Boundary-Layer Meteorol (2011) 140:263–294
DOI 10.1007/s10546-011-9607-3

ARTICLE

Evaluation of Statistical Distributions for the Parametrization of Subgrid Boundary-Layer Clouds

Emilie Perraud · Fleur Couvreux · Sylvie Malardel · Christine Lac · Valéry Masson · Odile Thouron

Received: 6 April 2010 / Accepted: 14 March 2011 / Published online: 30 April 2011

Abstract In numerical weather prediction and climate models, planetary boundary-layer (PBL) clouds are linked to subgrid-scale processes such as shallow convection. A comprehensive statistical analysis of large-eddy simulations (LES), obtained for warm PBL cloud cases, is carried out in order to characterize the distributions of the horizontal subgrid cloud variability. The production of subgrid clouds is mainly associated with the variability of the total water content. Nevertheless, in the case of PBL clouds, the temperature variability cannot be completely discarded and the saturation deficit, which summarizes both temperature and total water fluctuations, provides a better representation of the cloud variability than the total water content. The probability density functions (PDFs) of LES saturation deficit generally have the shape of a main asymmetric bell-shaped curve with a more or less distinct secondary maximum specific to each type of PBL clouds. Unimodal theoretical PDFs, even those with a flexible skewness, are not sufficient to correctly fit the LES distributions, especially the long tail that appears for cumulus clouds. They do not provide a unified approach for all cloud types. The cloud fraction and the mean cloud water content, diagnosed from these unimodal PDFs, are largely underestimated. The use of a double Gaussian distribution allows correction of these errors on cloud fields and provides a better estimation of the cloud-base and cloud-top heights. Eventually, insights for the design of a subgrid statistical cloud scheme are provided, in particular a new formulation for the weight of the two Gaussian distributions and for the standard deviation of the convective distribution.

E. Perraud (✉) · F. Couvreux · S. Malardel · C. Lac · V. Masson
CNRM-GAME, Météo-France and CNRS, 42 Avenue Gaspard Coriolis, 31057 Toulouse Cedex 1, France
e-mail: emilie.perraud@cnrm.meteo.fr

F. Couvreux
e-mail: fleur.couvreux@meteo.fr

O. Thouron
CERFACS, 42 Avenue Gaspard Coriolis, 31057 Toulouse Cedex 1, France

Keywords Double Gaussian distribution · Large-eddy simulations · Planetary boundary-layer clouds · Probability density function · Saturation deficit · Statistical cloud scheme

1 Introduction

Numerical weather prediction (NWP) models with resolutions of a few kilometres are now running operationally in a number of Meteorological Centres. Some of them are actually running without deep convection parametrization. However, in these models, planetary boundary-layer (PBL) clouds, e.g. sparse cumuli, are still linked to subgrid-scale processes. Shallow convection needs to be parametrized. Of course, this also applies to climate models. In both model types, the description of the PBL clouds is still not accurate suggesting a need for improving their representation. The challenge to correctly represent cloud life cycles and transitions between PBL cloud regimes is an additional motivation.

A method widely used in the GEWEX (for Global Energy and Water Cycle Experiment) Cloud System Study (GCSS) community for the development of cloud schemes starts with the statistical analysis of fine scale cloud data. Observational data resulting from in situ aircraft measurements or from satellite data (A-Train, Stephens et al. 2002) are collected. However, they usually do not give a broad cover of the cloud organization at fine enough resolution. Three-dimensional cloud simulations performed by large-eddy simulations (LES) provide complete cloud information with high resolution and regular sampling of PBL clouds. PBL case studies such as BOMEX (for Barbados Oceanographic Meteorological EXperiment), ARM/Cumulus (for Atmospheric Radiation Measurement) for shallow cumulus development or ACE-2 (for Second Aerosol Characterization Experiment) and DYCOMS (for DYnamics and Chemistry Of Marine Stratocumulus) for stratocumulus clouds are now well documented. They are already widely used as a basis for the validation of cloud parametrization. This study is based on such simulations. These cases are benchmark simulations. Even though they cannot cover all the PBL cloud types, in particular, more complex cases such as cumulus under stratocumulus or rapidly evolving cloud types, they form a first sample of cases to evaluate the variability of total water content in cloudy boundary layers.

Sommeria and Deardorff (1977) and Mellor (1977) illustrated the necessity for a statistical approach to describe the subgrid warm PBL clouds even with a relatively fine horizontal resolution. The computation of the mean cloud parameters is based on a unimodal Gaussian distribution and uses the conservative variables, total water content and liquid potential temperature, as predictors needed to compute the saturation deficit. Bougeault (1981) highlighted the interest of a skewed distribution and pointed out the drawbacks of the simple Gaussian distribution in case of low cloud fractions. For shallow cumulus clouds, the positively skewed distribution presents a long flat tail, closely related to shallow convection effects. In Bougeault (1982), for cumulus cases with a positive skewness, the unimodal and flexible gamma distribution is used. For quasi-resolved stratocumulus cases with a zero or negative skewness, the simple Gaussian distribution is retained. To adapt to the various cloud regimes, Cuijpers and Bechtold (1995) considered a unimodal distribution computed as a linear combination of a Gaussian distribution (for stratiform clouds) and an exponential distribution (for cumulus clouds). Chaboureau and Bechtold (2002, 2005) considered the effects of deep convection in the parametrization of the standard deviation of the saturation deficit. The combination of these three last studies was an attempt to take into account the variety of cloud regimes in NWP models. However, such a combination of distributions still frequently underestimates the cloud fields, especially in the case of sparse subgrid clouds.

In the case of resolved clouds, subgrid cloud schemes are based on a current binary scheme (Dirac distribution) considering only two values for the cloud fraction, 0 or 100%. This very simple assumption is the basis of the "All or Nothing" method, which is mainly used for cloud-resolving model simulations when the clouds are resolved at fine grid scales.

For larger scale models, simpler distributions are proposed and often used in NWP models. For example, Smith (1990) developed a statistical scheme based on a triangular distribution with the relative humidity as the predictor instead of the conservative variables. Other authors have proposed more sophisticated methods for global circulation models (GCMs), such as the Bony and Emanuel (2001) and Tompkins (2002) schemes based, respectively, on the unimodal log-normal distribution and the unimodal beta distribution coupled with a deep convection scheme. In both, the temperature variability is neglected and the total water content is the only predictor. The scheme of Tompkins (2002) takes into account, in particular, the effects of deep convection on the cloud life cycle by considering prognostic evolutions for the standard deviation and the skewness. It gives a better representation of the cloud evolution in time than do diagnostic statistical schemes for which the cloud formation is diagnosed at each timestep.

Some of the previous studies showed that a non-zero skewness of the total water content or the saturation deficit distribution is very often associated with the appearance of a second mode. Such a mode, often located on the cloudy side of the distribution, is not reproduced with the assumption of simple modality. To represent this second mode, several studies used a bimodal distribution inducing two local maxima. Lappen (1999) considered a double delta distribution equal to a linear combination of two Dirac delta functions, one for the updraft and one for the downdraft. The main drawback of such a scheme is to not allow subplume variability, leading to large errors in the cloud field estimations. Lewellen and Yoh (1993) established a subgrid scheme based on a joint double Gaussian distribution for the total water content, the liquid potential temperature and the vertical velocity. They considered (i) a bimodal distribution, and (ii) a supplementary predictor, the vertical velocity. This double Gaussian distribution is equal to a linear combination of two simple Gaussian distributions. The bimodal distribution relies on the first three moments of these three prognostic variables and assumes the same weight coefficient to compute the double Gaussian for all three variables. It was subsequently used by Golaz et al. (2002a,b) and Larson et al. (2002). The bimodality increases the number of free parameters to be computed, leading to a sophistication of the scheme. However, it also gives a gain in generality for the representation of cloud evolutions and transitions. Lewellen and Yoh (1993) showed that it is necessary to use a distribution with two modes, such as the double Gaussian distribution, and proposed assumptions to reduce the number of parameters without degrading the results by considering the same weight coefficient for all three double Gaussian distributions. Larson et al. (2002) also proposed different assumptions such as assuming a zero skewness of the liquid potential temperature and a skewness of the total water proportional to the skewness of the vertical velocity. In Golaz et al. (2002a), in addition to the previous assumption, the variances of the vertical velocity for the two Gaussian distributions are assumed equal to a fraction of the total variance. It is noted that these distributions rely on the first three moments that should be provided by a high-order turbulence scheme. For PBL convection cases, Neggers (2009) used a double Gaussian distribution for conservative variables, assigning each simple Gaussian probability density function (PDF) to a transport component of an eddy-diffusivity mass-flux scheme: the main mode is assigned to the eddy-diffusivity part and the second mode to the mass-flux part. The closure of its scheme relies on the determination of the total water variance and its updraft variance. This scheme has been validated for several PBL cloud cases, including a complex scenario of cloud transition between stratocumulus and cumulus.

In this study, we revisit the possibility of using most of the different distributions proposed in the literature for a statistical description of PBL clouds. Hence, this study is largely inspired by that of Larson et al. (2002). Nevertheless, instead of relying on the first three moments of three variables, we seek a PDF based only on the liquid potential temperature and the total water mixing ratio, therefore focusing on two variables. This reduces significantly the number of free parameters of the distribution. In addition, no ad-hoc assumptions on the values of the variances and of the skewness are made. Eventually, some guidance for the development of a cloud scheme are deduced from this study. The proposition is to determine the different parameters of the PDF based on information from a turbulence scheme and a shallow convection mass-flux scheme. This ensures coherences between the thermodynamical tendencies provided by those schemes and the cloud characteristics. A brief description of the distributions selected to fit the LES distributions is given in Sect. 2.3. We selected four cases of warm PBL cloud cases obtained with the Meso-NH model (Lafore et al. 1998) with a resolution of 50–100 m. These simulations were validated with the results of GCSS intercomparisons for the shallow cumulus cases, and the stratocumulus cases were validated in Chosson et al. (2007). These cases are presented in Sect. 2.1. Section 3 contains the results of the statistical analysis of the LES distributions, and also presents a study of sensitivity to the horizontal domain size to confirm the robustness of the results of the statistical analysis for 1–3 km resolutions. In Sect. 4, the different theoretical distributions are evaluated with respect to their ability to fit the LES distributions and to give the correct cloud fraction and cloud mean water content. This study has the advantage of being simpler than most previous study on bimodal distributions because it only considers a single statistical variable, the saturation deficit, which facilitates its use in NWP models.

2 Statistical Definitions and Data

2.1 Large-Eddy Simulations

Four cases were run with the LES version of the non-hydrostatic anelastic research model Meso-NH (Lafore et al. 1998) with a horizontal resolution varying from 50 to 100 m depending on cloud cases (see Table 1). The turbulent scheme was a three-dimensional (3D) turbulent kinetic energy (TKE) scheme (Cuxart et al. 2000) with a Deardorff mixing length. PBL clouds were assumed to be resolved at the LES resolution and we used an "All or Nothing" method for a one- or two- moment warm cloud microphysical scheme. The four warm PBL cloud cases were classic intercomparison cases that we selected for the variety of their cloud formations and time evolutions. Ice clouds are not considered.

