Process-oriented stochastic perturbations applied to the parametrization of turbulence and shallow convection for ensemble prediction

Axelle Fleury, François Bouttier, Fleur Couvreux

Process-oriented stochastic perturbations applied to the parametrization of turbulence and shallow convection for ensemble prediction.

Axelle Fleury[1*], François Bouttier[1], Fleur Couvreux[1]

[1] CNRM, University of Toulouse, Météo-France, CNRS, Toulouse, France

* Correspondence to: Axelle Fleury, CNRM, Météo-France, 42 Av. Gaspard Coriolis, 31057 Toulouse, France. E-mail: axelle.fleury@meteo.fr

Key-words: coarse-graining, ensemble weather prediction, large-eddy simulations, model uncertainty, shallow convection, stochastic parametrization, turbulence.

Acknowledgments: This study was financed by the French Government through Météo-France, CNRS and the University of Toulouse. It benefited from the LEFE/MANU IPSIPE research grant of CNRS/INSU.

Weather forecasting nowadays often requires some estimation of uncertainties associated with the output of meteorological models, in order to better inform decision making, especially in the context of intense weather events. Ensemble prediction systems provide such information through sets of possible scenarios which are designed to represent various uncertainty sources, including model uncertainties. A wide variety of methods have been proposed to estimate model uncertainties, among which perturbation methods targeting uncertain processes are a promising research field. In this study, we focus on the representation of small-scale variability by process-oriented perturbation schemes applied to two key physical processes, namely turbulence and shallow convection. The perturbations are applied to a single-column version of the convection-permitting AROME model, in three idealized boundary-layer cases. Large-eddy simulations (LESs) of the same cases serve as a reference for the subgrid variability that has to be represented, and the results are also compared to those given by the Stochastically Perturbed Parametrization Tendencies (SPPT) method, which is a method commonly used by weather forecast centers to represent model uncertainty. The spread produced by our process-oriented perturbations of turbulence and shallow convection does not represent all the small-scale variability implied by the LESs for temperature and humidity. However, it is of a similar order of magnitude for the wind, thanks to perturbations generated by the stochastic turbulence scheme. The dispersion is structurally different from what is obtained with SPPT. It is non-negligible in the lower levels, where SPPT perturbations are usually suppressed because of numerical instabilities, indicating a possible complementarity between the schemes.

1. Introduction

Probabilistic forecasts constitute one of the products commonly used to assess the forthcoming weather, and are becoming increasingly valuable in many areas, including the prediction of extreme meteorological events. They result from the concept of ensemble prediction, which has been developed over several decades as a way to estimate the degree of confidence one can have in a forecast. Predictions of weather forecasting systems are

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/qj.4242

indeed necessarily imperfect because of various errors arising, for example, from incomplete knowledge of the initial state of the atmosphere, or from approximations in the modeling of atmospheric processes. They can have a greater or lesser impact depending on the situation, time and location. Instead of a single forecast, an ensemble prediction system provides several forecasts, called members, thanks to random perturbations and therefore provides a continuous, prognostic measure of uncertainty. The random perturbations are designed to represent known error sources, among which those contained in initial and boundary model conditions have been shown to play a major role (Lorenz 1965). In addition, we know that the errors of the model design itself, called "model errors", must be addressed as well in order to account for the entire forecast uncertainty (Buizza et al. 2005, Wilks 2005, Houtekamer 2009). They are currently the target of active research.

One important source of model uncertainty lies in the parametrization of subgrid processes (Palmer et al. 2009), which are all the phenomena whose time and space scales are smaller than those used by the model but need to be parametrized because their effect on the resolved fields is not negligible. Such phenomena can be complex, imperfectly known and thus challenging to represent. Different strategies have been proposed to deal with parametrization uncertainty, like the 'multiphysics' approach, which is based on the idea of combining different sets of existing parametrizations (Stensrud et al. 2000, Charon et al. 2010, Berner et al. 2011, Duda et al. 2014, Jankov et al. 2017), or the broad category of stochastic perturbations, where stochastic terms are introduced in the model, with various purposes and behaviors (Buizza et al. 1999, Lin and Neelin 2002, Shutts 2005, Dorrestijn et al. 2013, Baker et al. 2014, among others). The common feature of these methods is that they can produce, for the same resolved state, a set of parametrized subgrid effects, thus allowing the uncertainty of the parametrizations to be represented. Nevertheless, each method targets the sources of uncertainty with different degrees of precision, introducing noise at various levels of the parametrization schemes.

A commonly used stochastic perturbation method is the Stochastically Perturbed Parametrization Tendencies (SPPT) scheme introduced by Buizza et al. (1999). In this method, a 'bulk' uncertainty coming from the parametrizations is represented by multiplying the total parametrization tendency by a random field, with temporal and spatial correlations. It has proved efficient in producing ensemble spread in several ensemble systems (Palmer et al. 2009, Bouttier et al. 2012, Batté and Doblas-Reyes 2015, Berner et al. 2015), and has the advantage of remaining relatively simple, and keeping the balance between the tendencies of the different parametrization schemes. However, its formulation implies hypotheses that may not always be justified, such as the proportionality of the error to the tendency, the correlation of the errors between variables, and the compensation effects induced by the use of the sum of all the parametrization tendencies.

More recently, new stochastic perturbation schemes have been developed which seek a more physically justified approach. For example, efforts have been undertaken to perturb free input parameters, instead of the parametrization output tendencies (Baker et al. 2014, Christensen et al. 2015, Ollinaho et al. 2017, Lang et al. 2021). These perturbation schemes, called Random Parameters (RP) or Stochastically Perturbed Parametrizations (SPP)

schemes, address the problem of parameter uncertainty inside the parametrization schemes, and have the advantage to act on specific physical processes.

We can identify another category of 'physically-based' perturbation schemes which aims specifically at representing 'random errors' coming from the spatial and temporal heterogeneity of the physical processes taking place in the atmosphere, and their representation in a model with a given grid size. They address the assumption of a deterministic link between resolved state and some subgrid state characteristics, which is made in many parametrization schemes. Indeed, because of the apparent randomness of some fine-scale phenomena, such as turbulence, and the multiple interactions between physical processes, two grid boxes with the same resolved state may not have the exact same subgrid state and feedback from subgrid processes. These fluctuations around an average contribution from subgrid processes may be negligible in the case of low resolutions, where the sample size of the subgrid phenomena present in a model grid box is large. With finer resolutions however, the sample size can become very small (e.g. the number of cumulus clouds in a 1 km^2 grid box) and the deterministic link between the resolved and subgrid state is more clearly violated. These subgrid state fluctuations may have an impact on resolved scales and therefore are likely to be a source of uncertainty (Lin and Neelin 2002). Stochastic schemes have therefore been proposed to represent it.

Some schemes target the variability linked to convective updrafts and downdrafts: they can consist of adding a noise to a variable diagnosed by a deterministic scheme (Lin and Neelin 2000) or of rendering key aspects of the parametrization stochastic: a stochastic trigger for deep convection is studied by Rochetin et al. (2014), Bengtsson et al. (2021) propose to use a cellular automata to generate subgrid plumes and derive stochastic formulations of the entrainment, triggering and mass-flux closure, and Craig and Cohen (2006) and Sakradzija et al. (2015) propose a stochastic mass-flux closure condition for deep and shallow convection respectively, further developed or tested in subsequent studies (Plant and Craig 2008, Groenemeijer et al. 2012, Sakradzija et al. 2016, Sakradzija and Klocke 2018, Machulskaya and Seifert 2019). These last schemes explicitly consider the limited number of clouds inside a model grid box and their heterogeneous characteristics.
Other schemes target the heterogeneity of small-scale eddies represented by turbulence schemes. Kober and Craig (2016) and Hirt et al. (2019) propose to perturb the output tendencies of the turbulence scheme in the COSMO model by adding a noise controlled by subgrid variances diagnosed by the turbulence scheme. Clark et al. (2021) study the representation of eddies as random events following a Poisson distribution inside a bulk model of the convective boundary layer.
For simplicity, these kinds of schemes will be referred to as 'process-oriented' schemes in this paper. They have the advantage of relying on physical arguments and of targeting a source of uncertainty – small-scale variability of physical processes – which is likely to become increasingly important to consider as the resolution of numerical weather prediction (NWP) models increases.

The evaluation of perturbation methods used to produce ensembles can be difficult, depending on the kind of uncertainty they are designed to represent. They can be validated by computing statistical scores on large data sets of real case simulations, but this does not allow one to look in detail at the effects of the perturbations on the model physics, and can

hide contrasting results between different meteorological situations and compensation effects. In this study, we choose to work in an idealized framework using single-column model (SCM) simulations of several idealized cases, compared to large-eddy simulations (LESs). Taking LESs as a reference to investigate boundary-layer situations and evaluate or improve parametrization schemes has already been done in a number of studies (Ayotte et al. 1996, Duynkerke et al. 2004, Cuxart et al. 2006, Nie and Kuang 2012, Lu et al. 2016, Angevine et al. 2018, Couvreux et al. 2021, to cite only a few). The high resolution of LESs allows one to resolve a substantial part of the phenomena parametrized in NWP models, and LESs have the advantage to provide continuous 3D variable fields from which various diagnostics, including process-oriented ones, can be derived and compared to SCM outputs. LESs have also been validated on several occasions against observations and have been shown to realistically represent organized structures and small-scale variability (Neggers et al. 2003, Couvreux et al. 2005, Heus et al. 2009). In our case, we are particularly interested in the information provided on the small-scale variability of variables. By applying, on a LES domain large enough to be several times the size of the SCM grid box, the same homogeneous initial conditions and large-scale forcing as the ones prescribed in the SCM, the variability of the LES fields coarse-grained to the SCM resolution can be taken as the reference for the variability we wish to represent with process-oriented perturbations. Our first question therefore is: are process-oriented perturbation schemes able to represent the small-scale variability of an LES domain?

To be useful in ensemble forecasts, perturbation schemes should be able to produce sufficient ensemble spread. The process-oriented perturbation schemes target one source of uncertainty among several, therefore it may be interesting to compare their results in terms of dispersion with methods that use a more general approach. This is our second question: how does the dispersion produced by process-oriented perturbation schemes compare with the dispersion produced by the SPPT method?

Two process-oriented perturbation schemes are tested in this study in a single-column version of the convection-permitting AROME model. They are based on schemes proposed in the literature and target the representation of turbulent motions, which are known to lack variability in their current parametrization in AROME at resolutions belonging to the grey-zone of turbulence (Honnert et al. 2016). Our perturbation schemes are closely related to Kober and Craig (2016), hereafter referred to as KC16, for the perturbation of turbulence, and we build on the work of Sakradzija et al. (2015, 2016) and Sakradzija and Klocke (2018) to design a very simplified version of their stochastic shallow-convection scheme, based on the analysis of LES fields.

The paper is organized as follow. The process-oriented perturbation schemes applied to turbulence and shallow convection are described in section two. The analysis of LESs used to design the perturbation of shallow convection is presented in section three. Section four documents the dispersion of different ensembles constructed with the new perturbation schemes, evaluated against the dispersion inside the LES domain, and also compared with ensemble spread produced by SPPT. Section five is dedicated to discussion and conclusion.

2. Description of the perturbation schemes and simulation setup

2.1. The mesoscale AROME model

The perturbation schemes tested in this study are applied to the convection-permitting limited area model AROME (Seity et al. 2011) used operationally at Météo-France for weather prediction at kilometric scales. The experiments are performed in a one-dimensional (1D) framework, using a single-column version of the model (AROME-SCM), using 90 atmospheric vertical levels. Prognostic variables are used for temperature, wind, water vapor, five condensed water species and the turbulent kinetic energy (TKE). In the 1D framework, the tendencies produced by the dynamical part of the model are replaced by specified large-scale and surface forcing. The rest of the evolution is due to the parametrization of subgrid processes. In AROME, the parametrization schemes include radiation, surface, microphysics, turbulence and shallow convection. These last two schemes follow the EDMF (Eddy Diffusivity/Mass Flux) approach (Soares et al. 2004) to parametrize turbulent motions: the turbulence scheme is a classical eddy-diffusivity scheme (Cuxart et al. 2000), used to represent unorganized, local turbulence, while the shallow-convection scheme is a mass-flux scheme (Pergaud et al. 2009) used for the representation of non local turbulence in the form of a single convective plume. In this work, we propose to perturb each of these two schemes.

