiation schemes by  
879 using three regions at each height. *J. Climate*, *21*(11), 2352 - 2370. doi:  
880 [10.1175/2007JCLI1940.1](https://doi.org/10.1175/2007JCLI1940.1)
- 881 Shonk, J. K. P., & Hogan, R. J. (2010). Effect of improving representation of  
882 horizontal and vertical cloud structure on the earth's global radiation budget.  
883 part ii: The global effects. *Q. J. R. Meteorol. Soc.*, *136*(650), 1205-1215. doi:  
884 <https://doi.org/10.1002/qj.646>
- 885 Siebesma, A. P., Bretherton, C. S., Brown, A., Chlond, A., Cuxart, J., Duynkerke,  
886 P. G., ... Stevens, D. E. (2003). A large eddy simulation intercomparison  
887 study of shallow cumulus convection. *J. Atmos. Sci.*, *60*(10), 1201 - 1219. doi:  
888 [10.1175/1520-0469\(2003\)60<1201:ALESIS>2.0.CO;2](https://doi.org/10.1175/1520-0469(2003)60<1201:ALESIS>2.0.CO;2)
- 889 Sulak, A. M., Calabrese, W. J., Ryan, S. D., & Heus, T. (2020). The Contributions  
890 of Shear and Turbulence to Cloud Overlap for Cumulus Clouds. *J. Geophys. Res.-*  
891 *Atm.*, *125*(10), e2019JD032017. doi: <https://doi.org/10.1029/2019JD032017>
- 892 Swales, D. J., Pincus, R., & Bodas-Salcedo, A. (2018). The cloud feedback  
893 model intercomparison project observational simulator package: Version 2.  
894 *Geosci. Model Dev.*, *11*(1), 77–81. doi: [10.5194/gmd-11-77-2018](https://doi.org/10.5194/gmd-11-77-2018)
- 895 Tompkins, A. M., & Di Giuseppe, F. (2007). Generalizing Cloud Overlap Treat-  
896 ment to Include Solar Zenith Angle Effects on Cloud Geometry. *J. Atmos. Sci.*,  
897 *64*(6), 2116-2125. doi: [10.1175/JAS3925.1](https://doi.org/10.1175/JAS3925.1)
- 898 Tompkins, A. M., & Di Giuseppe, F. (2015). An interpretation of cloud overlap  
899 statistics. *J. Atmos. Sci.*, *72*(8), 2877 - 2889. doi: [10.1175/JAS-D-14-0278.1](https://doi.org/10.1175/JAS-D-14-0278.1)
- 900 vanZanten, M. C., Stevens, B., Nuijens, L., Siebesma, A. P., Ackerman, A. S., Bur-  
901 net, F., ... Wyszogrodzki, A. (2011). Controls on precipitation and cloudiness in  
902 simulations of trade-wind cumulus as observed during RICO. *J. Adv. Model. Earth*  
903 *Syst.*, *3*(2). doi: <https://doi.org/10.1029/2011MS000056>
- 904 Villefranque, N., Blanco, S., Couvreux, F., Fournier, R., Gautrais, J., Hogan,  
905 R. J., ... Williamson, D. (2021). Process-based climate model development har-

906 nessing machine learning: III. the representation of cumulus geometry and their  
907 3D radiative effects. *J. Adv. Model. Earth Syst.*, 13(4), e2020MS002423. doi:  
908 <https://doi.org/10.1029/2020MS002423>  
909 Villefranque, N., Fournier, R., Couvreur, F., Blanco, S., Cornet, C., Eymet, V.,  
910 ... Tregan, J.-M. (2019). A Path-Tracing Monte Carlo Library for 3-D Radiative  
911 Transfer in Highly Resolved Cloudy Atmospheres. *J. Adv. Model. Earth Syst.*,  
912 11(8), 2449-2473. doi: <https://doi.org/10.1029/2018MS001602>  
913 Willèn, U., Crewell, S., Baltink, H. K., & Sievers, O. (2005). Assessing model  
914 predicted vertical cloud structure and cloud overlap with radar and lidar ceilome-  
915 ter observations for the baltex bridge campaign of cliwa-net. *Atmospheric Re-*  
916 *search*, 75(3), 227-255. (CLIWA-NET: Observation and Modelling of Liquid Water  
917 Clouds) doi: <https://doi.org/10.1016/j.atmosres.2004.12.008>  
918 Wu, X., & Liang, X.-Z. (2005). Radiative effects of cloud horizontal inhomogeneity  
919 and vertical overlap identified from a monthlong cloud-resolving model simulation.  
920 *J. Atmos. Sci.*, 62(11), 4105 - 4112. doi: 10.1175/JAS3565.1