1. ARM case: diurnal cycle of shallow cumulus over land. An idealized shallow cumulus convection case over land was derived from the Atmospheric Radiation Measurement (ARM) program carried out on 21 June 1997 in the Southern Great Plains. Details of this case are given in Brown et al. (2002).
2. BOMEX case: shallow cumulus over ocean. The BOMEX case study is a well-documented idealized stationary case of shallow cumulus convection over ocean derived from the Barbados Oceanographic Meteorological Experiment (BOMEX) which took place during 22–30 June 1969 (see Holland and Rasmusson 1973). Undisturbed trade wind cumulus convection under steady-state conditions is considered. Details of the case are given in Siebesma et al. (2003).
   For those two cases, an intercomparison of LES models was carried out. The MESO-NH simulations have been compared to the intercomparison results and lie in the range of

Table 1 Configuration of LES simulations

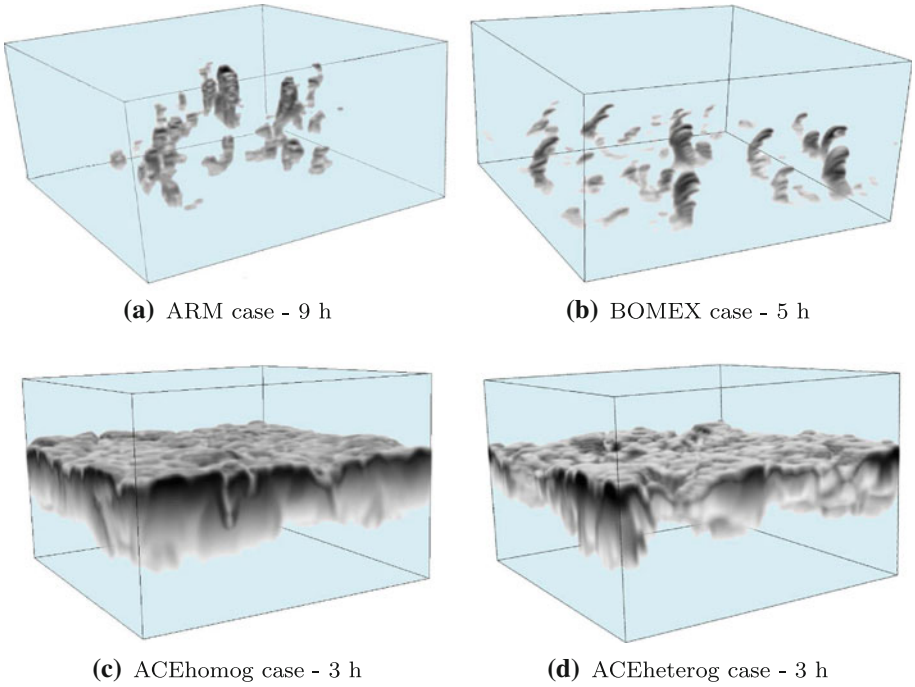
	Cu		Homogeneous Sc	Fractional Sc
	BOMEX	ARM	ACEhomog	ACEheterog
Length of the simulation	6 h	15 h	3 h	3 h
Δt	1 s	1 s	1 s	1 s
Horizontal Domain	$N_x = N_y = 128$	$N_x = N_y = 64$	$N_x = N_y = 100$	$N_x = N_y = 100$
$\Delta x = \Delta y$	50 m	100 m	50 m	50 m
Vertical levels	$N_z = 75$	$N_z = 100$	$N_z = 66$	$N_z = 66$
Δz in cloud	40 m	40 m	10 m	10 m
Δz otherwise	40 m	40 m	30 m	30 m
Domain size ($x \times y \times z$)	$6.4 \times 6.4 \times 3\text{km}^3$	$6.4 \times 6.4 \times 4\text{km}^3$	$5 \times 5 \times 1.5\text{km}^3$	$5 \times 5 \times 1.5\text{km}^3$

the different model results. It is noted that the dispersion among the different models of the intercomparison is relatively small (see Siebesma et al. 2003, and Brown et al. 2002). Sensitivity tests to the resolution show that the results are unchanged when considering a 50-m or a 100-m horizontal resolution.

3. ACEheterog case: heterogeneous stratocumulus. The ACEheterog case is a fractional stratocumulus case, based on the Second Aerosol Characterization Experiment (ACE-2) program which took place during 16–24 July 1997 on the Canary Islands and which had the objective of improving the description of the role of aerosols in GCMs (see Raes et al. 2000). This simulation has been validated against observations in Chosson et al. (2007).
4. ACEhomog case: homogeneous stratocumulus. This stationary case, called ACEhomog, is derived from the reference case ACEheterog for which the total water content in the free troposphere is modified. The initial value is increased from 0.005 to 0.007 kg kg^{-1} at the beginning of the simulation to humidify the free troposphere and to obtain a stationary stratocumulus that does not disappear during the simulation.

Figure 1 shows instantaneous 3D views of the cloud water content for the four cases used for describing qualitatively the various cloud layers. It is noted that, in this article, two simulation times are considered for the ARM case: 6 h for cloud formation and 9 h for the end of cloud growth. BOMEX, ACEheterog and ACEhomog are quasi-stationary cases studied after 5 h of simulation for the cumulus case and after 3 h for the two stratocumulus cases.

In the following, the horizontal subgrid cloud variability is estimated by the cloud variability in the entire LES domain as its domain is of the same order of magnitude than a grid size of a NWP model. At a given time for each vertical level, a statistical analysis is performed over the corresponding horizontal LES domain to provide the distribution (referred to LES distributions hereafter) and the different moments of the conservative variables as well as the cloud characteristics, in particular, the cloud fraction and the mean cloud liquid content. Those LES moments are used for defining the theoretical PDF described later. In this study, because we used LES moments to compute parametrized PDF and therefore parametrized cloud fraction and mean cloud liquid water content, we do not test errors in the host model's prediction of moments (due to errors in parametrization schemes, such as the turbulence and the shallow-convection parametrizations).



(a) ARM case - 9 h **(b)** BOMEX case - 5 h

(c) ACEhomog case - 3 h **(d)** ACEheterog case - 3 h

Fig. 1 3D views of the cloud water content for each PBL cloud case: the cumulus ARM case after 9 h of simulation, the cumulus BOMEX case after 5 h of simulation, and the ACEhomog and ACEheterog cases both after 3 h of simulation. In the ARM case, 6 h corresponds to the cloud formation and 9 h to the end of the cloud growth. BOMEX, ACEhomog and ACEheterog are quasi-stationary cases

2.2 Assumed PDF Method

One of our goals is to determine which statistical distribution should be chosen to mimic the LES distributions. For this selection, most of the studies, such as those of Sommeria and Deardorff (1977) or Bougeault (1981) use the Assumed PDF method, described by Golaz et al. (2002a), to compute the parameters of the different distributions. In this method, the first statistical moments (mean, standard deviation, skewness...) are obtained from the LES and used in the different theoretical distributions. Then, those theoretical distributions are compared to the LES distributions. For unimodal distributions, parameters may be estimated with the Method of moments (see details in Appendix) using expressions relating them to theoretical moments. For the double Gaussian distribution, an analytical iterative method (described in the Appendix) is used for computing the larger number of necessary parameters. In this article, we evaluate five families of unimodal statistical laws and one family of bimodal statistical laws, all of which are commonly used in the literature. They are described in the next subsection.

The comparison of the cloud fraction (CF) and the mean non-precipitating cloud water content computed from the LES data $CF_{(\mathrm{LES})}$ and $\overline{r_c}_{(\mathrm{LES})}$[1] and from the theoretical distributions

[1] With the "All or nothing" method, for each point of the LES domain, $r_c = r_t - r_{sat}(T)$. r_{sat} is the saturation mixing ratio depending on the temperature T. Then, for each vertical level, $\overline{r_c}_{(\mathrm{LES})}$ is deduced from horizontal averaging of r_c. We assume a cloudy point ($CF = 1$) if $r_c > 10^{-12}\,\mathrm{kg\,kg^{-1}}$. Then, $CF_{(\mathrm{LES})}$ is equal to the proportion of cloudy points relative to the total number of points at a given vertical level.

CF and $\overline{r_c}$ given by Eqs. 1 and 2, provide a practical way to evaluate these distributions (see Sect. 4)

$$CF = \int_{\alpha}^{+\infty} G(x)\,\mathrm{d}x \tag{1}$$

$$\overline{r_c} = \int_{\alpha}^{+\infty} (x-\alpha)\,G(x)\,\mathrm{d}x \tag{2}$$

where α is the liquid saturation threshold according to the chosen statistical variable X and G is the theoretical PDF. For warm cloud cases, the liquid condensation process leading to cloud formation begins if the air reaches a liquid saturation corresponding to 100% relative humidity.

The two statistical variables X often used in cloud statistical schemes are

- the non-precipitating total water content $r_t = r_v + r_c$ where r_v is the water vapor content and r_c is the cloud water content. In this case, the liquid saturation threshold is a function of the mean temperature in the grid box: $\overline{r_{sat}} = r_{sat}(\overline{T})$. This neglects the role of temperature variability for the cloud formation;
- the saturation deficit which quantifies the local difference to saturation inside the grid box. It depends on a combination of r_t and the liquid potential temperature θ_l and summarizes both the total water and the temperature variability inside the grid. This variable is denoted s deliberately, regardless of the usual formulations of the literature, and is expressed as follows:

$$s = a_l\,(r_t - r_{sat}(\theta_l)) \tag{3}$$

  where a_l is a coefficient obtained by combining the definition of saturation and the Clausius-Clapeyron relation, as shown in Chaboureau and Bechtold (2002). This variable accounts for the linearized temperature fluctuations on saturation. Note that Q_1, often used in the literature, corresponds to $\overline{s}/\sigma_s$. In a subgrid condensation scheme, conservative variables such as r_t and θ_l are commonly used as prognostic variables because they are conserved under evaporation/condensation processes (see Sommeria and Deardorff (1977) and Mellor (1977)). Supersaturation occurs when s becomes greater than 0. In this case, considering temperature fluctuations implies a variable saturation threshold $r_{sat}(\theta_l)$ inside the grid box.

The relevance of using these two variables in a statistical subgrid cloud scheme is discussed in Sect. 3.1.

2.3 Theoretical Distributions

This section briefly describes the different theoretical distributions for fitting the LES distributions with the Assumed PDF method. A full description of this method is given in the Appendix and uses the first two or three statistical moments and the bounds of the LES distributions.

1. The unimodal *simple Gaussian distribution* has an unbounded symmetric bell-shaped curve that is centred on its mean value. The two distribution parameters are also its first two statistical moments. It has been used by Sommeria and Deardorff (1977), Mellor (1977) and Bougeault (1981).

2. The unimodal *triangular distribution* is a simple symmetric triangle centred on its mean value. The two distribution parameters are again the first two statistical moments. It has been used by Smith (1990). It has the advantage of being easily implemented in a numerical model because the PDF shape facilitates the integral computations.
3. The unimodal *gamma distribution* is valid only for positive variables and defined by a shape parameter, which determines the skewness, and a scale parameter, which stretches or squeezes the PDF curve. It is a flexible distribution that is bounded on the left by zero and is positively skewed. It has been used by Bougeault (1982).
4. The unimodal *log-normal distribution* is valid only for positive variables and is defined by two parameters: the mean and the standard deviation of the logarithm of the statistical variable. It is flexible and can vary from a normal distribution to a positively skewed distribution according to the cloud type. It has been used by Bony and Emanuel (2001).
5. The unimodal *beta distribution* is defined by four parameters: two bound parameters and two positive shape parameters. The beta PDF is very flexible because it takes J-, U- or bell-shapes with a variable skewness. It was proposed by Tompkins (2002). Two different beta distributions are considered in this article (see Appendix): the first, called β_1, is constrained by the first two statistical moments and the second, called β_2, is constrained by the first three statistical moments.
6. The bimodal *double Gaussian distribution* is a linear combination of two individual simple Gaussian distributions. The distribution is defined by five parameters, which are a relative weight and the mean and standard deviation for each simple Gaussian distribution. The double Gaussian distribution may drift into a symmetric distribution if the two distributions overlap, or into a skewed distribution. It has been used by Lewellen and Yoh (1993) and Larson et al. (2001a, 2002) as a joint distribution for several statistical variables.

3 Statistical Analysis of the LES Data

3.1 Choice of the Statistical Variable

In this section, the sensitivity to the choice of the statistical variable using r_t or s for the cloud field computation is tested. Figure 2 shows vertical profiles of the cloud fraction (left column) and the mean cloud water content (right column), values deduced directly from LES data in red ($CF_{(LES)}$ and $\overline{r_c}_{(LES)}$), and diagnosed from LES distributions of s (in blue) and r_t (in green), using Eqs. 1 and 2. For the ARM and BOMEX cumulus cases, the cloud fraction computed with r_t is always underestimated compared with the LES cloud fraction CF_{LES}. The computation from s significantly improves the estimation of the cloud fraction. This is in agreement with Tompkins (2003) that showed with data that the temperature variability could not be neglected. The differences in the computation of $\overline{r_c}$ are less obvious but $\overline{r_c}$ is usually slightly overestimated with r_t. Table 2 gives the errors averaged over all vertical levels for the computation of CF and $\overline{r_c}$ for the four LES. Using r_t leads to averaged errors greater than 50% while using s generates averaged errors lower than 7%. Although errors are smaller (10–20%) for the stratocumulus cases than for cumulus cases, the use of r_t does not correctly estimate the cloud fields. Although errors for the mean cloud water content are less important than for the cloud fraction, the underestimation of CF with r_t impacts negatively the computation of the local cloud water content $r_c = \overline{r_c}/CF$ and this can lead to significant