2.2. Implementation of stochastic turbulence in AROME

The KC16 scheme is designed as a situation- and scale-adaptive perturbation scheme that consists of an additive noise applied to the output tendencies of a turbulence scheme (Kober and Craig 2016). Its adaptability comes from the formulation of the noise amplitude which depends on subgrid variances diagnosed by the turbulence scheme and on a measure of the relative size of the eddies compared to the horizontal resolution of the model. The formulation of the scheme is summarized in Equation 1.

$$\left(\frac{\partial \phi}{\partial t}\right)^{\text{stoch}} = \frac{\partial \phi}{\partial t} + \alpha \frac{1}{\Delta t} \frac{l_{\text{eddy}}}{\Delta x_{\text{eff}}} \left(\overline{\phi'^2}\right)^{\frac{1}{2}} \eta. \qquad (1)$$

ϕ is a model prognostic variable, $\partial \phi / \partial t$ is the tendency produced by the deterministic turbulence scheme and other symbols are defined below. The stochastic version of the tendency, $(\partial \phi / \partial t)^{\text{stoch}}$, is obtained by adding a random perturbation composed of a Gaussian random number field η, multiplied by the square root of the subgrid variance of the variable $\left(\overline{\phi'^2}\right)^{1/2}$, a scaling factor depending on time and length scales, and a scaling parameter α. This formulation has similarities with the stochastic perturbation scheme of Clark et al. (2021), where, starting from a simplified boundary layer model and the assumption that turbulent eddies can be considered as random events following a Poisson distribution, they mathematically derive a perturbation of the temperature tendency representing the fluctuations around the mean due to the random presence of eddies.

The implementation of the KC16 scheme in AROME is rather straightforward. The tendencies produced by the turbulence scheme of Cuxart et al. (2000) are perturbed following Equation 1. The key components of the method, the subgrid variances, are available in the scheme of Cuxart et al. (2000), which provides diagnostic equations for them. An example is given for potential temperature in Equation 2.

$$\overline{\theta'^2} = \frac{2}{3}\frac{1}{C_s}\frac{L^2}{C_\theta}\left(\frac{\partial\overline{\theta}}{\partial z}\right)^2 \varphi, \qquad (2)$$

where C_s and C_θ are constants, φ is a stability function and L is the Bougeault-Lacarrère mixing length (Bougeault and Lacarrère 1989). In AROME, the scheme is employed in its one-dimensional version, where only the vertical gradients are taken into account, as shown in Equation 2.

The other components of the perturbation are taken as in Hirt et al. (2019):

- Δt is the characteristic lifetime of an eddy, set to 10 minutes,
- l_{eddy} is the characteristic horizontal size of an eddy, set to 1000 m,
- Δx_{eff} is the effective resolution of the model, here set to$5\Delta x = 5 \times 1.3$ km,

except for the tuning parameter α which here is set to 1.

Since we work in a 1D framework, the random field η is specified as a random scalar, taken from an autoregressive process of order 1 with a time constant set to 10 minutes. This temporal correlation of the noise has been chosen by Kober and Craig (2016) and Hirt et al. (2019) considering the characteristic lifetime of an eddy. The perturbations are added at each time step to the tendency of temperature T, specific humidity q_v and the two components of horizontal wind u and v. The same η is used for the four variables, which means that at a given time step, the perturbations added to T, q_v, u and v tendencies always have the same sign. This may not be ideal and one could think of using different random processes for each variable in order to decorrelate the perturbations. However, in this work we choose to remain close to the scheme original setup and therefore use a unique η. No parametrization tendency is produced for the vertical wind in AROME, hence it is not perturbed as in the original scheme of Kober and Craig (2016). We acknowledge that the interpretation of horizontal wind perturbations in a single-column framework without any interaction with the model dynamics may be limited, but as the results show that KC16 scheme has the ability to produce significant wind perturbations, they are kept in our simulations. A more careful design may however be necessary in future 3D experiments to prevent undesirable behavior of the perturbations such as a fast decay due to the appearance of acoustic modes reported by Hirt et al. (2019).

2.3. Design of a simple stochastic convection scheme

Stochastic convection schemes such as proposed by Plant and Craig (2008) or Sakradzija and Klocke (2018) rely on the generation of a random population of clouds inside each model grid box, constrained by large-scale conditions. In the case of coarse resolutions

studied by Cohen and Craig (2006), it allows one to study statistical fluctuations around a bulk equilibrium state due to the finite size of cloud samples inside model grid boxes. In mesoscale models with kilometric-scale horizontal resolution, the cloud population is small so the scheme is further used to relax the assumption that a statistically representative sample of clouds, well represented by its averaged characteristics, is present in each model grid box (Sakradzija et al. 2015). In addition to generating a cloud population, each cloud is assigned a mass flux randomly chosen according to a statistical distribution, and a lifetime. This enables one to compute the total mass flux on a grid box at a given time step as the sum of the mass fluxes of all the clouds existing in this grid box at this time, which can then be used as a closure condition for the rest of the convection scheme.

The generation of clouds in each model grid box and their tracking over several time steps can be computationally expensive. Machulskaya and Seifert (2019) recently demonstrated that they could reproduce the essential behavior of the stochastic convection schemes through a set of stochastic differential equations (SDEs), and thus reduce the computational cost. In both cases, the key feature of the scheme is the stochastic mass-flux closure given to an otherwise deterministic convection scheme. Here, we choose to try a very simplified approach, where the mass-flux closure is directly perturbed according to a statistical distribution.

The mass-flux scheme of Pergaud et al. (2009), hereafter referred to as PM09, is used in AROME to represent dry and moist convective thermals in the boundary layer. A single ascending plume represents the bulk effects of all the updrafts within a grid box. It is initialized at the surface thanks to closure conditions on the mass flux, vertical velocity and thermodynamic characteristics. In order to try our simplified version of a stochastic convection scheme, we thus need to find an appropriate distribution of the mass flux at the surface. We cannot directly use the empirical distributions found by Sakradzija et al. (2015) as a result of their cloud generation scheme, for they are located at the cloud base where their convection scheme has its closure condition. Therefore, in this study we choose to rely on the analysis of LESs to determine the likely distributions of the surface mass fluxes. The LES analysis and the final formulation of the stochastic convection scheme are presented in section three.

2.4. The SPPT scheme

SPPT is the current perturbation scheme used to represent model errors for ensemble prediction at Météo-France. Its implementation in the AROME-EPS ensemble system is described in Bouttier et al. (2012). In this study, we use a configuration of the SPPT scheme slightly modified from the operational one. In the following, every mention of SPPT will refer to our modified configuration. It has the same temporal correlation of the noise, set to 6 hours, and the same standard deviation, set to 0.3, as the current operational configuration. However, the tapering function used to damp the perturbations inside the boundary layer in the operational configuration is deactivated, as we are mainly interested in perturbations of boundary-layer processes in this study. In addition, the perturbations are not added to the total net physical tendencies but to the sum of the tendencies produced by the turbulence and shallow-convection schemes only. This ensures a fair comparison between the ensemble spread produced by SPPT and the ensemble spread

produced by the two process-oriented schemes previously described, which are only perturbing these two parametrizations.

2.5. Cases

Three well-documented cases are examined, representing different boundary-layer situations:

- the ARMCu case is an idealization of observations from the Atmospheric Radiation Measurement (ARM) campaign carried out on June 21, 1997 in the Great Plains region of the United States (Brown et al. 2002). It features the development of shallow non-precipitating cumulus clouds at the top of a continental convective boundary layer ;
- the BOMEX case features a marine boundary layer topped by cumulus clouds under steady conditions in the trade-wind region (east of Barbados). This case is described in Siebesma et al. (2003) and is based on observations from the Barbados Oceanographic and Meteorological Experiment (Holland and Rasmusson 1973) ;
- the FIRE case represents the diurnal cycle of a stratocumulus-topped marine boundary layer. It is taken from the European Projection Cloud Systems in Climate Models (EUROCS) project and based on observations made off the coast of California in July 1987 (Albrecht et al. 1988, Duynkerke et al., 2004).

For each case, both SCM simulations and LESs are used, whose characteristics are summarized in Table 1. The LESs are made with the research model Méso-NH (Lac et al. 2018) and use the same initial profiles and large-scale forcing as the SCM simulations. Large-scale forcing include advective and, for some cases, radiation forcing, given directly to the model as tendency terms. A large-scale subsidence can also be prescribed (BOMEX, FIRE) through a large-scale vertical velocity, which is used by the model together with the vertical gradients of the variables to compute the corresponding tendency term. For each case, the geostrophic wind is also specified.

For the FIRE case, the large-scale forcing and the sea surface temperature are time-independent and the diurnal cycle is entirely controlled by the cyclic solar radiations provided by the radiation scheme. Hence, this case can be run for an arbitrary long time with AROME-SCM. After 48 hours of simulation, the cloud evolution simulated by AROME becomes perfectly regular between successive 24 hour-intervals. We thus choose to systematically drop the first 48 hours in our simulations of the FIRE case. All the perturbations applied to the model start at 48h. For the sake of readability, the first 48 hours are never plotted in the figures and the time axis starts at 0h.

The evolution of the domain-averaged cloud liquid water in the LES fields, as well as the cloud liquid water inside the column simulated by AROME-SCM are presented in Figure 1.

We can note that AROME-SCM over-estimates the cloud water for BOMEX with regard to the LES, and that the FIRE stratocumulus cloud top and base heights are too low. Otherwise, AROME-SCM and averaged LES fields are in relatively good agreement. These cases have already been used in a number of studies comparing SCM and LESs (Sušelj et al.

2013, De Rooy and Siebesma 2008, Huang et al. 2013, Rio and Hourdin 2008, among others), including the work of Pergaud et al. (2009) on the mass-flux scheme currently used in AROME for the parametrization of shallow convection.

For each case, four ensembles of 100 members are produced using AROME-SCM, with the different perturbation methods previously described: KC16 (perturbation of turbulence) only, stochastic convection only, both KC16 and stochastic convection, and SPPT. No initial or boundary perturbations are applied in these ensembles.

The small-scale variability inside the LES fields is evaluated with a coarse-graining procedure: small subdomains of 1.3km^2 are defined, and the variables of interest are averaged over each subdomain. The dispersion between subdomains provides an indication on the subgrid variability we seek to represent.

3. Calibration of the stochastic convection method

As indicated in section 2.3, we wish to test a simple stochastic convection scheme where the surface mass-flux closure of the PM09 scheme is perturbed according to a statistical distribution. To determine the appropriate distribution to use for the three boundary-layer cases described in section 2.5, we choose to rely on the corresponding LESs. Our strategy is the following: for each LES field we calculate a coarse-grained mass-flux field by dividing the LES domain into small subdomains of the same size as the AROME grid box (1.3 km^2), and estimating the total mass flux of each subdomain at each vertical level. The estimation of the mass flux is achieved in the following way:

1. updrafts objects are identified in the LES fields, according to the method of Couvreux et al. (2010) based on the emission of a passive tracer at the surface, with limited lifespan. Points are identified as part of an updraft when they exceed a certain threshold of positive tracer concentration anomaly relative to the field average, together with a positive vertical velocity ;

2. the fractional area of each updraft is computed, as well as its average vertical velocity (at each vertical level), and both quantities are multiplied in order to get an approximation of the updraft mass flux (density variations are neglected) ;

3. on each subdomain the total mass flux is obtained by summing over all the updrafts it contains.

These mass-flux values can then be summarized into histograms that inform us on the shape of the distribution, which is done here in Figure 2, for three vertical levels relative to the cloud base height.

The distribution of the mass flux is found to be highly dependent on the height. It has an exponential shape at cloud base, which is consistent with the distribution obtained by Sakradzija et al. (2015) for a resolution of 1.6 km. Near the surface, Gaussian distribution functions fit reasonably well the results.