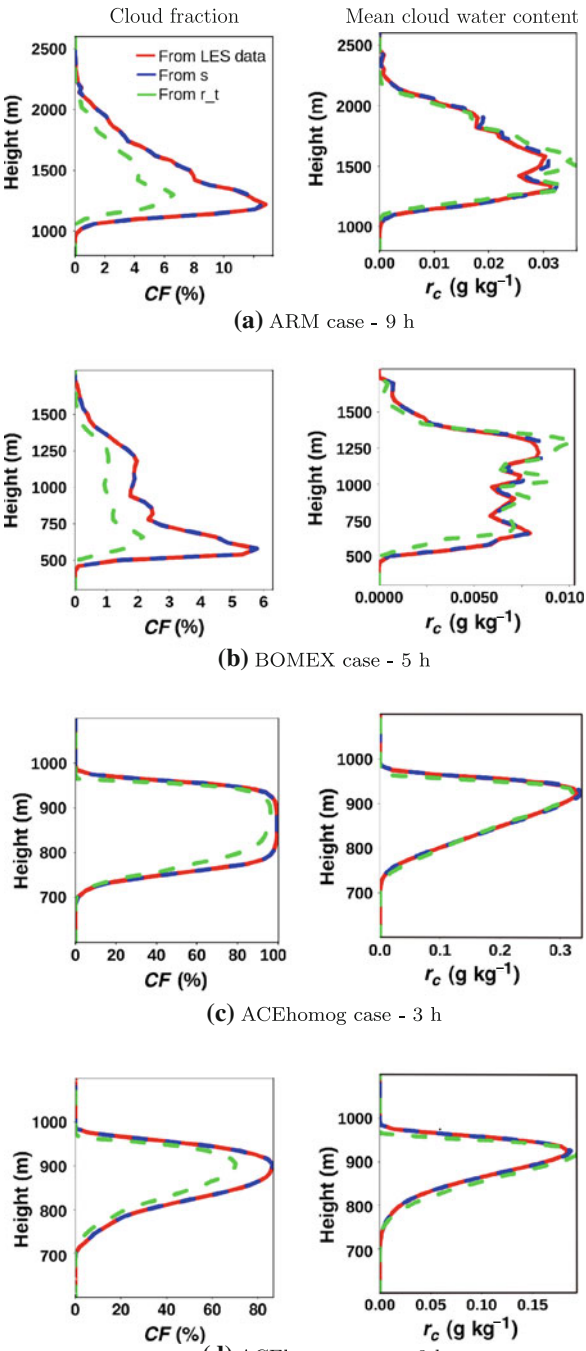


Fig. 2 Vertical profiles of the cloud fraction (*left column*) and the mean cloud water content (*right column*) for **a** the ARM case, **b** the BOMEX case, **c** the ACEhomog case and **d** the ACEheterog case. The *red* profile is for LES data, the *blue dashed* profile is computed from the LES distributions of s and the *green dashed* profile from the LES distributions of r_t

Table 2 Mean CF and $\overline{r_c}$ errors over all vertical levels in the cases where s and r_t are used as predictors in expressions 1 and 2

Cases	ARM		BOMEX	ACEhomog	ACEheterog
Hours of simulation	6 h	9 h	5 h	3 h	3 h
Mean of $\left\lvert \frac{CF_{(\mathrm{LES})} - CF_{(\mathrm{s})}}{CF_{(\mathrm{LES})}} \right\rvert$ (%)	4	1.21	2.2	3.18	3.05
Mean of $\left\lvert \frac{CF_{(\mathrm{LES})} - CF_{(r_t)}}{CF_{(\mathrm{LES})}} \right\rvert$ (%)	58.74	47	54.13	9.38	20.38
Mean of $\left\lvert \frac{\overline{r_c}_{(\mathrm{LES})} - \overline{r_c}_{(\mathrm{s})}}{\overline{r_c}_{(\mathrm{LES})}} \right\rvert$ (%)	0.36	6.15	0.81	1.87	2.02
Mean of $\left\lvert \frac{\overline{r_c}_{(\mathrm{LES})} - \overline{r_c}_{(r_t)}}{\overline{r_c}_{(\mathrm{LES})}} \right\rvert$ (%)	42.07	12.13	4.66	0.97	11.49

errors in the radiative transfer. In conclusion, using the total water content as a predictor gives significant errors in the estimation of the cloud fields. In the later sections, the cloud fraction and the mean cloud water content are computed using the saturation deficit s.

3.2 Study of the Shape of the LES Distributions

The shapes of the LES distributions deduced from LES are illustrated in Fig. 3 for various levels from the cloud base (e) to the cloud top (a) for the ARM case (left column) and the two stratocumulus cases (middle and right columns). The LES PDFs are non-symmetric bell-shaped curves with a positive or negative skewness. In most examples shown in Fig. 3, most values of the saturation deficit are concentrated in a main mode around the mean value. Nevertheless, the non-zero skewness, which appears for low or high values of s depending on the cloud type, is often associated with a second mode, which may evolve into a long tail for cumulus and stratocumulus cases. Note that the BOMEX case is not shown here as it is very similar to the ARM case.

For the ARM cumulus case, a second mode is clearly seen for high values of supersaturation in the lower part of the cloud layer (Fig. 3d), where the mean cloud water content is maximum. Higher in the cloud layer, the second mode vanishes into a long flat tail in the distribution of s. Inside the cloud layer, the skewness (not shown) is positive near the cloud base and increases up to the cloud top, where it reaches its maximum value. It is noted that the appearance of the second mode inside the cloud layer coincides with an increase of the distribution spread with height.

For the stratocumulus cases, the shapes of the LES distributions strongly vary with height and sometimes, a second mode appears. The skewness (not shown) is weakly positive in the lower part of the cloud layer and decreases upwards until it becomes zero or negative. Figure 3 shows an increase of the distribution spread with height. The PDFs of the two stratocumulus cases gradually shift towards high positive values of s. For the "quasi-resolved" homogeneous stratocumulus layer, the PDF moves totally beyond saturation with $CF = 100\%$ above the middle of the cloud except at cloud top. For the fractional stratocumulus layer, the model grid is very often unsaturated on average ($\overline{s} < 0$) inside the cloud layer, and the cloud fraction never reaches 100%.

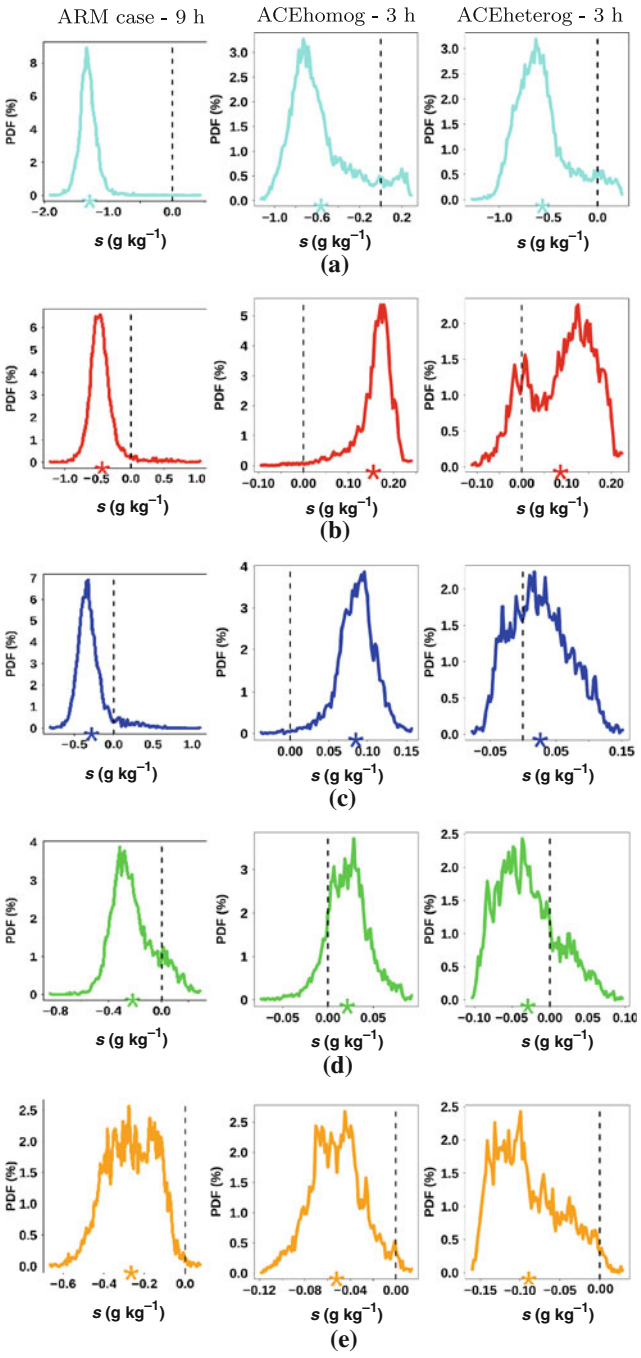


Fig. 3 LES distributions of s inside the cloud layer for the ARM case (*left column*) and the ACEhomog and ACEheterog cases (*middle* and *right columns*). The *top panels* **a** correspond to the top of the cloud layer and the *panel* **e** to the base of the cloud layer. The *panels* **b**–**d** correspond approximatively to the three quarters, the half and the first quarter of the thickness of the cloud layer (see Fig. 2). The *vertical dashed line* represents saturation ($s = 0$), and the *star* on the x-axis represents the mean value of s. Note that the y-axis is different for each panel to highlight the distribution characteristics

For cumulus and stratocumulus clouds, the distribution of the saturation deficit considerably varies with height, and is often bimodal. PBL cloud parametrizations have to take this evolution into account, particularly for the second mode, which generally covers the saturated part of the distribution.

3.3 PDF Shape and Cloud Type

This section explores the link between the evolution of the PDF shape and the atmospheric processes leading to the cloud evolution.

3.3.1 Cumulus Cloud

For shallow cumulus clouds, the first mode of the distribution of s is on the dry side of the distribution. The second mode, which may take the shape of a long flat tail, is essential for the description of shallow cumuli because it corresponds to the cloudy values of s. This particular shape of the PDF produces the low cloud fraction and is characteristic of shallow convective situations. Figure 4b–e shows horizontal cross-sections of s, r_t, θ_l and w, for the ARM case inside the cloud layer at the level of maximum mean cloud fraction. The high values of supersaturation corresponding to the second mode (red zone on the cross-section 4b), are clearly correlated with high values of r_t, negative anomalies of θ_l and positive anomalies of w. They are the signatures of cloudy shallow convective thermals bringing relatively cooler and moister air above the inversion. In the upper part of the cloud, where the second mode becomes more flattened, the effects of the shallow convection are still visible but sparser. Investigation of other times for the ARM and BOMEX cases leads to the same conclusions: the second mode is related to convective processes. The cloud fraction and the mean cloud water content are fully related to the description of the tail of the distribution of the saturation deficit for shallow cumulus cases.

3.3.2 Stratocumulus Cloud

A conditional sampling of LES data for the two stratocumulus cases is carried on in order to characterize the processes of the two modes of the LES distributions. Results are shown in Fig. 5. The LES horizontal domain at a given level is partitioned into three classes: clear sky ($r_c = 0$) on the dry part of the distribution, subsiding cloudy zones ($r_c > 0$ and $w < 0$) and ascending cloudy zones ($r_c > 0$ and $w > 0$). Figure 5 shows the distributions of s at various levels for the entire domain in black and for each of the three classes in colours. For the two stratocumulus cases, the ascending cloudy zones are moister and correspond to higher values of supersaturation than the subsiding cloudy zones. Near the cloud base (Fig. 5c), the ascending zones predominate over the subsiding zones. The subsiding cloudy zones appear at increasingly greater heights than the ascending cloudy zones. In the heterogeneous stratocumulus cloud, the ascending cloud zones always predominate over the subsiding cloudy zones whereas in the homogeneous cases both cloudy zones are more equally distributed. The contrast between the two stratocumulus cases is due to the characteristic of the free troposphere, which is drier for the fractional stratocumulus cases. As shown in Fig. 1d, large variability of the cloud base exist in this case, a consequence of the intrusion of drier air at cloud top. This also leads to a smaller fraction of subsiding cloudy zones. For the moister homogeneous stratocumulus cloud, *CF* equals 100%, and the effects of the intrusion of dry air are much less visible. The parametrization of fractional stratocumuli clouds in NWP models

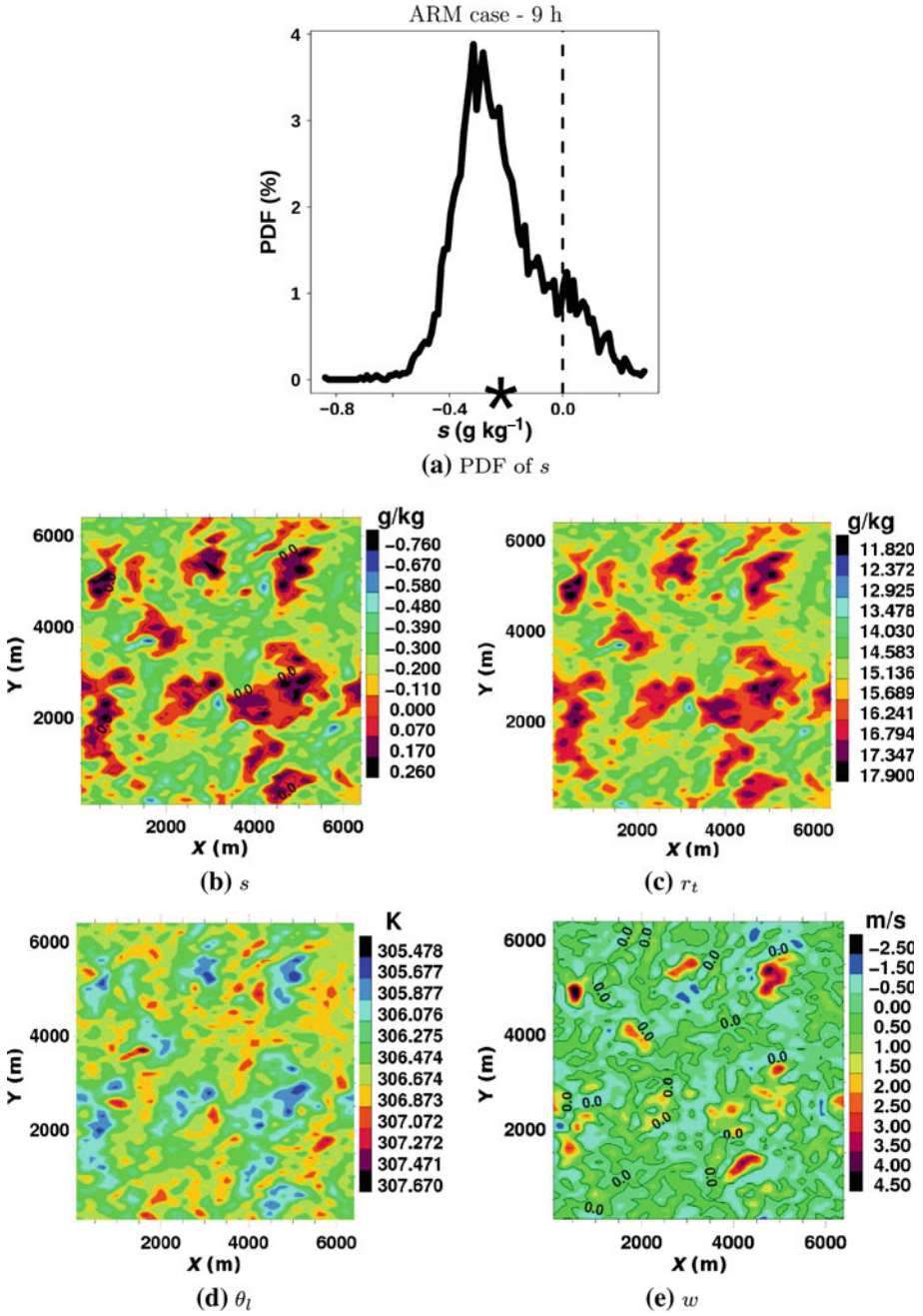


Fig. 4 ARM cumulus case: **a** LES distribution of s at the level of maximum mean cloud water content (corresponding to the quarter of the thickness of the cloud layer in Fig. 3d for the ARM case); the *dashed line* represents saturation ($s = 0$). Horizontal cross-sections of s **b**, r_t **c**, θ_l **d** and w **e** at the same level. In **b** and **e**, the *black line* corresponds to the 0-isocontour

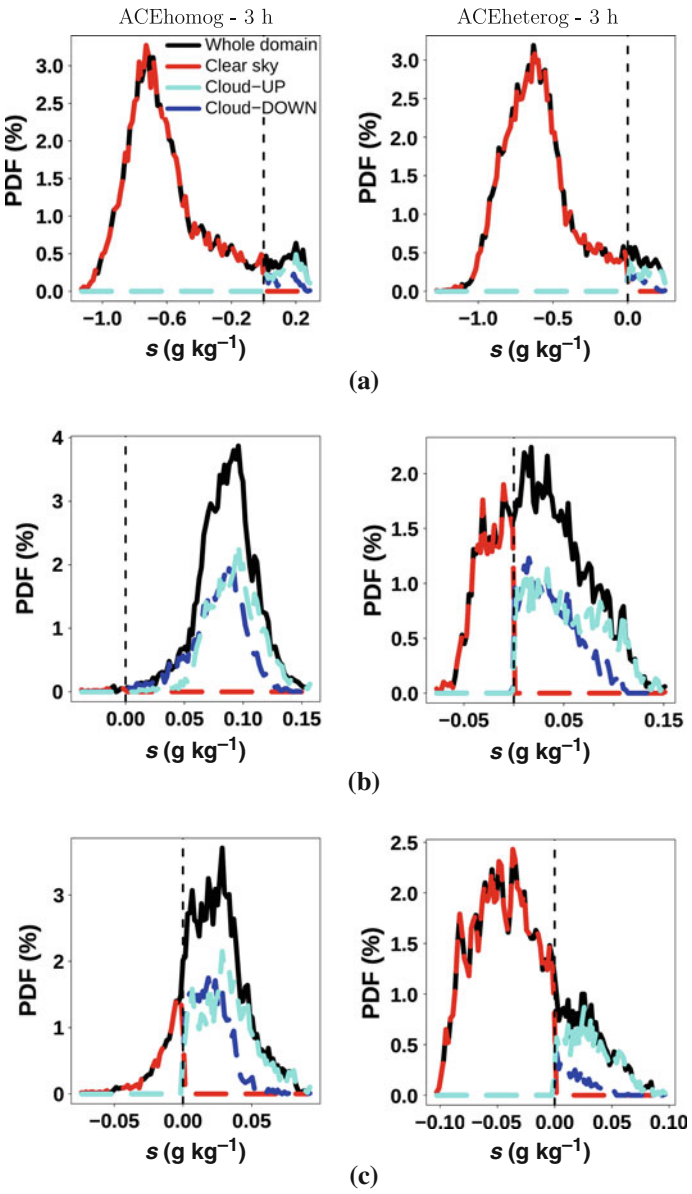


Fig. 5 LES distributions of s inside the cloud layer for the ACEhomog and ACEheterog cases **a** at cloud top, **b** at the half of the thickness of the cloud layer and **c** at cloud base. The distribution of the whole horizontal domain is in *black*. The *red dashed* distribution corresponds to the clear sky, the *azure dashed* distribution to the ascending cloudy zones and the *dark blue* dashed distribution to the subsiding cloudy zones. The *vertical dashed line* represents saturation ($s = 0$)

may require a subgrid cloud scheme because an "All or Nothing" method, which could be used for resolved clouds such as stationary stratocumuli, does not consider the irregularities of the cloud layer when $CF < 100\%$.

3.4 Sensitivity to the Horizontal Domain Size

The horizontal resolution used in NWP models is constantly evolving and will be more fine in the future. For instance, the representation of cumulus clouds as subgrid processes is questionable with decreasing grid size, because of their size being a few hundreds of metres. In this section, the sensitivity to the horizontal domain size of the LES is evaluated. The previous statistical analysis was performed for the entire horizontal LES domain with a size of $6.4 \times 6.4\,\mathrm{km}^2$ for cumulus cases and $5 \times 5\,\mathrm{km}^2$ for stratocumulus cases. Here, we tackle the question of the robustness of previous results with respect to the horizontal subgrid variability for potentially finer model grids. For this, the initial LES domain is divided into 4 or 16 identical sub-domains ($\Delta x = 3.2$ or 2.5 km for four sub-domains and $\Delta x = 1.6$ or 1.25 km for 16 sub-domains) and the same statistical analysis is carried out for each sub-domain.

For the ARM and BOMEX cases, the results hold for smaller horizontal domain sizes (not shown here). The horizontal cloud fluctuations are important for shallow cumulus clouds and they are still subgrid clouds for horizontal resolutions of 3.2 and 1.6 km. As for the entire LES domain, using s as the statistical variable to consider the variability of both temperature and total water content (Sect. 3.1) provides better results according to the cloud fraction and the mean cloud water content. For these finer resolutions, the LES distributions also show two local maxima and the second, less marked, can evolve into a long flat tail.

For the ACEhomog case, the horizontal fluctuations inside a model grid are weak. The stratocumulus cloud is completely resolved horizontally in terms of the condensation/evaporation processes for resolutions of 2.5 and 1.25 km in the sense that the horizontal size of the cloud is larger than the grid mesh. Nevertheless, no assumption is made about the moisture transport such as eddy circulation, which is still partly subgrid. The choice of the theoretical distribution for the subgrid cloud scheme is arbitrary since the PDF is totally beyond saturation with $CF = 100\%$. However, the description of the fractional ACEheterog case requires use of a subgrid cloud scheme as the cloudy cells are smaller due to the alteration of the cloud layer from its base. As for cumulus cases, the LES distributions also show two local maxima for finer resolutions. A subgrid scheme based on the saturation deficit and a bimodal distribution also provides better results at 2.5 and 1.25 km.

The previous results obtained with the statistical analysis of the LES data are robust for sub-domains of the initial LES domain.

4 Evaluation of the Theoretical Distributions

This section evaluates how the theoretical unimodal and bimodal distributions described in Sect. 2.3 are able to fit the LES distributions of the saturation deficit. The PDF shapes studied in Sect. 3.3 for shallow cumulus and stratocumulus cases are not easily fitted by most of the classical distributions, which are essentially constrained by the main mode. It is noted that the results obtained in this study are also valid for the total water content for which the gamma and the lognormal laws were also tested.

4.1 The Unimodal Distributions

The unimodal distributions commonly used in the literature (Sect. 2.3) are evaluated against the LES distributions (Sect. 3). Figure 6 presents the vertical profiles of the cloud fraction and the mean cloud water content computed from the LES distributions and the various unimodal distributions. Figure 7 shows scatter plots of these two cloud fields for various

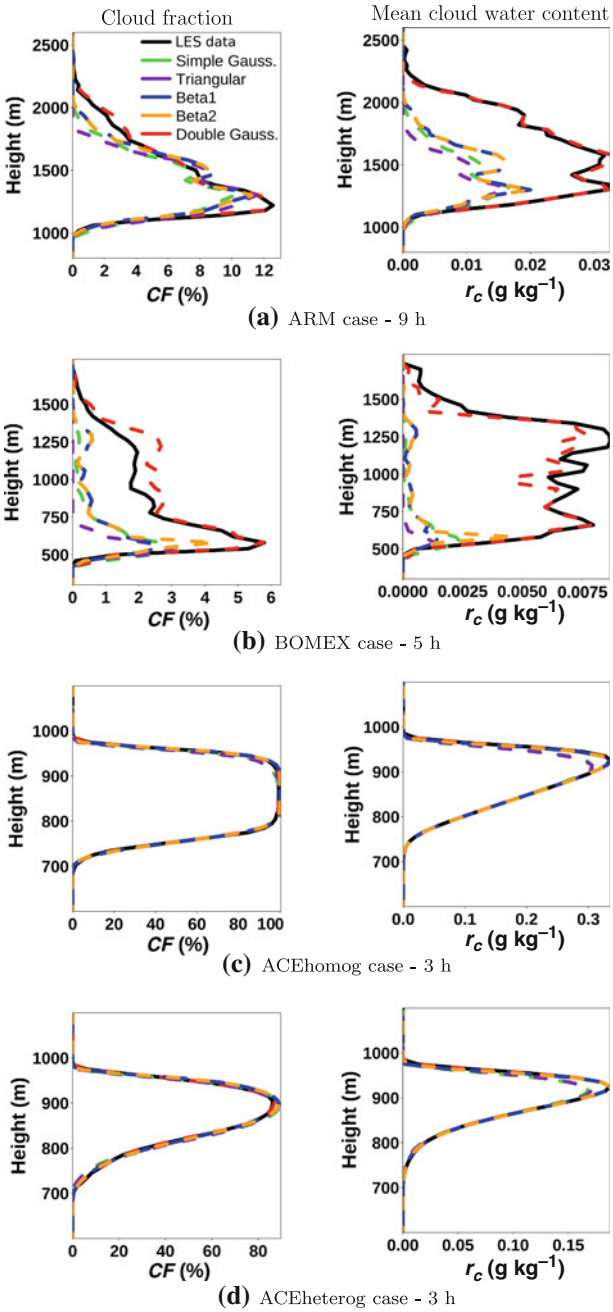


Fig. 6 Vertical profiles of the cloud fraction (*left column*) and the mean cloud water content (*right column*) for **a** the ARM case, **b** the BOMEX case, **c** the ACEhomog case and **d** the ACEheterog case, deduced from the LES distributions (*black*), from the simple Gaussian distribution (*green*), from the triangular distribution (*purple*), from the β_1 distribution (*blue*), from the β_2 distribution (*orange*) and from the double Gaussian distribution (*red*)