According to these results, we want our stochastic convection scheme to produce surface mass-flux values following a Gaussian distribution. In order to avoid introducing bias into the model because of systematic differences between AROME-SCM and the LESs, the surface mass-flux values are not sampled directly from the LES distribution, but a Gaussian noise η is applied to the AROME surface mass-flux initialization, following Equation 3. The standard deviation of the noise is set to $\frac{\sigma_{\mathrm{LES}}}{\mu_{\mathrm{LES}}}$, where σ_{LES} is the standard deviation and μ_{LES} the experimental mean of the LES coarse-grained mass-flux distributions shown in Figure 2. Their values are given for each case in the appendix. The noise distribution has a zero mean so that the stochastic surface mass flux $M_{\mathrm{surf}}^{\mathrm{stoch}}$ distribution should be centered around the deterministic surface mass flux M_{surf} diagnosed by the shallow-convection scheme.

$$M_{\mathrm{surf}}^{\mathrm{stoch}} = M_{\mathrm{surf}} * (1+\eta), \text{ with } \eta \sim N\left(0, \frac{\sigma_{\mathrm{LES}}}{\mu_{\mathrm{LES}}}\right) \quad (3)$$

Here, η is again an autoregressive process of order 1. By analogy with the KC16 scheme, we choose the time constant to be approximately the lifetime of a convective updraft. In our experiments, it is set to 40 minutes, close to the 45 minutes chosen by Plant and Craig (2008). However, this lifetime may be overestimated in the case of shallow convection, and other values around 20 minutes could be considered as well (Zhao and Austin 2005, Sakradzija et al. 2015).

4. Evaluation of the perturbation schemes

In this section, the SCM ensembles produced with the process-oriented perturbation schemes (KC16 for the perturbation of turbulence and stochastic convection) are compared with the coarse-grained LES fields to examine whether or not their dispersion is realistic. They are also compared with ensembles produced with SPPT, and we study how the perturbation methods differ from each other. We emphasize that no initial or boundary perturbations are used in our study, therefore the spread between ensemble members is solely due to the perturbations applied to the parametrization schemes. The control simulation that is referred to in the following is, for each case, one AROME-SCM simulation without any perturbations (the deterministic simulation).

4.1. KC16 turbulence perturbation

First, we examine the impact of the KC16 scheme. We start by looking at the amplitude and location of the perturbations added to the tendencies (second term on the right hand side of Equation 1), and then we look at the ensemble spread induced by these perturbations.

4.1.1. A limited impact of KC16 on temperature and humidity

The KC16 perturbation amplitude is controlled by the square-root of subgrid variances, which are given for temperature in Figure 3, for the three cases.

In the ARMCu case, most of the variance is located in the first levels above the ground, which is linked to strong heat fluxes from the continental surface. The variance close to the

surface is lower for BOMEX and FIRE which are oceanic cases. The profiles are otherwise similar between ARMCu and BOMEX, with significant variances in the cloud layer, except that the ARMCu cloud extends higher. The FIRE profile is very different, with high variances around the stratocumulus top due to the radiative cooling process occurring there.

To see what the perturbations controlled by these variances look like, we present in Figure 4 instantaneous vertical profiles of their standard deviation. In the same graphs are shown the standard deviations of SPPT perturbations, and the temperature absolute tendency of the control simulation. We recall that the tendency here is the sum of the tendencies produced by the turbulence and the shallow-convection schemes, instead of the total net tendency.

As shown by the red and black profiles, the standard deviation of SPPT perturbations follows the tendency profile, which is expected since the perturbations are multiplicative. The standard deviation is about 30% of the tendency, consistent with the settings. KC16 perturbations on the contrary are not linked to the tendency values and their profiles are logically close to those of the subgrid standard deviation shown in Figure 3. The perturbation values are in general lower than the SPPT perturbations, and rather small compared to the tendency.

The temperature spread in the 100-member ensembles obtained with these perturbations is illustrated in Figure 5. It is quantified by the tenth to ninetieth percentile ensemble range, which will be used throughout this study as a measure of the ensemble spread. In Figures 5b and 5c we can see that the vertical structure of the SPPT ensemble spread globally follows that of the perturbations shown in Figures 4a and 4b. For the KC16 ensemble, where the perturbation peaks are more localized, the perturbations seem to have spread vertically. The local perturbation peak at the surface visible in Figures 4a and 4b does not result in high temperature spread at this level ; the spread is rather homogeneous throughout all the subcloud layer suggesting a propagation of the perturbation by the turbulent transport in this layer. Furthermore, we observe temperature spread up to 3000 m in ARMCu after 9 hours (Figure 5b) and at 2000 m for BOMEX after 14 hours (Figure 5c), whereas no perturbations have been introduced at this level at this time (Figures 4a and 4b). This also shows that the perturbations can propagate upward and even be amplified in the cloud as the peak around 2000 m for BOMEX suggests. The results of FIRE (Figure 5d) also indicate that the perturbations propagate but no amplification can be seen.

Compared to the variability of the coarse-grained LES fields, the SPPT ensemble spread is generally too large and KC16 ensemble spread too low. For the cumulus cases, SPPT ensemble spread is especially large in the subcloud layer, whereas the LES variability there is smaller than above, probably because of turbulent mixing which tends to homogenize the layer. The KC16 ensemble spread in this layer is closer to the LESs. In the cloud layer however, the KC16 method does not seem to be able to produce as much spread as observed in the LESs, contrary to SPPT.

For FIRE, both SPPT and KC16 produce very similar temperature dispersion during both night (Figure 5d) and day (not shown), of the same order of magnitude as the LES dispersion, except at the top of the clouds. The LES dispersion has a strong peak there not at all captured by any of our ensembles. This is probably because radiation and condensation processes play a critical role at this level, and these schemes should be perturbed as well in order to get significant dispersion.

Similar observations can be made for the specific humidity (Figure 6). The main difference is that specific humidity tendencies are low in the subcloud layer, therefore the SPPT ensemble spread is reduced at this level contrary to the previous results on temperature. Otherwise the SPPT ensemble is over-dispersive in the cumulus clouds compared to the LES variability, whereas KC16 ensemble tends to be under-dispersive. The humidity dispersion profile for FIRE is homogeneous throughout the first 500 m, similarly to temperature.

4.1.2. Significant wind spread produced by KC16

Contrary to the thermodynamic variables, we observe important wind disturbance with KC16. The stochastic term added to the tendency with KC16 method is large (not shown), because of large wind subgrid standard deviation values. This leads to the high zonal wind spread in the KC16 ensembles seen in Figure 7 (similar results are obtained for the meridional wind), which can sometimes exceed that of the SPPT ensembles. We can also note that the KC16 perturbations, which are added mainly in the subcloud layer (not shown), slightly propagate vertically but do not amplify in the cloud layer, which suggests that the amplification observed for temperature and humidity resulted from interaction with the microphysics.

To summarize, the KC16 method seems to produce limited temperature and humidity spread, especially in the case of the continental cumulus ARMCu. On the other hand, it can produce significant wind spread, of the same order of magnitude as the LES variability in the low levels. The SPPT method on the contrary produces high temperature and humidity dispersion in the cumulus cases, but it is often over-estimated compared to the LES variability. The dispersion obtained for the stratocumulus is more similar between KC16 and SPPT, except for the wind, for which KC16 method leads to greater spread.

The low spread observed in the cloud layer when using KC16 method is a motivation for investigating other perturbation methods targeting different physical processes more likely to bring perturbations in clouds, such as shallow convection.

4.2. Impact of the stochastic convection scheme

Ensembles produced with the stochastic convection scheme are now examined. We first investigate the physical consistency of our scheme by examining the parametrized updraft characteristics compared to the LES updrafts, and then look at the spread of SCM ensembles.

4.2.1. Impact of the scheme on the updraft characteristics

The mass-flux profile of the SCM ensemble made with the stochastic convection scheme is presented in Figure 8a for the ARMCu case.

The range of mass-flux values in the ensemble is large below 1000 m, but decreases sharply toward the top of the profile suggesting that all the members of our ensemble have updrafts that reach similar altitude, but with various strengths inside the first 1000 m. This is linked to the equation governing the mass flux in the shallow convection scheme:

$$\frac{1}{M_{\mathrm{up}}}\frac{\partial M_{\mathrm{up}}}{\partial z} = \epsilon - \delta, \quad (4)$$

where M_{up} is the mass flux of the updraft, ϵ the entrainment rate and δ the detrainment rate. The exponential form of the solution of this equation and the evolution of the sign of $(\epsilon - \delta)$ along the vertical (Figure 8b) imply that any perturbation at the surface will be amplified in the middle part of the updraft and will shrink again toward the updraft top. Thus, to have additional spread around the updraft top, modifying the surface mass flux is not enough, the entrainment and detrainment rate should vary as well. We see very little spread in the entrainment and detrainment values of our SCM ensemble (Figure 8b), indicating that the perturbations of the mass-flux closure do not have a strong feedback on the entrainment and detrainment rates. This suggests the need for further research on the perturbation of these two variables, which are known to be difficult to accurately model (Rio et al 2019).

Figure 8a also shows that the mass-flux values of the SCM ensemble mean, as well as that of the control simulation, are strongly biased with respect to the LES. This over-estimation of the mass flux was observed by Honnert et al. (2016) when using PM09 shallow-convection scheme at resolutions belonging to the grey-zone of convective boundary-layer thermals. The use of the stochastic mass-flux closure does not change the average behavior of PM09 scheme, but it enables one to explore more values belonging to the LES range. The distribution of the values, however, is not entirely satisfactory. Having been perturbed by a Gaussian perturbation at the surface, the mass-flux values keep a Gaussian-like distribution at higher levels, whereas in the LES they evolve toward an exponential distribution (see Figure 2). Similar observations can be made in BOMEX and FIRE (not shown).

4.2.2. Impact of the scheme on the main variables tendencies

The ensemble spread generated by the stochastic convection scheme is of the same order of magnitude (ARMCu) or lower (BOMEX, FIRE) than the spread generated by the KC16 scheme for humidity (Figure 9) and temperature (not shown). The vertical structure however can be rather different. In the FIRE case for example, there is a spread peak for specific humidity at the top of the stratocumulus, not observed in the KC16 ensemble spread profile.

As for the wind, the spread range with the stochastic convection scheme does not exceed half that of the KC16 ensemble, and is negligible in the FIRE case. Its vertical structure is also quite different (Figures 9d-f): stochastic convection has an impact on the wind mainly

near the ground and at cloud base, whereas KC16 perturbations act on all the mixed layer, with a maximum impact in the middle of the subcloud layer.

4.2.3. Impact of the scheme on cloud water

Although the impact of the stochastic convection scheme on the resolved temperature, humidity and wind is not greater than that of KC16 method, there is a strong effect on the cloud liquid water. This is due to the parametrization of subgrid condensation and subgrid cloud in AROME which includes a direct contribution from the mass-flux scheme, of the form:

$$r_c = r_{c_{\mathrm{up}}} C_{\mathrm{cf}} a_{\mathrm{up}}, \quad (5)$$

where $r_{c_{\mathrm{up}}}$ is the updraft liquid water mixing ratio, a_{up} the updraft area and C_{cf} a constant set to 2.5. The updraft area is directly linked to the updraft mass flux (along with the updraft vertical velocity), so that a_{up} in our ensemble can vary by more than 100%. This has a significant impact on the cloud water as can be seen in Figure 10 representing the vertical profiles of cloud liquid water spread in the three cases.

The majority of the spread is located in the lower part of the cumulus clouds as for KC16, but it is greater and extends higher in the cloud layer. It has a more realistic shape, for ARMCu and BOMEX, than the SPPT ensemble which has a spread peak in the upper part of the cloud not observed in the LESs. For the stratocumulus case, the stochastic convection ensemble has more spread in the upper part of the cloud, and is again more realistic when compared to the LES than the other ensembles. However, even if the spread of cloud water is improved with the stochastic convection scheme, it does not cover all the variability observed in the LES field, which is linked to the high spatial heterogeneity of the clouds.