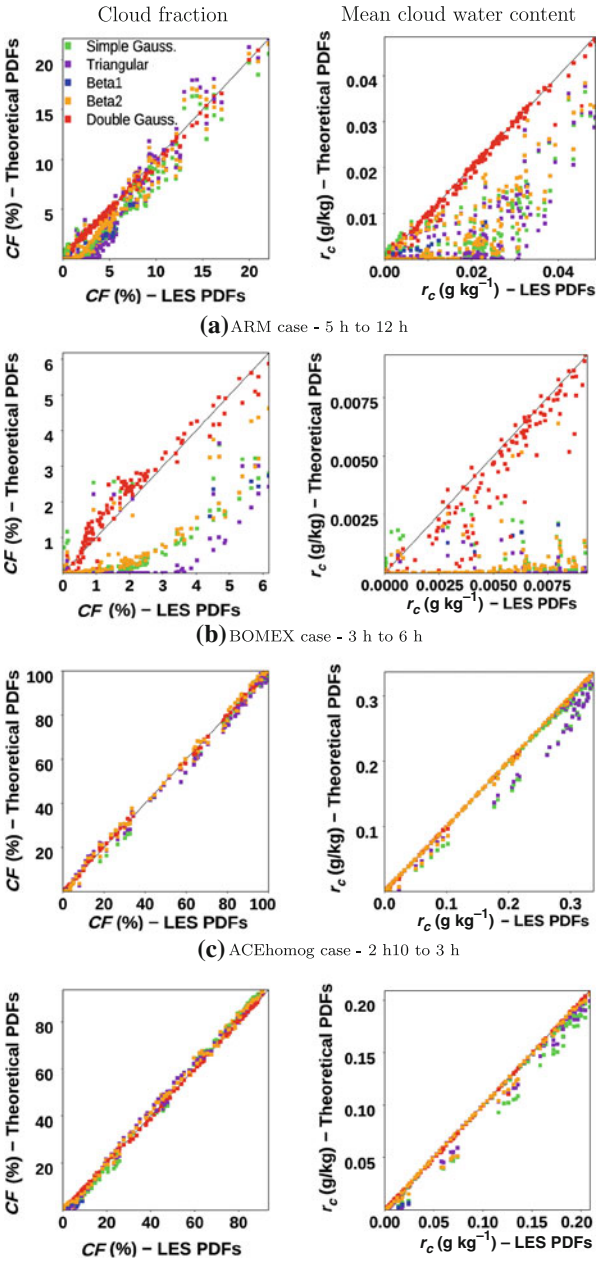


Fig. 7 Scatter plots of the cloud fraction (*left column*) and the mean cloud water content (*right column*) for **a** the ARM case, **b** the BOMEX case, **c** the ACEhomog case and **d** the ACEheterog case. The simple Gaussian distribution (*green*), the triangular distribution (*purple*), the β_1 distribution (*blue*), the β_2 distribution (*orange*) and the double Gaussian distribution (*red*) are compared to the LES distributions. Each *squared symbol* represents a vertical level in the cloud layer at various times of simulation: between 5 and 12 h of simulation for ARM, 3 and 6 h of simulation for BOMEX and 2 h 10 min and 3 h of simulation for the two stratocumulus cases

times and levels for the different cumulus and stratocumulus cases (each point of the scatter plots represents an instantaneous value of the cloud field for a given hour of simulation and a given vertical level). Table 3 quantifies the errors obtained with the different theoretical distributions for each case.

For ARM and BOMEX (Fig. 6a, b), all the unimodal distributions clearly underestimate the cloud fraction and the mean cloud water content. Symmetric distributions, such as the simple Gaussian (in green), the triangular (in purple) and the β_1 (in blue) distributions are not flexible enough and therefore not able to reproduce the tail of the skewed distribution. The β_2 distribution (in orange) improves the computation of the cloud fraction for the ARM case because it conserves the skewness of the LES data and the variation of the PDF shape with height (not shown), but it still gives large errors for BOMEX (Fig. 7a, b). Errors on $\overline{r_c}$ are often greater than 50% in the ARM case and 90% in the BOMEX case with all unimodal distributions. Thus, none of the unimodal distributions, including the skewed β_2 distribution, are able to provide correct values of cloud fraction and mean cloud water content in the case of sparse subgrid clouds like shallow cumuli. All the unimodal distributions also systematically underestimate the height of the cloud top (Fig. 6). This is mainly explained by the fact that the shape of a unimodal PDF is constrained by the main mode in the dry part of the domain in shallow cumulus cases.

For the two stratocumulus cases, errors on the computation of cloud fields from unimodal distributions rarely exceed 5% (Table 3). In Fig. 6c and d, the heights of the cloud base and the cloud top are correctly estimated. Considering the shapes of the previous LES distributions (see Fig. 3), the unimodal distributions may be adequate for the description of stratocumulus clouds, except for values of cloud fields associated with a PDF showing two modes with comparable amplitudes (Fig. 3b for the ACEheterog case). As for the ARM case, it seems that the β_2 distribution gives the best estimate of cloud fields (Fig. 7c, d).

A unimodal distribution seems sufficient to describe the stratocumulus clouds but not the cumulus clouds. A more independent representation of the second mode would provide a better description of the cloudy part of the distribution, particularly in the case of a long flat distribution tail for cumulus clouds.

4.2 The Double Gaussian Distribution

To improve the representation of the second mode, the double Gaussian distribution is now evaluated and the vertical profiles of the cloud fraction and the mean cloud water content computed from this bimodal distribution are given in red in Fig. 6 for the four cases. In all cases, the heights of the cloud base and the cloud top are correctly estimated with the double Gaussian distribution. The cloud field estimates are obviously improved in cumulus cases with the double Gaussian distribution (Fig. 7a–d), and the errors are significantly reduced (Table 3) compared to those generated by unimodal distributions. They do not exceed 4%, except for the BOMEX case where the cloud fraction and mean liquid water content are small leading to large relative errors in% even though the difference between this distribution and the LES distribution are never larger than respectively 1% and $3 \times 10^{-6}\,\mathrm{kg\,kg^{-1}}$ for the cloud fraction and the mean liquid water content (Fig. 7). The benefits of the double Gaussian distribution are less evident (Fig. 6c, d) in stratocumulus cases but it still provides the best estimates of cloud fields. Errors are globally lower than 1%.

The comparison of the simple and the double Gaussian distributions for the four cases (not shown here) highlights that the double Gaussian distribution is able to correctly approach more or less skewed distributions, as in Fig. 3b the ACEhomog cases, or distributions with two marked second modes as in Fig. 3c for the ARM case and in Fig. 3b for the ACEheterog cases. The second mode of the double Gaussian distribution is very often well positioned with the correct amplitude. For cumulus cases, it evolves in a long flat tail in the upper part of the cloud layer. The main mode of the bimodal PDF is also in better agreement with the LES distributions. The independent computation of the parameters (mean and standard deviation)

Table 3 Mean errors for CF and $\overline{r_c}$ over all vertical levels for each theoretical distribution

Cases		ARM		BOMEX	ACEhomog	ACEheterog
Hours of simulation		6 h	9 h	5 h	3 h	3 h
Mean of $\left\lvert \frac{\overline{r_c}(\mathrm{REF}) - \overline{r_c}(\mathrm{THEOR})}{\overline{r_c}(\mathrm{REF})} \right\rvert$ (%)	Simple Gaussian	9.69	23.7	70.72	0.79	1.95
	Triangular	68.42	36.01	22.92	6.59	9.03
	β_1	4.71	15.32	69.43	0.28	2.22
	β_2	0.55	15.32	66.16	0.26	2.55
	Double Gaussian	0.42	0.95	15.02	0.41	0.25
Mean of $\left\lvert \frac{CF(\mathrm{REF}) - CF(\mathrm{THEOR})}{CF(\mathrm{REF})} \right\rvert$ (%)	Simple Gaussian	34.68	68.56	92.14	5.07	7.05
	Triangular	11.06	58.31	89.90	2.68	2.31
	β_1	32.60	64.27	92.85	1.48	3.20
	β_2	28.86	64.27	90.74	1.48	2.83
	Double Gaussian	3.65	1.77	13.91	0.70	1.07

for each simple Gaussian law allows the choice of two more or less narrow modes with the right amplitudes. Nevertheless, it is noted here that the better behaviour of this distribution is expected due to more free parameters (two for the simple Gaussian and five for the double Gaussian distribution). Nevertheless, according to the shape of the LES distribution, a unimodal distribution even with more free parameters will never allow the reproducion of the bimodality. The double Gaussian law is the simplest law that allows taking into account this characteristic. It is therefore a trade-off between complexity and efficiency.

The double Gaussian distribution improves the description of sparse subgrid clouds such as shallow cumuli and fractional stratocumuli. The construction of a statistical cloud scheme based on the double Gaussian distribution seems the best alternative to obtain a unified approach for the treatment of PBL clouds. If necessary, the double Gaussian distribution correctly switches over to a simple distribution showing a single maximum. Some guidance for the development of a cloud scheme, based on those results, are given in Sect. 5.

4.3 Sensitivity to the Vertical Grid

In the previous section, the statistical analysis was performed for each level of the LES vertical grid. The vertical resolutions of the four LES used as reference (Table 1) are much finer than the resolution generally used in a NWP model. In order to test the sensitivity of the previous results to the number of levels in the cloud layer, several layers of the LES were averaged to obtain a new sample of data with a vertical resolution similar to the one used for the operational models. The new vertical grid contains 40 levels with a vertical resolution ranging from 40 m near the ground to 500 m above 2 km. Figure 8 shows vertical profiles of the cloud fraction and the mean cloud water content averaged on the new vertical grid deduced from LES data (in black) and computed from the simple Gaussian distribution (in green) and from the double Gaussian distribution (in red) for the two cumulus cases and for the ACEheterog case.

For ARM and BOMEX (Fig. 8a, b), cloud fields are correctly diagnosed (Table 4) and the height of the cloud top is better estimated with the double Gaussian distribution than with the simple one. As for cumulus cases, the use of a bimodal distribution improves the horizontal description of stratocumulus clouds, especially for the fractional stratocumulus

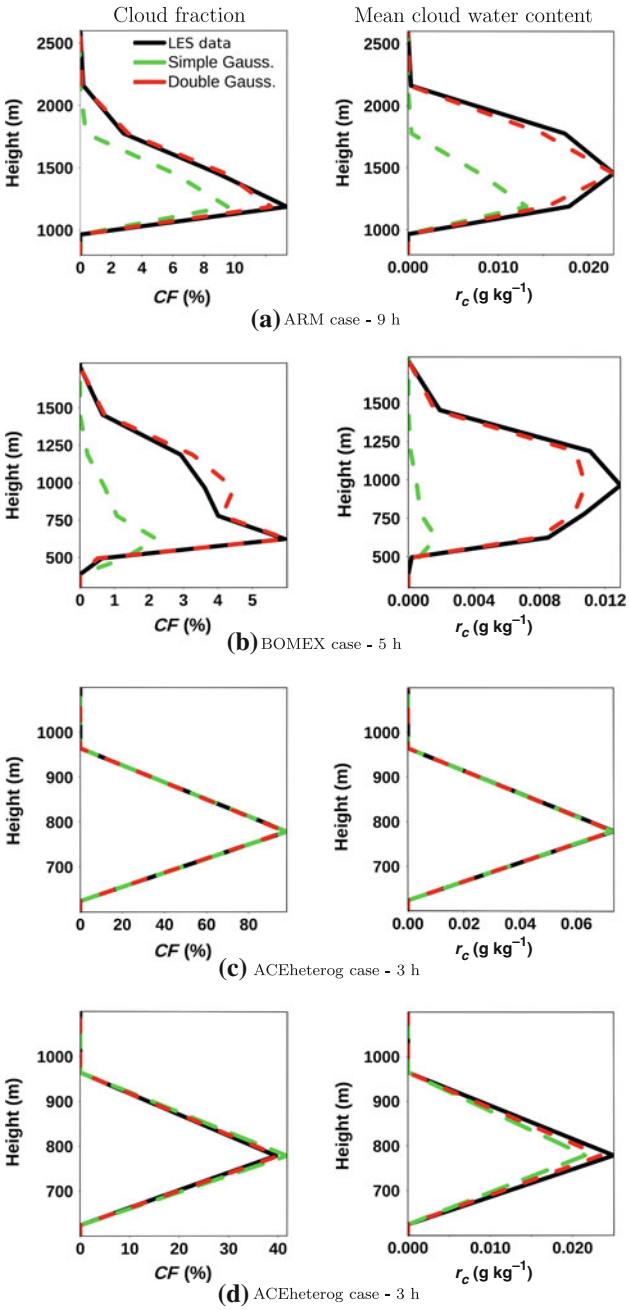


Fig. 8 Vertical profiles of the cloud fraction (*left column*) and the mean cloud water content (*right column*) for **a** the ARM case, **b** the BOMEX case, **c** the ACEhomog case and **d** the ACEheterog case and deduced from the LES distributions (*black*), the simple Gaussian distribution (*green*) and the double Gaussian distribution (*red*) for a vertical grid typical of NWP models

Table 4 As in Table 3 for the simple Gaussian distribution and the double Gaussian distribution for a vertical grid typical of NWP models

Cases		ARM		BOMEX	ACEhomog	ACEheterog
Hours of simulation		6 h	9 h	5 h	3 h	3 h
Mean of $\left\lvert \frac{CF_{(\mathrm{REF})} - CF_{(\mathrm{THEOR})}}{CF_{(\mathrm{REF})}} \right\rvert$ (%)	Simple Gaussian	26.99	43.93	69.21	0.08	5.66
	Double Gaussian	6.72	4.01	14.01	0.57	1.39
Mean of $\left\lvert \frac{\overline{r_c}_{(\mathrm{REF})} - \overline{r_c}_{(\mathrm{THEOR})}}{\overline{r_c}_{(\mathrm{REF})}} \right\rvert$ (%)	Simple Gaussian	40.48	67.97	90.88	0.64	13.88
	Double Gaussian	20	16.87	26.56	0.35	6.24

case (see errors on Table 4). All the conclusions drawn for the statistical analysis with the fine resolution of the LES are still valid for a coarser description of the cloud layer for the four cases.

Nevertheless, Fig. 8c and d raises the problem of the treatment of subgrid clouds in the vertical dimension. The change to a coarser vertical grid causes a smoothing of the vertical profiles of the cloud fraction and the mean cloud water content, leading to a loss of information. It is particularly problematic in the case of shallow stratocumulus clouds, which have a smaller vertical extent than cumulus clouds. This is evidenced by comparing the maximum values of cloud fraction on vertical profiles obtained for the LES vertical grid (red profile in Fig. 2) and for the coarser vertical grid (black profile in Fig. 8): an underestimation of about 50% appears. This large error is due to the lack of information about the vertical extent of the cloud layer between two successive vertical levels. In this article, we focus on the description of the horizontal subgrid cloud variability, but this sensitivity study to the vertical grid shows the need to also consider the vertical subgrid cloud variability in cloud schemes in future studies, to improve the representation of stratocumulus clouds.

5 Keys for the Improvement of Statistical Cloud Schemes

This section provides some guidance for the implementation of a double Gaussian statistical cloud scheme. First, it is shown that the grid cell can be decomposed into an environmental (non-convective) and convective fraction, with a single Gaussian fitted for each sub-domain as suggested in Sect. 3.3.1. Such hypothesis has been also recently used in Neggers (2009) and Jam et al. (2010). Then, it is shown that when splitting the grid cell into those two domains, r_t instead of s can be the statistical variable, reducing the number of prognostic variables from 2 to 1. Eventually, a parametrization for the standard deviation of the convective single Gaussian is provided as well as a formulation for the coverage fraction of the convective domain. This proposition of cloud scheme is built to only require the convective mass flux M and thermodynamical characteristics (θ^{up} and r_t^{up}) in the parametrized cloudy updraft. This way, it can be usable by any model having a shallow-convection parametrization (either classical or EDMF ones).

5.1 Determination of the Two Single Gaussian Distributions

In this section, only cumulus cases (ARM and BOMEX) are considered. First, as shown in Sect. 3.3.1, the second mode in cases of cumulus corresponds to the signature of shallow

convective thermals. This has been confirmed by Jam et al. (2010) using a conditional sampling defined by Couvreux et al. (2010) to distinguish the thermals and the environment fraction. Therefore, to implement the double Gaussian distribution, G_{PDF}, as a statistical cloud scheme, one would need the definition of five parameters: the convective area, α, the mean m and standard deviation σ of the variable for each Gaussian distribution:

$$G_{\mathrm{PDF}} = \alpha G_{\mathrm{convective}} + (1-\alpha) G_{\mathrm{nonconvective}} \tag{4}$$

where $G_{\mathrm{convective}}$ and $G_{\mathrm{nonconvective}}$ are simple Gaussian distribution. As the cloud fraction and the liquid water content are determined linearly from the Gaussian distributions, the total liquid water content and total cloud fraction can be computed separately by each Gaussian distributions independently. A physically based proposition is adopted to determine the mean and the standard deviation of the non-convective domain via the turbulence parametrization and the mean and the standard deviation of the convective domain with the shallow-convection parametrization. The non-convective mean can be assumed equal to the entire domain mean if the convective fraction remains relatively small. We further assume that the turbulence scheme provides the standard deviation for the non-convective zone. Therefore, a formulation for only three remaining parameters must be proposed: the convective fraction, the mean and standard deviation for the convective distribution.

5.2 s or r_{t}?

In Sect. 3, we have shown that the temperature variability cannot be neglected relatively to the humidity variability when considering the whole domain. Nevertheless, once we separate the domain into two sub-domains considered independently, the use of the saturation deficit s is not anymore necessary (the differences are significantly reduced compared to Fig. 2, not shown). This is illustrated in Fig. 9 that compares the s distribution (upper panel) and the r_{t} distribution (lower panel) for the entire domain, the convective and non-convective domains. The saturation threshold in the convective domain is much smaller than the saturation threshold in the other two domains ensuring selection of the whole convective fraction even when using r_{t}. In other words, the temperature variability is mainly accounted for by a top-hat representation between environment and thermals, and therefore the temperature variability in each sub-domain can be neglected. It is noted that, as a consequence, saturation vapour ratio needed to evaluate cloud cover and cloud water content from the PDF of r_{t} must be computed using the average temperature in the sub-domain corresponding to the convective part alone (red vertical line in lower panel in Fig. 9).

5.3 Cloud and Cloud Core

We focus here more deeply on the analysis of the cumulus cloud itself. In cumulus, the dynamics is well known to be driven by vertical updrafts often called “core” of the cloud: this core, more buoyant than the environment, transports vertically the water from the boundary layer upper in the atmosphere. To determine the characteristics of the cloud from the LES, we refine the sampling proposed by Couvreux et al. (2010). The first variable we focus on is the buoyancy mass flux, as it is the first key process in the cloud.

In fact, as shown in Fig. 10a, the distribution of the buoyancy flux over the convective part selected by this sampling (extracting mostly all cloudy points) shows bimodality. There is a

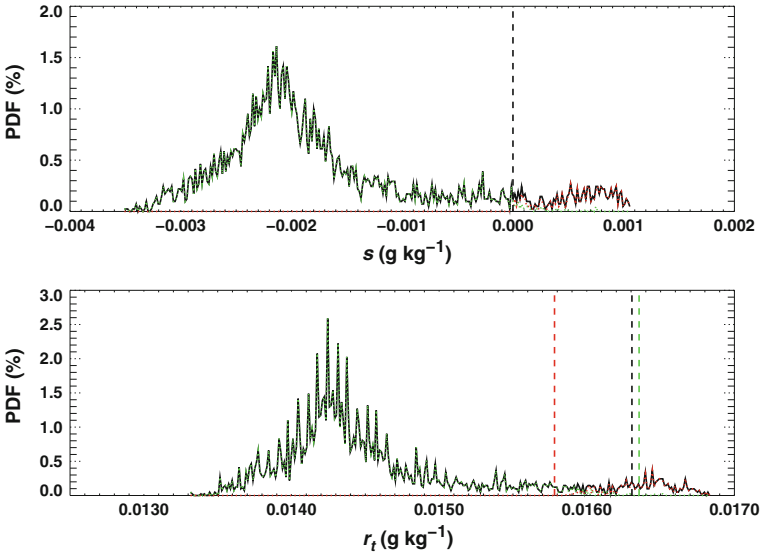


Fig. 9 Distribution of the saturation deficit (*upper panel*) and the total water mixing ratio (*lower panel*) in the lower part of the cloud for ARM case at 9 h. The total distribution is in *black*, the distribution over the convective part in *red* and over the non-convective part in *green*. the convective and non-convective fractions are determined using the conditional sampling proposed by Couvreux et al. (2010). In the *upper panel*, the *vertical line* corresponds to the saturation threshold ($s = 0$). In the *lower panel*, the *vertical lines* correspond to the saturation mixing ratio for the whole domain (*black*), the convective part (*red*) and the non-convective part (*green*)

Gaussian distribution centred over 0 and a long flat tail of buoyant parcels. The buoyant core can then be defined as the points selected by the sampling that in addition have a buoyancy greater than $0.5\,\mathrm{K\,kg\,s^{-1}\,m^{-2}}$. The remaining points are the edges of the cloud. In Fig. 10b and c, the distributions of total water mixing ratio and liquid potential temperature is shown for the total cloud and the cloud core. The core set has the moister and cooler (in term of liquid potential temperature) points, with a relatively small dispersion.

As we see the core is not the total cloud that we seek to parametrize in our cloud scheme. However, as it is the active part of the cumulus, all the mass flux transport occurs in this core. As such, it is this core that is parametrized in shallow convection schemes (particularly in mass-flux approaches). One can therefore consider that the outputs of these schemes are representative of the core of the cloud. From these outputs, one will have to define a cloud scheme representing all the cumulus (core and edges).

To describe the first required variable, the mean of the (Gaussian) distribution associated to the convective part, one can use the mean value of the total water content provided by the shallow-convection parametrization in the core ($r_{\mathrm{t}}^{\mathrm{up}}$). This value may overestimate the mean cloud content (as core is the moister part of the cloud), but this first approximation of the cloud water content allows to discriminate the cloud from the environment in the free troposphere (the latter being much dryer by a large margin in the case of shallow cumulus). It gives a base for the localization of the second Gaussian distribution. It could be improved by setting an offset depending on the standard deviation of the distribution. Nevertheless, determining this offset is not trivial as it depends with height in the cloud, being smaller near cloud base and larger near cloud top. This will be pursued in future studies. Hence, in the following, we assume that $r_{\mathrm{t}_{\mathrm{convective}}} = r_{\mathrm{t}}^{\mathrm{up}}$.

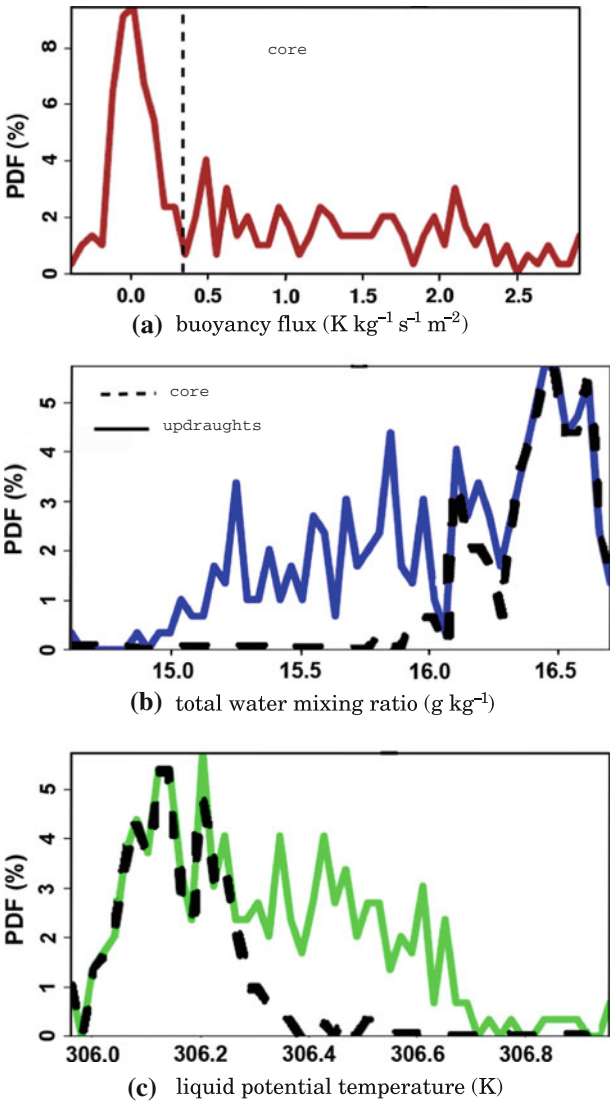


Fig. 10 Distribution of **a** buoyancy, **b** total water content and **c** liquid potential temperature at the level of maximum cloud fraction for ARM at 9 h. The *dashed black lines* for **b** and **c** correspond to the distribution obtained from the updraught core. The core is defined according to **a**

5.4 Formulation for the Thermal Standard Deviation

Several definitions of the standard deviation for the distribution of r_t due to the convection have already been proposed. Lenderink and Siebesma (2000) have proposed a standard deviation of the total water mixing ratio that takes into account the convective impact. Nevertheless, they do not use a bimodal distribution; this standard deviation is then made use of to derive directly the cloud fraction:

$$\sigma^2 = \frac{M(r_t^{up} - \langle r_t \rangle) l_{cloud}}{w_*^{up}} \left(\frac{\partial r_t}{\partial z} \right) \tag{5}$$

where l_{cloud} is the cloud depth, M is the mass flux diagnosed in the shallow convection scheme (it is noted that here $M = \alpha w^{up}$), and w_*^{up} is a vertical velocity scale defined as

$$w_*^{up} = \left(\int_{cloud} \frac{g}{\theta} M \Delta\theta_v \, dz \right)^{1/3} \tag{6}$$

where $\Delta\theta_v$ is the difference in virtual potential temperature between the convective part and the environment. Similarly to Lenderink and Siebesma (2000) and due to the underestimation of the moisture fluctuations, Chaboureau and Bechtold (2005) proposed to take into account the effect of convection on the moisture variability through (but again for a simple Gaussian distribution for the whole grid):

$$\sigma = cMf(z) \tag{7}$$

with c, a tuned constant equal to 3×10^{-3} and $f(z) = 1/a$ with $a = (1 + L\partial r_{sat}(T_l)/C_{pm}\partial T)^{-1}$ a thermodynamic function arising from a linearization of the function for the water saturation mixing ratio.