4.3. Combination of KC16 and stochastic convection schemes

In the previous sections, KC16 and the stochastic convection schemes have been tested separately on each boundary layer case in order to understand their impact on the model variables and their ability to produce ensemble spread. However, the aim is to use them jointly in the same ensemble since they are designed to represent distinct small-scale variability sources. Here, we look at the spread of ensembles made with both KC16 and stochastic convection schemes activated. They will be denoted as POP (for Process-Oriented Perturbations) ensembles in the following.

For temperature and humidity spread, the interaction between the two schemes is not trivial (Figure 9). Overall, it is difficult to say whether the differences in the spread profiles are statistically significant, given that we are working with ensembles of finite size. Re-running the four ensembles for the ARMCu case shows that we can not conclude on any increase or decrease in the temperature and specific humidity spread when combining KC16 and stochastic convection schemes. For the FIRE case however, the dispersion from KC16 and the dispersion from stochastic convection seem to add up to each other. We also observe that, generally, the POP ensemble wind spread is lower than the KC16 ensemble wind spread (see Figure 9 for the zonal wind). Even if the vertical structure of the wind

spread in the KC16 ensembles and stochastic convection ensembles is very different, combining both schemes does not seem to bring additional spread and there even seems to be some form of compensation.

We finally summarize the results of POP and SPPT ensembles and of the coarse-grained LES in Figure 11. The time-height diagrams represent the tenth to ninetieth percentile range of the cloud water, potential temperature, humidity and wind for the ARMCu case.

In the LES domain, we observe two local maximums of temperature and humidity spread on the vertical, whose altitudes evolve during the simulation. The first peak appears two hours after the beginning of the simulation, and rises to about 2500 m after 15 hours. It can be associated with the top of the convective boundary layer, reached by many updrafts that bring to this altitude air moister and cooler than the environment. The second peak appears later, after about 6 hours, and is located around 1000 m height which corresponds to the cloud base height. It may therefore be associated with some variability in the condensation level height. The SCM ensembles reproduce reasonably well the position of the highest peak, although with either too much (SPPT ensemble) or too little (POP ensemble) spread. The second peak of temperature and humidity dispersion is less visible in the SCM ensembles. They rather produce dispersion within the subcloud layer for the temperature, because of the positive tendencies in the case of the SPPT ensemble, and because of the propagation of surface perturbations in the case of the POP ensemble. This is not necessarily a desirable feature, as we observe that in the LES, on the contrary, the subcloud layer is more homogeneous, probably because of active turbulent mixing.

Contrary to temperature and humidity, the wind spread range is similar between the LES and the SCM ensembles, but the vertical structure shows some important differences. The wind spread in the LES is large in the entire subcloud layer, especially near the surface. The SPPT ensemble does not reproduce this surface dispersion, contrary to the POP ensemble. However, in the POP ensemble the large wind spread values extend higher up in the subcloud layer because of high KC16 perturbations following the TKE profile, which may be less realistic.

The same kind of analysis for BOMEX (not shown here) leads to very similar results. The results for FIRE are however more difficult to interpret, because of the low spread generated by both SPPT and POP ensembles, except for the wind spread of the POP ensemble which has the same order of magnitude as the LES spread.

5. Discussion and conclusions

Two process-oriented perturbation schemes have been introduced in the AROME model, and used to produce single-column simulation ensembles of three idealized boundary-layer cases. The objective was to introduce perturbations representing the subgrid variability of turbulent and convective processes which is missing in deterministic parametrization schemes. This is achieved by different means in the two perturbation schemes. For the turbulence, the method of Kober and Craig (2016), which has been adapted here for AROME, relies on the use of second-order moments, diagnosed by the deterministic

scheme, to perturb its grid to subgrid relation. For shallow convection, the work of Sakradzija et al. (2015) enables one to relax some assumptions of classical convection schemes, such as the existence of a statistical sample of clouds inside a model grid box, by building a stochastic mass-flux closure condition. We have followed the same strategy, in a very simplified way, by using LESs to determine statistical distributions of the mass flux close to the surface in our experiments.

To evaluate the ability of these schemes to represent small-scale variability, we have relied on SCM/LES comparison for three well-known boundary-layer cases. Identical homogeneous forcing are applied to LES and SCM and the heterogeneity of the LES fields after the application of a coarse-graining procedure is taken as a reference for the subgrid state variability that the perturbation schemes should represent. Additional SCM ensembles have been performed with the 'bulk' perturbation scheme SPPT, which is an operational method to represent model uncertainty, modified in this study to target only turbulence and shallow convection.

The perturbation scheme targeting turbulence (KC16) proved to be effective in disturbing the horizontal wind variables in the three cases, which is a promising result. It is yet limited by the fact that our experiments only included single-column simulations where the horizontal advection is prescribed, hence no interaction with the dynamics could be studied, although we know that it can have important impacts (Lin and Neelin 2002). Future experiments with 3D simulations of real cases are planned, where the wind perturbation impact on the model will be further examined. The scheme may also have to be adapted in the light of Hirt et al. (2019) findings on the importance of having balanced 3D wind perturbations to prevent the appearance of acoustic modes leading to a fast decay of the perturbations.

The second process-oriented perturbation scheme, stochastic convection, had a much lower impact on the wind, but was able to generate substantial cloud water spread. Both stochastic convection and KC16 schemes, however, did not give satisfactory results regarding temperature and humidity spread. It was generally under-estimated, except for the temperature spread in the subcloud layer which could be too large. For KC16, the under-estimation is linked to the small values of temperature and humidity subgrid standard deviations, leading to perturbations that are small compared to the tendencies. As subgrid standard deviations are the central part of the KC16 scheme for controlling the perturbation amplitude, improving the scheme so that it produces more spread is not trivial. One could think of increasing the scale parameter α of Equation 1, however Kober and Craig (2016) emphasize that this parameter has no physical meaning, and should not be too different from 1. One possibility for future tests would be to decorrelate temperature, humidity (and wind) perturbations by using different random processes for each variable instead of a single one. On the other hand, Clark et al. (2021) argue that KC16 perturbations are, at least for temperature, too large at the top and bottom of the boundary layer. In our study, we found indeed too large temperature spread in the subcloud layer in some of our ensembles. Clark et al. (2021) suggest that a height-dependent length scale could be used in the scheme. We did some tests where we replaced the constant eddy length scale of 1000 m by the Bougeault-Lacarrère mixing length. The perturbations were indeed reduced at the surface, but it decreased the impact of the scheme on all the model

variables.
Regarding the stochastic convection scheme, in this study only the mass-flux closure condition was perturbed. However, we saw that the impact on the updraft characteristics was limited, therefore, perturbing other elements of the parametrization scheme such as the entrainment and detrainment rates should be considered. Another limitation of the stochastic convection scheme in this study is the determination of an appropriate distribution for the surface mass-flux perturbations. The results of the LES analysis showed that for the three cases it can be chosen as a Gaussian distribution function, and that its standard distribution was not too different between the cases. Some tests showed that replacing the σ_{LES}/μ_{LES} value of 0.33 by 0.4 for the BOMEX case resulted in ensembles with very similar dispersion. However, more boundary-layer situations should be examined to generalize these results and to check the behavior of the scheme when simulating real cases.

In this study, the SPPT method has been applied on the net tendencies of turbulence and shallow-convection parametrizations, instead of the total net tendencies. Therefore, both SPPT and POP methods perturb the same processes, but not in the same way. SPPT is globally a more efficient method to produce ensemble spread, which can be linked to the fact that it is designed to represent various errors coming from the parametrization schemes, whereas the process-oriented perturbation schemes studied here target only one possible source of error coming from the random occurrence of small-scale processes. The differences between POP and SPPT ensembles spread can also be explained by the temporal correlation of the noise which is set to 10 minutes in the KC16 scheme and 40 minutes in the stochastic convection scheme, whereas it is 6 hours in SPPT. SPPT perturbations thus accumulate on a longer time period. In several stochastic perturbation schemes, the temporal and spatial correlation scales of the uncertainty to be represented can not be easily determined. The noise correlation scales are thus chosen according to practical reasons and can be enlarged to ensure a significant impact on the resolved flow (Buizza et al. 1999, Ollinaho et al. 2017). In the process-oriented schemes presented here however, the temporal correlation of the perturbations is set following physical considerations: 10 minutes in KC16 scheme is the characteristic lifetime of an eddy, and 40 minutes for stochastic convection is chosen to represent the lifetime of a convective updraft.

In our experiments, all the perturbations were applied on turbulence and shallow-convection processes only. Although they are the main physical processes governing the evolution of the boundary layer in ARMCu and BOMEX cases, other processes are parametrized whose subgrid variability should also be accounted for. It appears to be critical in the FIRE case, where subgrid condensation and radiation are major contributors to the physical tendencies, and could also have an impact in the cloud layer of cumulus cases. Currently, the representation of uncertainty in these processes is done through random perturbations of some of their uncertain parameters (with RP, SPP schemes), and to the authors' knowledge, no process-oriented schemes similar to those presented in this study have been tested. Further development on the representation of subgrid variability for these processes would therefore be interesting, as well as the possible combination of RP and process-oriented perturbations. More generally, combining different perturbation

methods to produce ensembles seems to be an interesting – yet challenging – approach. It is probably difficult to combine perturbations that are not redundant, but it offers the possibility to take advantage of the benefits of each type of scheme. For example, the process-oriented schemes tested in this study produced non-negligible spread in the boundary layer, where SPPT perturbations are operationally switched off to avoid instabilities. However, by design they produce perturbations only once the processes in question have started, and seem less effective in modifying the timing of their triggering than SPPT. Christensen (2020) on the other hand argues that a multiplicative noise may not be suited to represent the uncertainty of all the parametrizations, and that a diversity of approaches could be beneficial. The iSPPT scheme (Christensen et al. 2017) , where each parametrization scheme is independently perturbed instead of the total tendencies, could be a convenient framework to test such combinations.

Appendix

The mean and standard deviation of the surface mass-flux distribution presented in section 3 are given in Table A1. These values have been used to draw the red curves on Figure 2.

References

Albrecht, B. A., Randall, D. A., & Nicholls, S. (1988). Observations of marine stratocumulus clouds during FIRE. *Bulletin of the American Meteorological Society*, *69*(6), 618-626.

Angevine, W. M., Olson, J., Kenyon, J., Gustafson, W. I., Endo, S., Suselj, K., & Turner, D. D. (2018). Shallow cumulus in WRF parameterizations evaluated against LASSO large-eddy simulations. *Monthly Weather Review*, *146*(12), 4303-4322.

Ayotte, K. W., Sullivan, P. P., Andren, A., Doney, S. C., Holtslag, A. A., Large, W. G., ... & Wyngaard, J. C. (1996). An evaluation of neutral and convective planetary boundary-layer parameterizations relative to large eddy simulations. *Boundary-Layer Meteorology*, *79*(1), 131-175.

Baker, L. H., Rudd, A. C., Migliorini, S., & Bannister, R. N. (2014). Representation of model error in a convective-scale ensemble prediction system. *Nonlinear Processes in Geophysics*, *21*(1), 19-39.

Batté, L., & Doblas-Reyes, F. J. (2015). Stochastic atmospheric perturbations in the EC-Earth3 global coupled model: Impact of SPPT on seasonal forecast quality. *Climate Dynamics*, *45*(11), 3419-3439.

Bengtsson, L., Dias, J., Tulich, S., Gehne, M., & Bao, J. W. (2021). A stochastic parameterization of organized tropical convection using cellular automata for global forecasts in NOAA's Unified Forecast System. *Journal of Advances in Modeling Earth Systems*, *13*(1), e2020MS002260.

Berner, J., Ha, S. Y., Hacker, J. P., Fournier, A., & Snyder, C. (2011). Model uncertainty in a mesoscale ensemble prediction system: Stochastic versus multiphysics representations. *Monthly Weather Review*, *139*(6), 1972-1995.

Berner, J., Fossell, K. R., Ha, S. Y., Hacker, J. P., & Snyder, C. (2015). Increasing the skill of probabilistic forecasts: Understanding performance improvements from model-error representations. *Monthly Weather Review*, *143*(4), 1295-1320.