Jam et al. (2010) proposed a formulation for the standard deviation of the second mode of a double Gaussian parametrization as

$$\sigma^2_{convective} = max \left(\frac{0.015}{\sqrt{\alpha}} \left(r_t^{up} - \langle r_t \rangle \right)^2, \left(\frac{r_t^{up}}{100} \right)^2 \right). \tag{8}$$

Inspired by the study of Lenderink and Siebesma (2000), here we propose, but for the second mode of the distribution only, the following definition for the standard deviation of r_t:

$$\sigma^2_{convective} = \frac{M \left(r_t^{up} - \langle r_t \rangle \right) l_{cloud}}{20 w_*^{up}} \left(\frac{\partial r_t}{\partial z} \right). \tag{9}$$

This formulation has the advantage of depending only on parameters provided by any type of shallow convection scheme. It takes into account not only local updraft quantities, but also integrated quantities such as the cloud depth and the vertical velocity scale. This is a way to take into account the whole physical structure of the shallow cumulus. It is validated for ARM 9 h in Fig. 11b against the standard deviation of the second mode derived from LES.

5.5 Formulation for the Area Coverage of the Thermals

In many shallow convection schemes (Bechtold et al. 1995), the area coverage of the updraughts is not a diagnostic parameter. Nevertheless, to define the double Gaussian distribution, this parameter is needed to determine the weight in-between the two single Gaussian distributions. The following formulation is proposed:

$$\alpha_{convective} = \frac{M}{2 w_*^{up}}. \tag{10}$$

It is evaluated against LES results and the fraction used in Pergaud et al. (2009) in Fig. 11a. In Fig. 12, the cloud fields are evaluated for ARM (two different hours) and BOMEX. The different formulations of the literature are also plotted for comparisons. Cloud fraction and

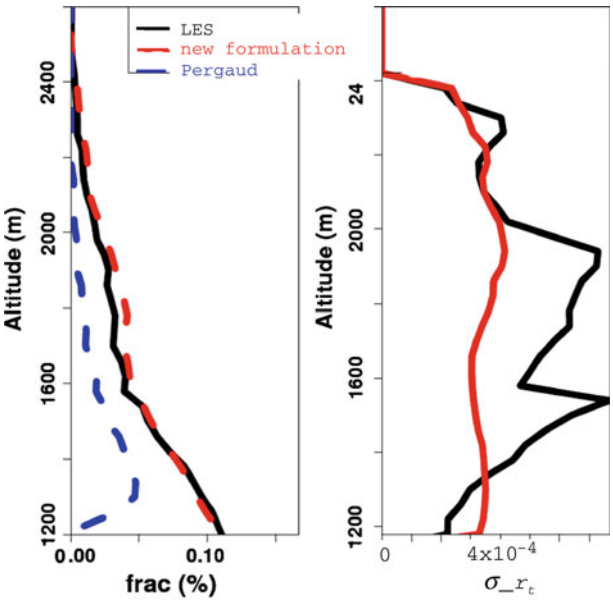


Fig. 11 Vertical profiles of **a** the updraught fraction and **b** the standard deviation of r_t in the updraughts: deduced from LES in *black*, used in Pergaud et al. (2009) in *blue* and proposed in the new formulation in *red*

cloud water content are greatly variable from case to case, and the cloud scheme from formulation to formulation. However, some patterns seem to emerge. Cloud cover is naturally better represented by the cloud schemes than cloud water content (which is a moment of higher order). In the lower part of the cloud, the new formulation is in the range of previous schemes. It tends to underestimate the cloud cover (especially for the BOMEX case), and this should be looked into in the future studies. In the middle part of the cloud, the shape of cloud cover and cloud fraction profiles are better for the schemes describing the second mode. The single Gaussian cloud schemes (Lenderink and Siebesma 2000, Chaboureau and Bechtold 2005) present a spiky shape, with too rapid a decrease above the maximum for both quantities.

The vertical extension of the clouds is underestimated by the single Gaussian cloud schemes. For the cloud scheme, Pergaud et al. (2009), while supposing to describe only the convective part, also underestimate the cloud height. Only Jam et al. (2010) and this study allow for reproducing correctly the upper part of the cloud. This new scheme is based on the convective processes themselves, and is general enough to be used by any model.

6 Conclusions

In both NWP models (even those with kilometre resolutions) and climate models, it is still necessary to correctly parametrize the horizontal subgrid variability of PBL clouds. This study provides a preliminary step in developing a new cloud scheme to improve the representation of PBL clouds.

The search for the best distribution to use in a statistical cloud scheme is based on the statistical analysis of four LES of classical warm PBL cases (cumulus and stratocumulus)

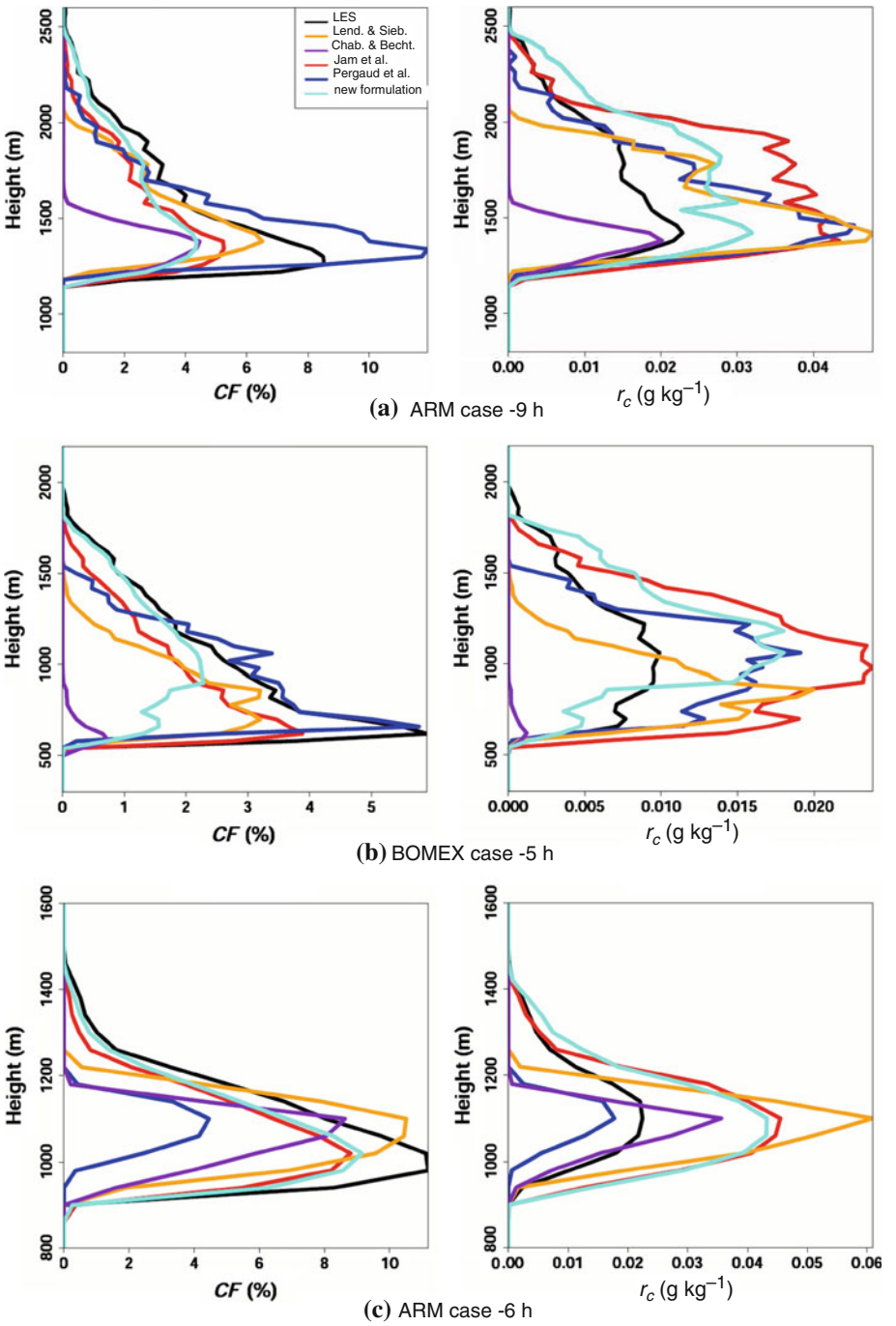


Fig. 12 Vertical profiles of the cloud fraction (*left column*) and the mean cloud water content (*right column*) for **a** the ARM case at 9 h, **b** the BOMEX case, **c** the ARM case at 6 h, the time of formation of cumulus clouds. The reference values obtained from the LES distributions are in *black*, the new formulation is in *light blue* and other distributions from the literature (see text) are in different *colours* (see legend)

performed with the research model Meso–NH. The simulated horizontal cloud variability inside the LES domain is summarized by a LES distribution, which may be fitted to theoretical distributions constrained by the first statistical moments of the LES distribution.

Our analysis shows that, for PBL clouds, the total water variability alone is not the best predictor to correctly diagnose the cloud fraction and the mean cloud water content. The saturation deficit, which combines the temperature variability and the total water variability, has to be taken into account. Nevertheless, when considering a convective and a non-convective domains separately, it is sufficient to consider only the total water variability as most of the temperature variability is mainly accounted for by the top-hat representation (convective and non-convective distinction).

LES distributions of the saturation deficit often show an asymmetric bell-shaped curve with varying skewness depending on the cloud type and also the level and the phase of the cloud life cycle. Most distributions show a primary main mode and a second mode, which may evolve into a long flat tail. This second mode plays an important role in the cloud field computation. In the cases of sparse cumuli, the main mode corresponds to non-cloudy points in the LES. The narrow cloud updrafts are described by the tail on the cloudy side of the distribution. For a fractional stratocumulus cloud, the first mode is often located on the dry side of the distribution and the cloudy second mode depends on the updraft/downdraft activity inside the cloud layer. It reflects, in particular, the intrusion of dry air from the cloud layer top. For a stationary stratocumulus, the model grid is saturated on average and the cloud fraction is nearly 100%.

Several theoretical distributions commonly used in the literature have been tested to find the best fit for the LES distributions. The "Assumed PDF method" was used for determining the parameters of each distribution constrained by the first two or three statistical moments predicted from LES data. All unimodal distributions, even the beta law with a controlled skewness, are unable to correctly represent both the primary mode and the important second mode, especially for cumulus clouds. They would not give a unified approach valid for all PBL cloud types. A double Gaussian distribution with two modes and based on a linear combination of two simple Gaussian laws provides a much better description for all types of PBL clouds. It corrects the large underestimation of the cloud fraction and the mean cloud water content given by unimodal distributions in the case of sparse cumuli and always gives a better estimate of the heights of the base and the top of the cloud layer. The second mode is well positioned with correct amplitude and can become a long flat tail. This study is simplified compared to Lewellen and Yoh (1993) or Golaz et al. (2002a) as we only consider a single statistical variable, the saturation deficit, and only five parameters are necessary to define the double Gaussian distribution. Moreover, it has been verified that all conclusions based on the statistical analysis for LES configuration are still valid for a coarser vertical description of the cloud layer. Eventually, it appears that the use of a bimodal distribution is not essential for the horizontal description of homogeneous stratocumulus clouds but their correct representation would require to take into account the vertical subgrid variability.

In the previous section, the results have been discussed, and some guidance has been indicated for the development of a cloud scheme usable by any shallow-convection model. The proposition is to use a combination of two single independent Gaussian distributions, one describing the convective domain and the other the non-convective domain. A formulation for the standard deviation of the thermals and their coverage has been proposed using outputs from shallow convection schemes (see Eqs. 9 and 10).

The next step in the development of a PBL cloud scheme is to confirm those results on other cases. The case of RICO on which an intercomparison has been carried out by van Zanten et al. (2011) could be considered to validate the new formulations. Then, more

complicated cases could be analyzed such as transition from stratocumulus to shallow cumulus, the dissipation of clouds and the transition from shallow to deep convection. Nevertheless, in the framework of NWP resolving more and more deep convection, the extension of this study to deep convection will not be considered. For all these transitions, it should also be investigated whether a diagnostic scheme can handle those situations or whether a prognostic scheme Tompkins (2002) is mandatory.