Bougeault, P., & Lacarrere, P. (1989). Parameterization of orography-induced turbulence in a mesobeta–scale model. *Monthly weather review*, *117*(8), 1872-1890.

Bouttier, F., Vié, B., Nuissier, O., & Raynaud, L. (2012). Impact of stochastic physics in a convection-permitting ensemble. *Monthly Weather Review*, *140*(11), 3706-3721.

Brown, A. R., Cederwall, R. T., Chlond, A., Duynkerke, P. G., Golaz, J. C., Khairoutdinov, M., … & Stevens, B. (2002). Large-eddy simulation of the diurnal cycle of shallow cumulus convection over land. *Quarterly Journal of the Royal Meteorological Society*, *128*(582), 1075-1093.

Buizza, R., Miller, M., & Palmer, T. N. (1999). Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, *125*(560), 2887-2908.

Buizza, R., Houtekamer, P. L., Pellerin, G., Toth, Z., Zhu, Y., & Wei, M. (2005). A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems. *Monthly Weather Review*, *133*(5), 1076-1097.

Charron, M., Pellerin, G., Spacek, L., Houtekamer, P. L., Gagnon, N., Mitchell, H. L., & Michelin, L. (2010). Toward random sampling of model error in the Canadian ensemble prediction system. *Monthly Weather Review*, *138*(5), 1877-1901.

Cheng, Y., Canuto, V. M., & Howard, A. M. (2002). An improved model for the turbulent PBL. *Journal of the Atmospheric sciences*, *59*(9), 1550-1565.

Christensen, H. M., Moroz, I. M., & Palmer, T. N. (2015). Stochastic and perturbed parameter representations of model uncertainty in convection parameterization. *Journal of the Atmospheric Sciences*, *72*(6), 2525-2544.

Christensen, H. M., Lock, S. J., Moroz, I. M., & Palmer, T. N. (2017). Introducing independent patterns into the stochastically perturbed parametrization tendencies (SPPT) scheme. *Quarterly Journal of the Royal Meteorological Society*, *143*(706), 2168-2181.

Christensen, H. M. (2020). Constraining stochastic parametrisation schemes using high-resolution simulations. *Quarterly Journal of the Royal Meteorological Society*, *146*(727), 938-962.

Clark, P. A., Halliwell, C. E., & Flack, D. L. A. (2021). A Physically Based Stochastic Boundary Layer Perturbation Scheme. Part I: Formulation and Evaluation in a Convection-Permitting Model. *Journal of the Atmospheric Sciences*, *78*(3), 727-746.

Cohen, B. G., & Craig, G. C. (2006). Fluctuations in an equilibrium convective ensemble. Part II: Numerical experiments. *Journal of the Atmospheric Sciences*, *63*(8), 2005-2015.

Couvreux, F., Guichard, F., Redelsperger, J. L., Kiemle, C., Masson, V., Lafore, J. P., & Flamant, C. (2005). Water-vapour variability within a convective boundary-layer assessed by large-eddy simulations and IHOP_2002 observations. *Quarterly Journal of the Royal Meteorological Society*, *131*(611), 2665-2693.

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 meteorology*, *134*(3), 441-458.

Couvreux, F., Hourdin, F., Williamson, D., Roehrig, R., Volodina, V., Villefranque, N., … & Xu, W. (2021). Process-based climate model development harnessing machine learning: I. A calibration tool for parameterization improvement. *Journal of Advances in Modeling Earth Systems*, *13*(3), e2020MS002217.

Craig, G. C., & Cohen, B. G. (2006). Fluctuations in an equilibrium convective ensemble. Part I: Theoretical formulation. *Journal of the Atmospheric Sciences*, *63*(8), 1996-2004.

Cuxart, J., Bougeault, P., & Redelsperger, J. L. (2000). A turbulence scheme allowing for mesoscale and large-eddy simulations. *Quarterly Journal of the Royal Meteorological Society*, *126*(562), 1-30.

Cuxart, J., Holtslag, A. A., Beare, R. J., Bazile, E., Beljaars, A., Cheng, A., ... & Xu, K. M. (2006). Single-column model intercomparison for a stably stratified atmospheric boundary layer. *Boundary-Layer Meteorology*, *118*(2), 273-303.

De Rooy, W. C., & Siebesma, A. P. (2008). A simple parameterization for detrainment in shallow cumulus. *Monthly weather review*, *136*(2), 560-576.

Dorrestijn, J., Crommelin, D. T., Siebesma, A. P., & Jonker, H. J. (2013). Stochastic parameterization of shallow cumulus convection estimated from high-resolution model data. *Theoretical and Computational Fluid Dynamics*, *27*(1), 133-148.

Duda, J. D., Wang, X., Kong, F., & Xue, M. (2014). Using varied microphysics to account for uncertainty in warm-season QPF in a convection-allowing ensemble. *Monthly Weather Review*, *142*(6), 2198-2219.

Duynkerke, P. G., de Roode, S. R., van Zanten, M. C., Calvo, J., Cuxart, J., Cheinet, S., … & Sednev, I. (2004). Observations and numerical simulations of the diurnal cycle of the EUROCS stratocumulus case. *Quarterly Journal of the Royal Meteorological Society*, *130*(604), 3269-3296.

Groenemeijer, P., & Craig, G. C. (2012). Ensemble forecasting with a stochastic convective parametrization based on equilibrium statistics. *Atmospheric Chemistry and Physics*, *12*(10), 4555-4565.

Heus, T., J. Pols, C. F., J. Jonker, H. J., A. Van den Akker, H. E., & H. Lenschow, D. (2009). Observational validation of the compensating mass flux through the shell around cumulus clouds. *Quarterly Journal of the Royal Meteorological Society*, *135*(638), 101-112.

Hirt, M., Rasp, S., Blahak, U., & Craig, G. C. (2019). Stochastic parameterization of processes leading to convective initiation in kilometer-scale models. *Monthly Weather Review*, *147*(11), 3917-3934.

Holland, J. Z., & Rasmusson, E. M. (1973). Measurements of the atmospheric mass, energy, and momentum budgets over a 500-kilometer square of tropical ocean. *Monthly Weather Review*, *101*(1), 44-55.

Honnert, R., Masson, V., & Couvreux, F. (2011). A diagnostic for evaluating the representation of turbulence in atmospheric models at the kilometric scale. *Journal of the Atmospheric Sciences*, *68*(12), 3112-3131.

Honnert, R., Couvreux, F., Masson, V., & Lancz, D. (2016). Sampling the structure of convective turbulence and implications for grey-zone parametrizations. *Boundary-layer meteorology*, *160*(1), 133-156.

Houtekamer, P. L., Mitchell, H. L., & Deng, X. (2009). Model error representation in an operational ensemble Kalman filter. *Monthly Weather Review*, *137*(7), 2126-2143.

Huang, H. Y., Hall, A., & Teixeira, J. (2013). Evaluation of the WRF PBL parameterizations for marine boundary layer clouds: Cumulus and stratocumulus. *Monthly weather review*, *141*(7), 2265-2271.

Jankov, I., Berner, J., Beck, J., Jiang, H., Olson, J. B., Grell, G., … & Brown, J. M. (2017). A performance comparison between multiphysics and stochastic approaches within a North American RAP ensemble. *Monthly Weather Review*, *145*(4), 1161-1179.

Kober, K., & Craig, G. C. (2016). Physically based stochastic perturbations (PSP) in the boundary layer to represent uncertainty in convective initiation. *Journal of Atmospheric Sciences*, *73*(7), 2893-2911.

Lac, C., Chaboureau, J. P., Masson, V., Pinty, J. P., Tulet, P., Escobar, J., ... & Wautelet, P. (2018). Overview of the Meso-NH model version 5.4 and its applications. *Geoscientific Model Development*, *11*(5), 1929-1969.

Lang, S. T., Lock, S. J., Leutbecher, M., Bechtold, P., & Forbes, R. M. (2021). Revision of the Stochastically Perturbed Parametrisations model uncertainty scheme in the Integrated Forecasting System. *Quarterly Journal of the Royal Meteorological Society*, *147*(735), 1364-1381.

Lin, J. W. B., & Neelin, J. D. (2000). Influence of a stochastic moist convective parameterization on tropical climate variability. *Geophysical research letters*, *27*(22), 3691-3694.

Lin, J. W. B., & Neelin, J. D. (2002). Considerations for stochastic convective parameterization. *Journal of the Atmospheric Sciences*, *59*(5), 959-975.

Lorenz, E. N. (1965). A study of the predictability of a 28-variable atmospheric model. *Tellus*, *17*(3), 321-333.

Lu, C., Liu, Y., Zhang, G. J., Wu, X., Endo, S., Cao, L., ... & Guo, X. (2016). Improving parameterization of entrainment rate for shallow convection with aircraft measurements and large-eddy simulation. *Journal of Atmospheric Sciences*, *73*(2), 761-773.

Machulskaya, E., & Seifert, A. (2019). Stochastic differential equations for the variability of atmospheric convection fluctuating around the equilibrium. *Journal of Advances in Modeling Earth Systems*, *11*(8), 2708-2727.

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

Nie, J., & Kuang, Z. (2012). Responses of shallow cumulus convection to large-scale temperature and moisture perturbations: A comparison of large-eddy simulations and a convective parameterization based on stochastically entraining parcels. *Journal of the atmospheric sciences*, *69*(6), 1936-1956.

Ollinaho, P., Lock, S. J., Leutbecher, M., Bechtold, P., Beljaars, A., Bozzo, A., … & Sandu, I. (2017). Towards process-level representation of model uncertainties: stochastically perturbed parametrizations in the ECMWF ensemble. *Quarterly Journal of the Royal Meteorological Society*, *143*(702), 408-422.

Palmer, T. N., Buizza, R., Doblas-Reyes, F., Jung, T., Leutbecher, M., Shutts, G. J., … & Weisheimer, A. (2009). Stochastic parametrization and model uncertainty.

Pergaud, J., Masson, V., Malardel, S., & Couvreux, F. (2009). A parameterization of dry thermals and shallow cumuli for mesoscale numerical weather prediction. *Boundary-layer meteorology*, *132*(1), 83-106.

Plant, R. S., & Craig, G. C. (2008). A stochastic parameterization for deep convection based on equilibrium statistics. *Journal of the Atmospheric Sciences*, *65*(1), 87-105.

Redelsperger, J. L., & Sommeria, G. (1981). Méthode de représentation de la turbulence d'échelle inférieure à la maille pour un modèle tri-dimensionnel de convection nuageuse. *Boundary-Layer Meteorology*, *21*(4), 509-530.

Rio, C., & Hourdin, F. (2008). A thermal plume model for the convective boundary layer: Representation of cumulus clouds. *Journal of the atmospheric sciences*, *65*(2), 407-425.

Rio, C., Del Genio, A. D., & Hourdin, F. (2019). Ongoing breakthroughs in convective parameterization. *Current Climate Change Reports*, *5*(2), 95-111.

Rochetin, N., Couvreux, F., Grandpeix, J. Y., & Rio, C. (2014). Deep convection triggering by boundary layer thermals. Part I: LES analysis and stochastic triggering formulation. *Journal of the Atmospheric Sciences*, *71*(2), 496-514.

Sakradzija, M., Seifert, A., & Heus, T. (2015). Fluctuations in a quasi-stationary shallow cumulus cloud ensemble. *Nonlinear Processes in Geophysics*, *22*(1), 65-85.

Sakradzija, M., Seifert, A., & Dipankar, A. (2016). A stochastic scale-aware parameterization of shallow cumulus convection across the convective gray zone. *Journal of Advances in Modeling Earth Systems*, *8*(2), 786-812.

Sakradzija, M., & Klocke, D. (2018). Physically constrained stochastic shallow convection in realistic kilometer-scale simulations. *Journal of Advances in Modeling Earth Systems*, *10*(11), 2755-2776.

Seity, Y., Brousseau, P., Malardel, S., Hello, G., Bénard, P., Bouttier, F., ... & Masson, V. (2011). The AROME-France convective-scale operational model. *Monthly Weather Review*, *139*(3), 976-991.

Shutts, G. (2005). A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quarterly Journal of the Royal Meteorological Society*, *131*(612), 3079-3102.

Siebesma, A. P., Bretherton, C. S., Brown, A., Chlond, A., Cuxart, J., Duynkerke, P. G., ... & Stevens, D. E. (2003). A large eddy simulation intercomparison study of shallow cumulus convection. *Journal of the Atmospheric Sciences*, *60*(10), 1201-1219.

Soares, P. M. M., Miranda, P. M. A., Siebesma, A. P., & Teixeira, J. (2004). An eddy-diffusivity/mass-flux parametrization for dry and shallow cumulus convection. *Quarterly Journal of the Royal Meteorological Society*, *130*(604), 3365-3383.

Stensrud, D. J., Bao, J. W., & Warner, T. T. (2000). Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective systems. *Monthly Weather Review*, *128*(7), 2077-2107.

Sušelj, K., Teixeira, J., & Chung, D. (2013). A unified model for moist convective boundary layers based on a stochastic eddy-diffusivity/mass-flux parameterization. *Journal of the Atmospheric Sciences*, *70*(7), 1929-1953.

Wilks, D. S. (2005). Effects of stochastic parametrizations in the Lorenz'96 system. *Quarterly Journal of the Royal Meteorological Society*, *131*(606), 389-407.

Zhao, M., & Austin, P. H. (2005). Life cycle of numerically simulated shallow cumulus clouds. Part I: Transport. *Journal of the atmospheric sciences*, *62*(5), 1269-1290.

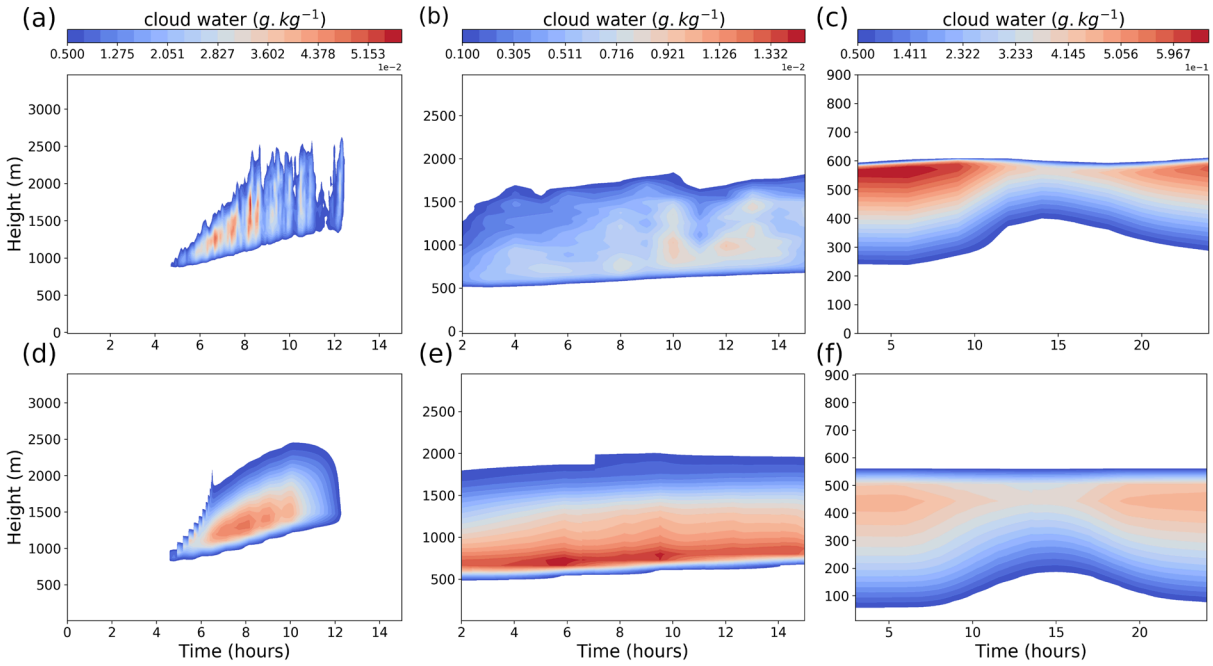

FIGURE 1 *Time-height diagrams of the specific mass of cloud liquid water, averaged on the LES domain (a, b, c) and in the column of AROME-SCM (d, e, f) for ARMCu (a, d), BOMEX (b, e) and FIRE (c, f). Note that the y-axis does not extend to the same altitude for the three cases and each case has a specific color bar.*

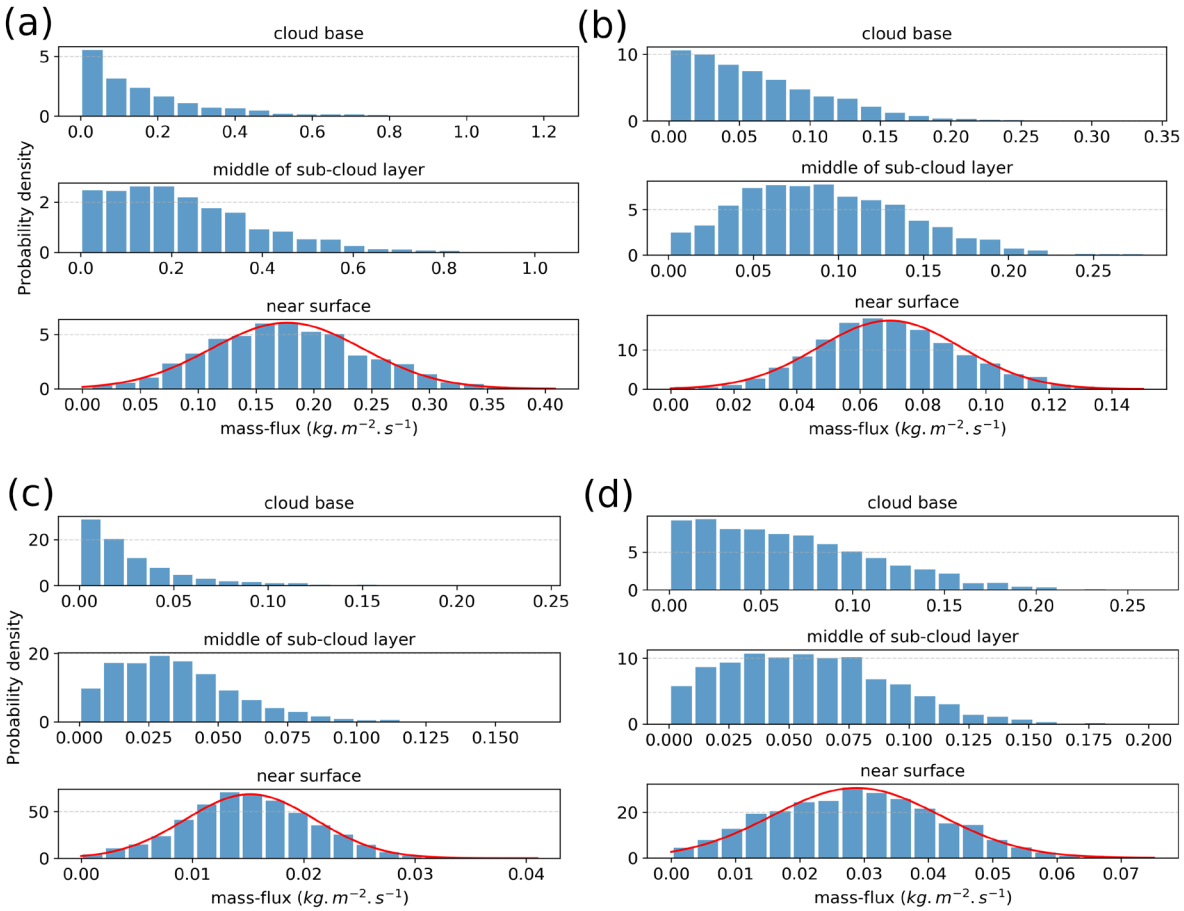

FIGURE 2 *Histograms of mass-flux values on the LES domains at three different levels: cloud base, middle of the subcloud layer (0.5 x cloud base height) and near surface (0.1 x cloud base height), for (a) ARMCu between +5 and +10h, (b) BOMEX between +2 and +15h, (c) FIRE between +12 and +16h (day) and (d) FIRE between +20 and +24h (night). The results of the FIRE case have been split in two periods corresponding to day and night because the behavior of the updrafts changes according to a diurnal cycle. The mass-flux values have been computed assuming a constant density ρ=1. The red curve in the bottom histograms corresponds to a Gaussian fit to the data.*

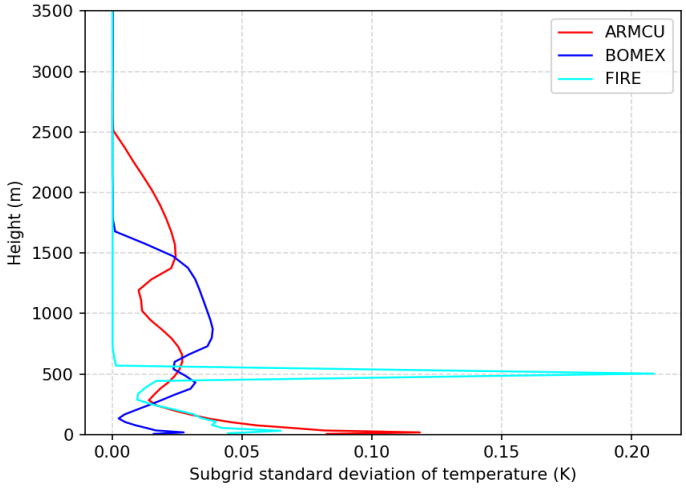

FIGURE 3 *Time-averaged vertical profiles of temperature subgrid standard deviation in the deterministic AROME-SCM simulation for ARMCu, BOMEX and FIRE.*

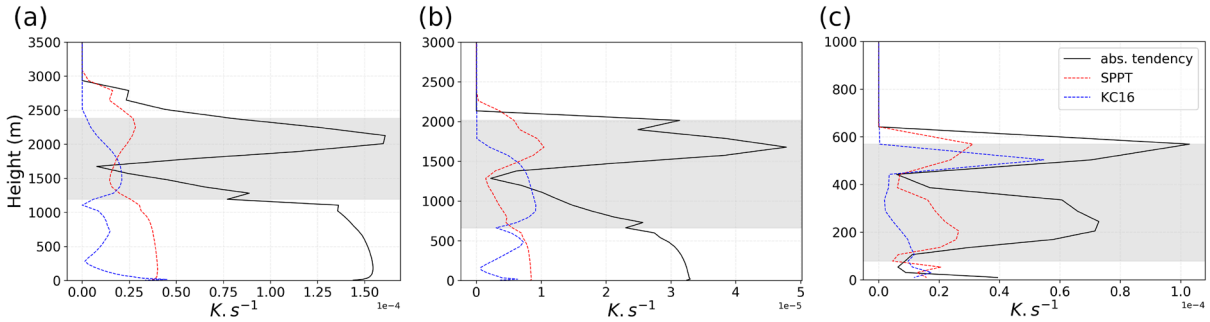

FIGURE 4 *Vertical profiles of the absolute net temperature tendency from turbulence and shallow convection in the control simulation (black line), the standard deviation of KC16 perturbations (blue-dashed line) in a 100-member ensemble, and the standard deviation of SPPT perturbations (red-dashed line) in a 100-member ensemble. SPPT perturbations are defined as the tendency difference before/after the application of the SPPT scheme. The grey shaded area represents the vertical extension of the cloud in the control simulation. Profiles are shown for (a) ARMCu at +9h, (b) BOMEX at +14h and (c) FIRE at +24h.*