A Appendix: Computation of the Theoretical Distributions

In order to fit to the five unimodal distributions, estimators (noted with $\widehat{\ }$) of their parameters are computed with the "Method of moments". It consists of resolving an equation system built from the empirical expressions of statistical moments as a function of the distribution parameters. It is more or less complex according to the number of unknown parameters and to the PDF family considered. For this appendix, the first three statistical moments are, respectively, noted as μ, σ and ζ.

A.1 The Simple Gaussian Distribution

The two distribution parameters are easily computed to fit the LES PDFs because they are equal to the first two predicted statistical moments: $\widehat{\mu} = \mu$ and $\widehat{\sigma} = \sigma$.

A.2 The Triangular Distribution

As for the Gaussian law, the two distribution parameters of the triangular distribution are deduced directly from the first two statistical moments : $\widehat{\mu} = \mu$ and $\widehat{\sigma} = \sigma$.

A.3 The Gamma Distribution

If k and δ are, respectively, the positive shape parameter and the positive scale parameter of a gamma distribution, the parameter estimators are deduced from the expressions of μ and σ by: $\widehat{k} = \mu^2/\sigma^2$ and $\widehat{\delta} = \mu/\sigma^2$.

A.4 The Log-normal Distribution

A positive statistical variable X follows a log-normal distribution if $\ln(X)$ follows a Gaussian law. In order to fit to the log-normal distribution, we obtain the two parameter estimators $\widehat{\eta}$ and $\widehat{\delta}$ from the expressions of μ and σ:

$$\widehat{\eta} = \ln(\mu) - \frac{1}{2}\ln\left(1 + \frac{\sigma^2}{\mu^2}\right) = \overline{\ln X}, \tag{11}$$

$$\widehat{\delta} = \left(\ln\left(1 + \frac{\sigma^2}{\mu^2}\right)\right)^{\frac{1}{2}} = \sqrt{\overline{(\ln X)'^2}}. \tag{12}$$

A.5 The Beta Distribution

If a and b are the bounds and p and q are the positive shape parameters of a beta distribution, then two methods are available to compute them to fit the LES distributions. Note that, in this study, a bell-shaped curve is considered so $p \geq 2$ and $q \geq 2$ are imposed.

– If the parameters are constrained by μ and σ, then $\widehat{a}$ and $\widehat{b}$ are respectively, equal to the minimum and maximum values of the statistical variable, and the shape parameters are computed from their expressions (distribution called β_1):

$$\widehat{p} = \frac{\mu^2\,(1-\mu)}{\sigma^2} - \mu, \tag{13}$$

$$\widehat{q} = \frac{\mu\,(1-\mu)^2}{\sigma^2} - (1-\mu), \tag{14}$$

– If the parameters are constrained by μ, σ and ζ, and $\widehat{a}$ is equal to the minimum value of the statistical variable and $\widehat{b}$, and the two shape parameters are computed from their expressions (distribution called β_2):

$$\widehat{p} = \frac{2(\mu-\widehat{a})\,[\sigma^2 - (\mu-\widehat{a})^2 - \zeta\sigma(\mu-\widehat{a})]}{\zeta\sigma[(\mu-\widehat{a})^2-\sigma^2] - 4(\mu-\widehat{a})\sigma^2}, \tag{15}$$

$$\widehat{q} = \frac{\widehat{p}(\widehat{p}+1)\,\sigma^2}{(\mu-\widehat{a})^2 - \widehat{p}\sigma^2}, \tag{16}$$

$$\widehat{b} = \widehat{a} + \frac{\widehat{p}+\widehat{q}}{\widehat{p}}\,(\mu-\widehat{a}). \tag{17}$$

It is noted that, in a few cases, there is no mathematical solution for this computation which is compatible with the physical constraints imposed by this study. In that case, the β_2 PDF cannot be drawn and is taken equal to the previous β_1 PDF that keeps only μ and σ.

A.6 The Double Gaussian Distribution

The double Gaussian distribution is equal to a linear combination of two individual Gaussian distributions $G = aG_1 + (1-a)G_2$ where G_1 and G_2 are the simple Gaussian PDFs, a and $(1-a)$ are the relative weights of each distribution and μ_1, μ_2, σ_1 and σ_2 are the means and the standard deviations of G_1 and G_2. The computation of these parameters is based on the "Expectation-Maximization method" (EM hereafter) (see Dempster et al. 1997; Hogg and Craig 2005 and McLachlan and Krishnan 1997). It is an analytical iterative method that consists of maximizing the likelihood of having the desired bimodal distribution corresponding to the known sample of LES data. It requires the knowledge of first-guess values for the distribution parameters to start the iterative computations, built as follows:

– We suppose that $a = 0.5$, i.e. the two simple Gaussian distributions have the same weight.
– the two means μ_1 and μ_2 are approximately equal to the two values of s corresponding to the local maxima
– we use the definition of a Gaussian standard deviation to determine the value of σ_1 and σ_2 for each mode: it is proportional to the width, H, of the curve at a height equal to the half the maximum of the mode considered ($H = 2.3548\,\sigma_s$).

To obtain a successful EM algorithm, assumptions are made to ensure its convergence and adapt it to the irregularity of the LES distributions, e.g. minimal values for standard deviations. Then, after some sensitivity tests for the number of iterations, it seems that 12 iterations are enough to correctly fit the LES distributions. The five parameters so computed verify the following expressions:

$$\mu = a\mu_1 + (1-a)\mu_2, \tag{18}$$

$$\sigma^2 = a\sigma_1^2 + (1-a)\sigma_2^2 + a(1-a)(\mu_2-\mu_1)^2. \tag{19}$$

References

Bechtold P, Cuijpers JWM, Mascart P, Trouilhet P (1995) Modeling of trade wind cumuli with a low-order turbulence model: toward a unified description of Cu and Sc clouds in meteorological models. J Atmos Sci 52:455–463

Bony S, Emanuel KA (2001) A parametrization of the cloudiness associated with cumulus convection; evaluation using Toga Coare data. J Atmos Sci 58:3158–3183

Bougeault P (1981) Modeling the trade-wind cumulus boundary layer. Part 1: testing the ensemble cloud relations against numerical data. J Atmos Sci 38:2414–2428

Bougeault P (1982) Cloud-ensemble relations based on the gamma probability distribution for the higher-order models of the planetary boundary layer. J Atmos Sci 39:2691–2700

Brown AR, Cederwall RT, Chlond A, Duynkerke PG, Golaz JC, Khairoutdinov M, Lewellen DC, Lock AP, Macvean MK, Moeng CH, Neggers RAJ, Siebesma AP, Stevens B (2002) Large-eddy simulation of the diurnal cycle of shallow cumulus convection over land. Q J Roy Meteorol Soc 128:1075–1093

Chaboureau JP, Bechtold P (2002) A simple cloud parameterization derived from cloud resolving model data: diagnostic and prognostic applications. J Atmos Sci 59:2362–2372

Chaboureau JP, Bechtold P (2005) Statistical representation of clouds in a regional model and the impact on the diurnal cycle of convection during tropical convection, cirrus and nitrogen oxides (troccinox). J Geophys Res 110:D17103

Chosson F, Brenguier JL, Schuller L (2007) Entrainment-mixing and radiative transfer simulation in boundary layer clouds. J Atmos Sci 64:2670–2682

Couvreux F, Hourdin F, Rio C (2010) Resolved versus parametrized boundary-layer plumes. part i: A parametrization-oriented conditional sampling in large-eddy simulations. Boundary-Layer Meterol 134:441–458

Cuijpers JWM, Bechtold P (1995) A simple parameterization of cloud water related variables for use in boundary layer models. J Atmos Sci (Notes and correspondence) 52:2486–2490

Cuxart J, Bougeault P, Redelsperger JL (2000) A turbulence scheme allowing for mesoscale and large-eddy simulations. Q J Roy Meteorol Soc 126:1–30

Dempster A, Laird N, Rubin D (1997) Maximum likelihood from incomplete data via the em algorithm. J Roy Stat Soc Ser B 39:1–38

Golaz JC, Larson VE, Cotton W (2002a) A PDF-based model for boundary layer clouds. Part I: method and model description. J Atmos Sci 59:3540–3551

Golaz JC, Larson VE, Cotton W (2002b) A PDF-based model for boundary layer clouds. Part II: model results. J Atmos Sci 59:3552–3571

Hogg RV, Craig AI (2005) Introduction to mathematical statistics. Pearson Education, Upper Saddle River

Holland JZ, Rasmusson EM (1973) Measurement of atmospheric mass, energy and momentum budgets over a 500-kilometer square of tropical ocean. Mon Weather Rev 101:44–55

Jam A, Hourdin F, Rio C, Couvreux F (2010) Vers un modele de nuage pour la representation des cumulus. In: Preprints of Ateliers de Modelisation de l'Atmosphere, Annual meeting, Meteo-France, Toulouse

Lafore JP, Stein J, Asencio N, Bougeault P, Ducrocq V, Duron J, Fischer C, Héreil P, Mascart P, Masson V, Pinty JP, Redelsperger JL, Richard E, Vilà-Guerau de Arellano J (1998) The meso-NH atmospheric simulation system. Part 1: adiabatic formulation and control simulations. Ann Geophys 16:90–109

Lappen CL (1999) The unification of mass flux and higher-order closure in the simulation of boundary layer turbulence. PhD thesis, Colorado State University

Larson VE, Wood R, Field PR, Golaz JC, Vonder Haar TH, Cotton WR (2001) Small-scale and mesoscale variability of scalars in cloudy boundary layers: one-dimensional probability density functions. J Atmos Sci 58:1978–1994

Larson VE, Golaz JC, Cotton WR (2002) Small-scale and mesoscale variability in cloudy boundary layers: Joint probability density functions. J Atmos Sci 59:3519–3539

Lenderink G, Siebesma P (2000) Combining the mass-flux approach with a statistical cloud scheme. Proceedings of the 14th symposium on Boundary Layers and Turbulence. Aspen, Co, USA, pp 66–69

Lewellen WS, Yoh S (1993) Binormal model of ensemble partial cloudiness. J Atmos Sci 50:1228–1237

McLachlan G, Krishnan T (1997) The EM algorithm and extensions. Wiley series in probability and statistics. Wiley, New York, 274 pp

Mellor GL (1977) The gaussian cloud model relations. J Atmos Sci 34:356–358

Neggers RAJ (2009) A dual mass flux framework for boundary layer convection. Part II: clouds. J Atmos Sci 66:1489–1506

Pergaud J, Masson V, Malardel S, Couvreux F (2009) A parameterization of dry thermals and shallow cumuli for mesoscale numerical weather prediction. Boundary-Layer Meteorol 132:83–106

Raes F, Bates T, McGovern F, Van Liedekerke M (2000) The 2nd aerosol characterization experiment (ACE2): general overview and main results. Tellus 52B:111–125

Siebesma AP, Bretherton CS, Brown A, Chlond A, Cuxart J, Duynkerke PG, Jiang H, Khairoutdinov M, Lewellen D, Moeng CH, Sanchez E, Stevens B, Stevens DE (2003) A large eddy simulation intercomparison study of shallow cumulus convection. J Atmos Sci 60:1201–1219

Smith RNB (1990) A scheme for predicting layer clouds and their water content in a general circulation model. Q J Roy Meteorol Soc 116:435–460

Sommeria G, Deardorff JW (1977) Subgrid-scale condensation in models of nonprecipitating clouds. J Atmos Sci 34:344–355

Stephens G, Vane D, Boain R, Mace G, Sassen K, Wang Z, Illingworth A, O'Connor E, Rossow W, Durden S, Miller S, Austin R, Benedetti A, Mitrescu C, The CloudSat Science Team (2002) The cloudsat mission and the A-train: a new dimension of space-based observations of clouds and precipitation. Bull Am Meteorol Soc 83:1771–1790

Tompkins AM (2002) A prognostic parameterization for the subgrid-scale variability of water vapor and clouds in large-scale models and its use to diagnose cloud cover. J Atmos Sci 59:1917–1942

Tompkins AM (2003) Impact of temperature and humidity variability on cloud cover assessed using aircraft data. Q J Roy Meteorol Soc 129:2151–2170

van Zanten MC, Stevens B, Nuijens L, Siebesma AP, Ackerman AS, Burnet F, Cheng A, Couvreux F, Jiang H, Khairoutdinov M, Kogan Y, Lewellen C, Mechem D, Nakamura K, Noda A, Shipway BJ, Slawinska J, Wang S, Wyszogrodzki A (2011) Controls on precipitation and cloudiness in simulations of trade-wind cumulus as observed during RICO. J Adv Model Earth Syst 3 (in press)