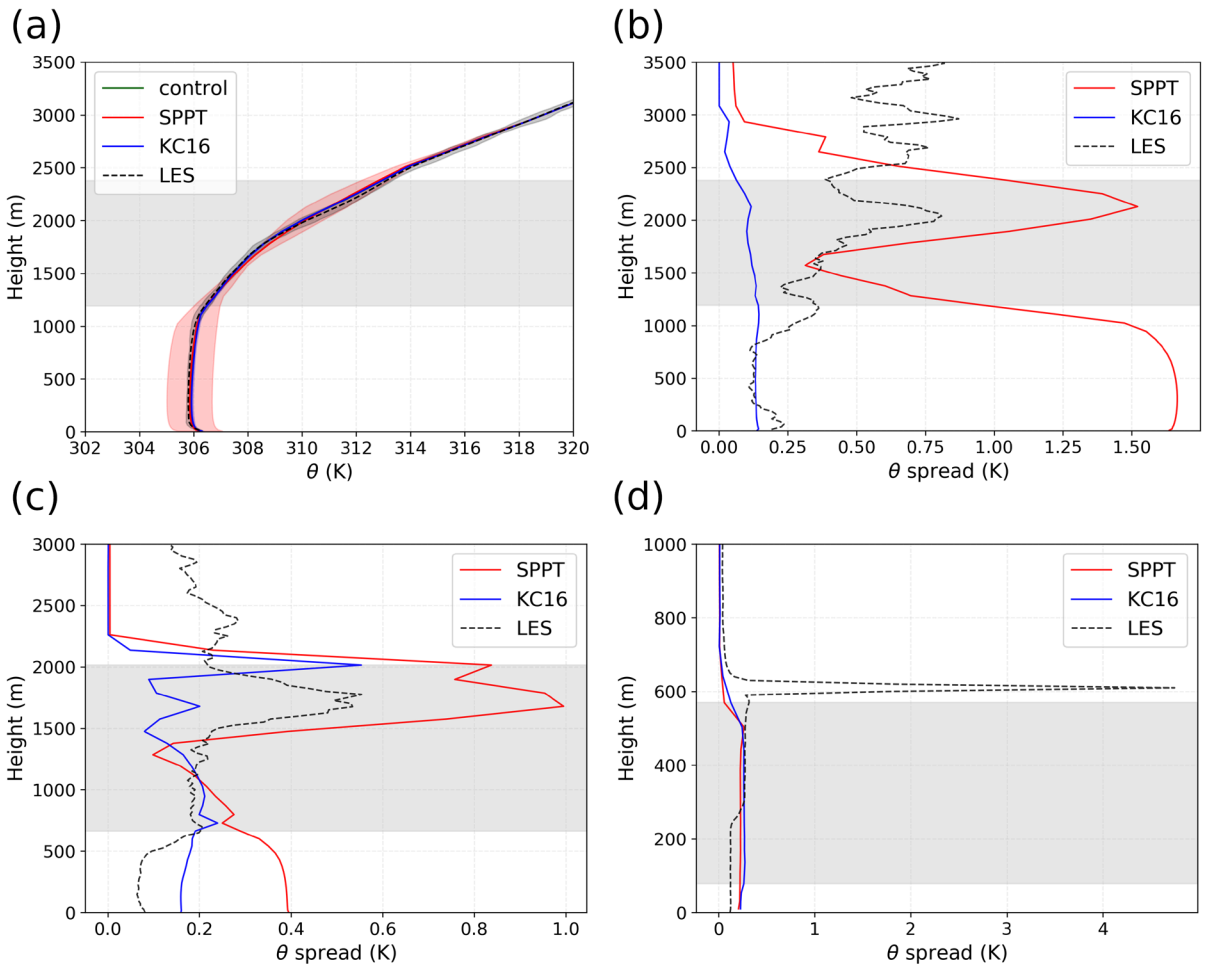

FIGURE 5 *(a) Potential temperature profiles of the control simulation (green line), the SPPT ensemble mean (red line), the KC16 ensemble mean (blue line) and the coarse-grained LES mean (black dashed line) for ARMCu at +9h. The red shaded area is the tenth to ninetieth percentile range of the SPPT ensemble. This range is much smaller for the KC16 ensemble and for the coarse-grained LES so the corresponding shaded areas are not visible on the graph. (b,c,d) Vertical profiles of the potential temperature spread in SCM ensembles and coarse-grained LES for (b) ARMCu at +9h, (c) BOMEX at +14h and (d) FIRE at +24h. The spread is calculated as the tenth to ninetieth percentile range. The horizontal grey shaded area represents the vertical extension of the cloud in the control simulation.*

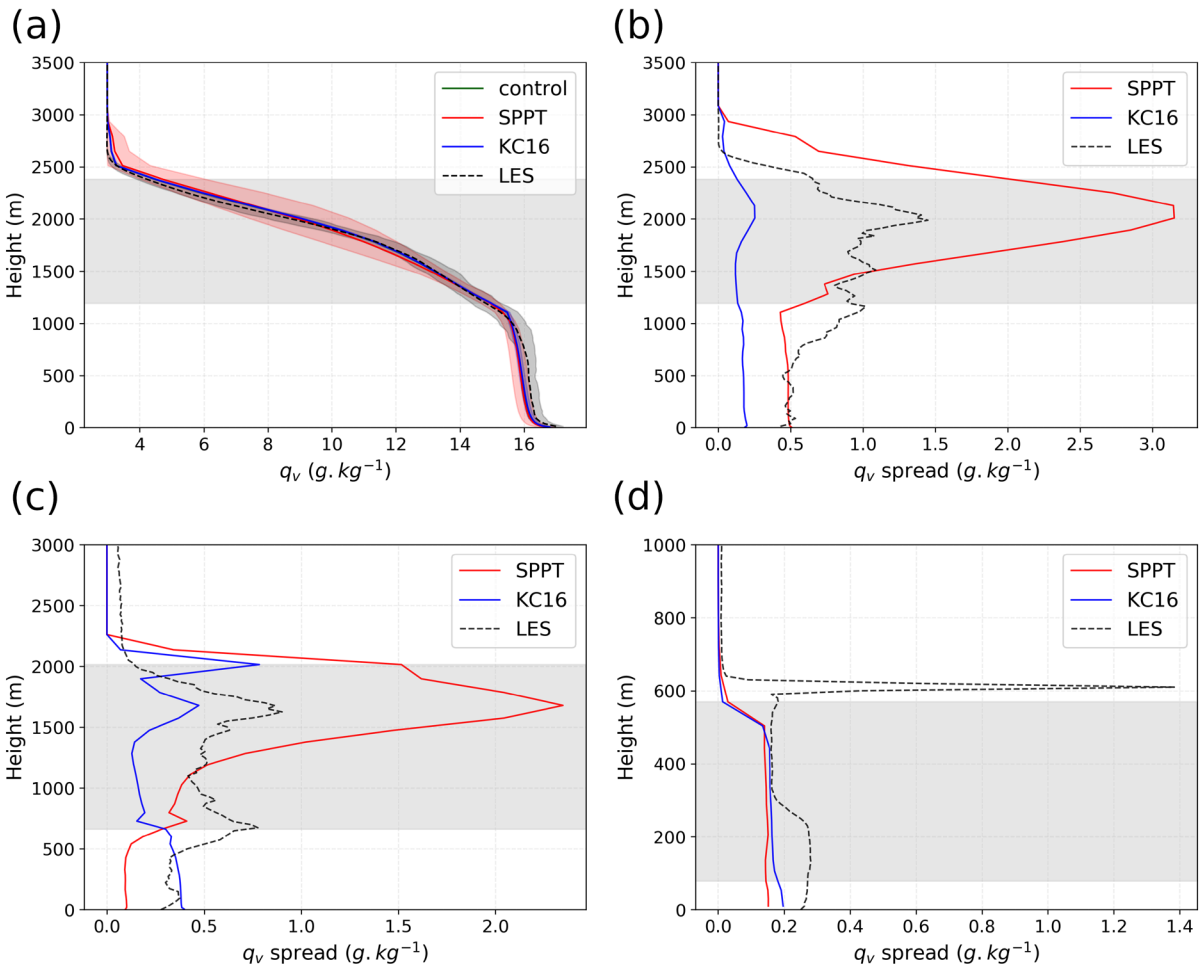

FIGURE 6 *Same as Figure 5 but for specific humidity.*

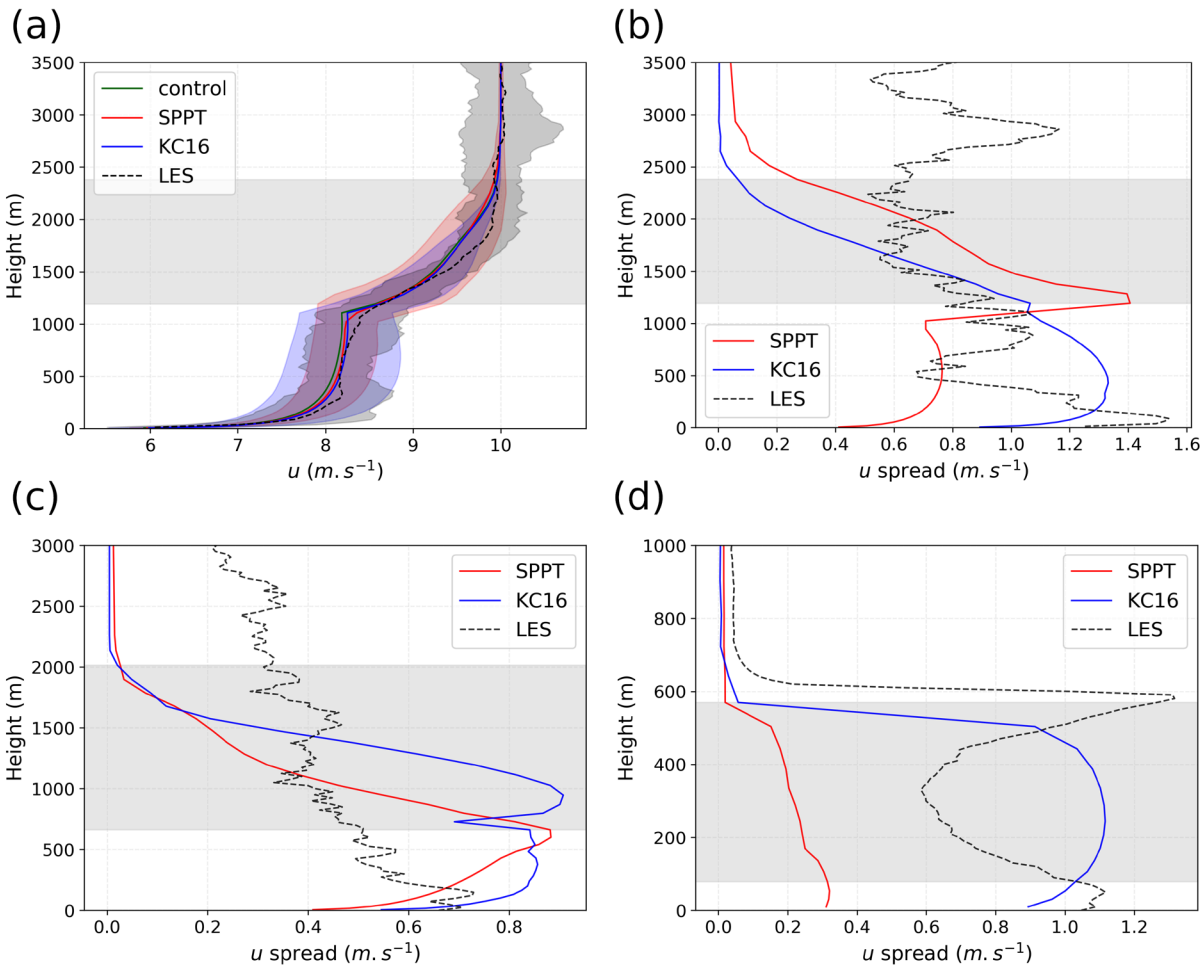

FIGURE 7 *Same as Figure 5 but for zonal wind.*

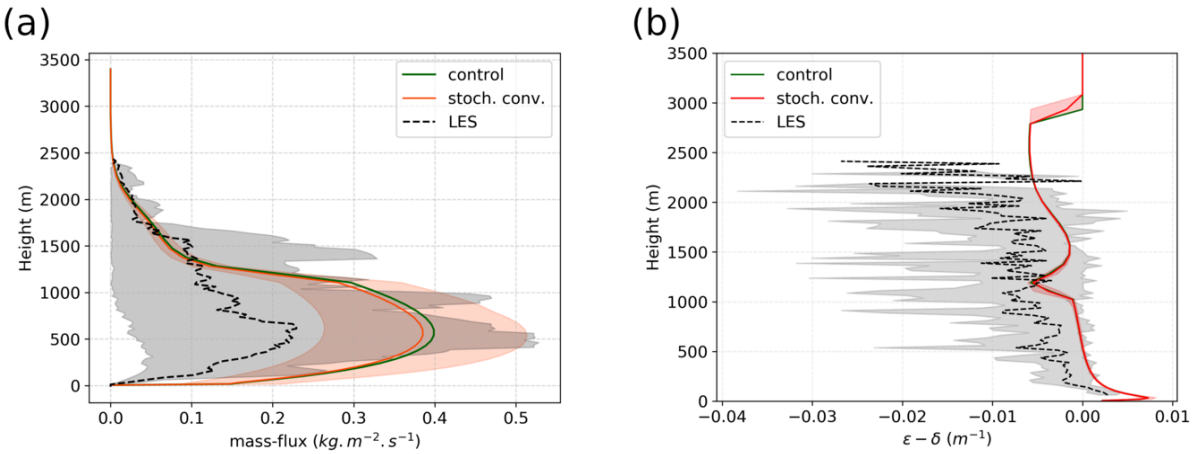

FIGURE 8 *Vertical profiles of (a) mass flux and (b) entrainment (ϵ) minus detrainment (δ) rate for a 100-member AROME-SCM ensemble produced with stochastic convection (in red) and coarse-grained LES (in grey), for the ARMCu case at +9h. Lines are ensemble averages and shaded areas represent tenth to ninetieth percentile ranges. The green lines are the profiles of the control simulation.*

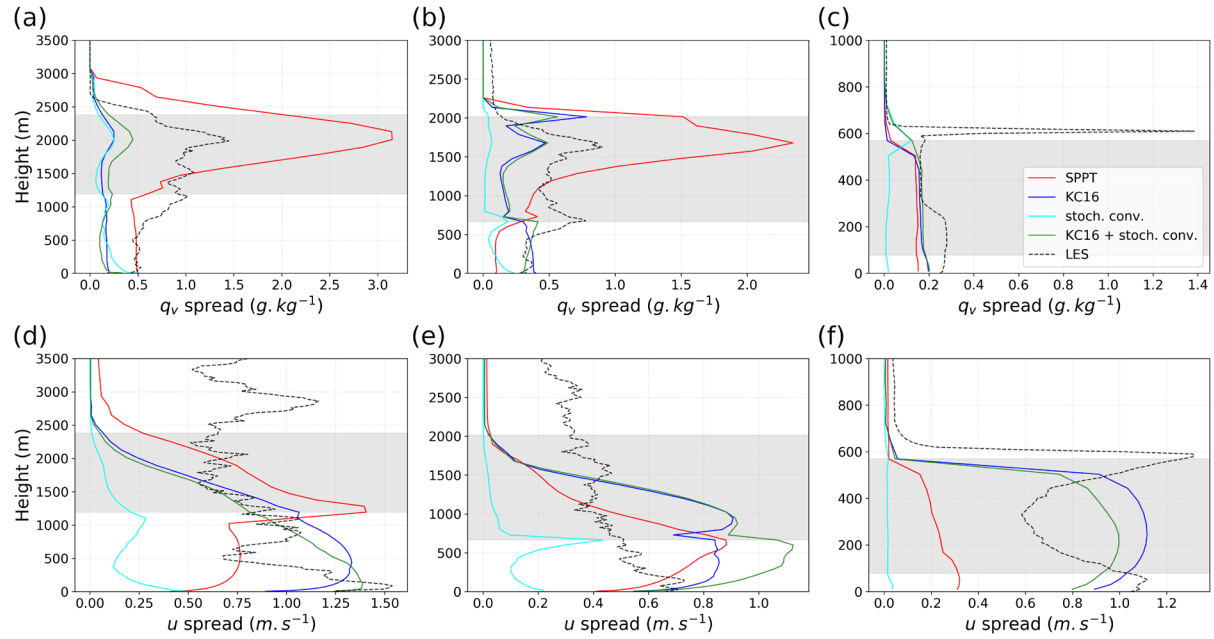

FIGURE 9 *Tenth to ninetieth percentile range profiles of (a,b,c) specific humidity and (d,e,f) zonal wind for the AROME-SCM ensembles and the coarse-grained LES, for (a,d) ARMCu at +9h, (b,e) BOMEX at +14h, (c,f) and FIRE at +24h.*

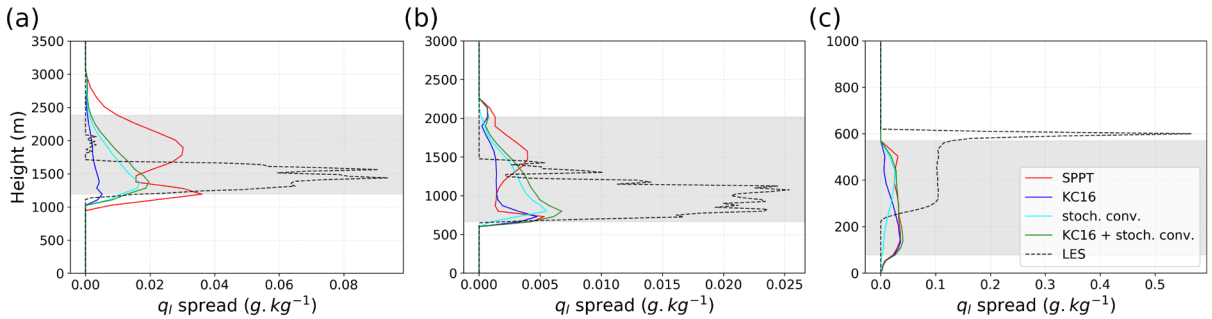

FIGURE 10 *Same as Figure 9 but for cloud liquid water.*

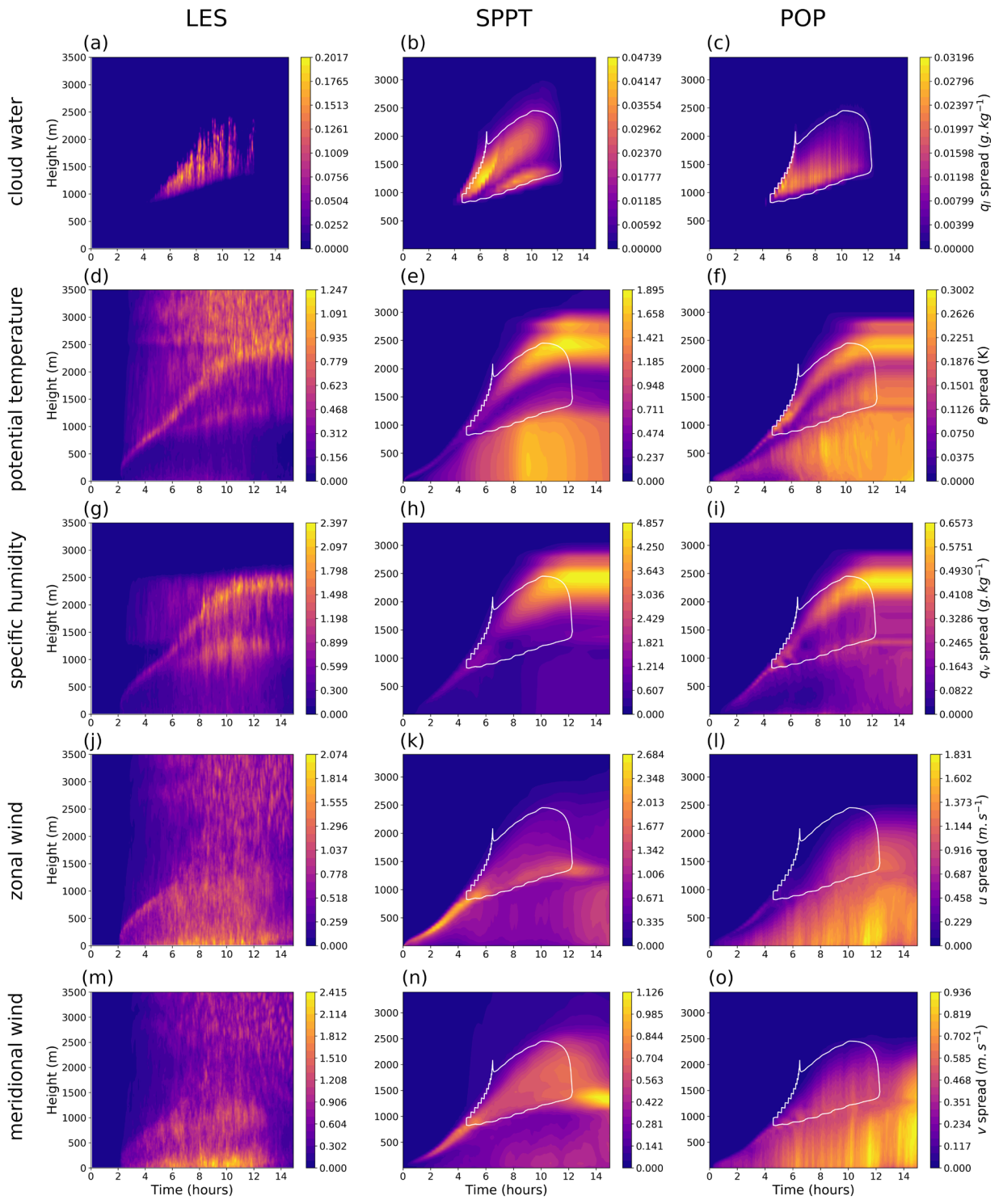

FIGURE 11 *Time-height diagrams of the tenth to ninetieth percentile range of cloud water (a, b, c), potential temperature (d, e, f), specific humidity (g, h, i), zonal wind (j, k, l) and meridional wind (m, n, o) in the LES field (a, d, g, j, m), the SPPT ensemble (b, e, h, k, n) and the*

POP ensemble (c, f, i, l, o), for the ARMCu case. The white contour corresponds to the cloud limits in the control simulation.

TABLE 1 *Description of LES and SCM simulation setup for each case.*

	ARMCU	BOMEX	FIRE
LES domain	6.4 x 6.4 x 4 km^3	12.8 x 12.8 x 4 km^3	25.6 x 25.6 x 1.2 km^3
LES resolution	$\Delta x = \Delta y = \Delta z = 25$ m	$\Delta x = \Delta y = \Delta z = 25$ m	$\Delta x = \Delta y = 50$ m, $\Delta z = 10$ m
Time range	15 hours: 05:30 LT – 20:30 LT	14 hours	24 hours: 00:00 LT – 24:00 LT
Time availability of LES fields	every five minutes	hourly	hourly
Large-scale forcing	evolving temperature and humidity advection, radiative forcing, steady geostrophic wind	steady temperature radiative forcing, humidity advection, subsidence and geostrophic wind	steady temperature advection, humidity advection, subsidence and geostrophic wind
Surface forcing	sensible and latent heat fluxes	sensible and latent heat fluxes	sea surface temperature
Parametrization schemes activated in AROME-SCM	shallow convection, turbulence, condensation and cloud microphysics	shallow convection, turbulence, condensation and cloud microphysics	radiation, shallow convection, turbulence, condensation and cloud microphysics

TABLE A1 Empirical mean and standard deviation values of the mass-flux distribution near the surface, for the three cases.

	ARMCU	BOMEX	FIRE	
			day	night
μ_{LES} ($kg.m^{-2}.s^{-1}$)	0.18	0.07	0.02	0.03
σ_{LES} ($kg.m^{-2}.s^{-1}$)	0.07	0.02	0.01	0.01
σ_{LES}/μ_{LES}	0.37	0.33	0.4	0.45