



**HAL**  
open science

# Deep Convection Triggering by Boundary Layer Thermals. Part I: LES Analysis and Stochastic Triggering Formulation

Nicolas Rochetin, Fleur Couvreur, Jean-Yves Grandpeix, Catherine Rio

► **To cite this version:**

Nicolas Rochetin, Fleur Couvreur, Jean-Yves Grandpeix, Catherine Rio. Deep Convection Triggering by Boundary Layer Thermals. Part I: LES Analysis and Stochastic Triggering Formulation. *Journal of the Atmospheric Sciences*, 2014, 71 (2), pp.496-514. 10.1175/JAS-D-12-0336.1 . hal-01083240

**HAL Id: hal-01083240**

**<https://polytechnique.hal.science/hal-01083240>**

Submitted on 12 Nov 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Deep Convection Triggering by Boundary Layer Thermals. Part I: LES Analysis and Stochastic Triggering Formulation

NICOLAS ROCHETIN

*Laboratoire de Météorologie Dynamique, Paris, France*

FLEUR COUVREUX

*CNRM-GAME, Météo-France and CNRS, Toulouse, France*

JEAN-YVES GRANDPEIX AND CATHERINE RIO

*Laboratoire de Météorologie Dynamique, Paris, France*

(Manuscript received 11 December 2012, in final form 3 September 2013)

## ABSTRACT

This paper proposes a new formulation of the deep convection triggering for general circulation model convective parameterizations. This triggering is driven by evolving properties of the strongest boundary layer thermals. To investigate this, a statistical analysis of large-eddy simulation cloud fields in a case of transition from shallow to deep convection over a semiarid land is carried out at different stages of the transition from shallow to deep convection. Based on the dynamical and geometrical properties at cloud base, a new computation of the triggering is first proposed. The analysis of the distribution law of the maximum size of the thermals suggests that, in addition to this necessary condition, another triggering condition is required, that is, that this maximum horizontal size should exceed a certain threshold. This is explicitly represented stochastically. Therefore, the new formulation integrates the whole transition process from the first cloud to the first deep convective cell and can be decomposed into three steps: (i) the appearance of clouds, (ii) crossing of the inhibition layer, and (iii) deep convection triggering.

## 1. Introduction

Many features of tropical deep convection are accounted for by the quasi-equilibrium (QE) hypothesis. According to this hypothesis, deep convection responds very rapidly to changes in tropospheric stability because of large-scale circulation and radiative forcing, and the tropical troposphere is thus permanently close to a state of equilibrium. However, several authors have emphasized that an atmosphere in a permanent QE state would exhibit an exceedingly low variability at small and large scales (Neelin et al. 2008; Jones and Randall 2011; Raymond and Herman 2011). Using cloud-resolving model (CRM) simulations, Raymond and Herman (2011) showed that the response of deep convection to a perturbation was

very fast (hours) only in the lower half of the troposphere and was much slower in the upper half. This points to the importance of the depth of moist convection and suggests that the QE hypothesis is valid in the region of the troposphere reached by cumulus clouds but not in the region reached only by deep convection. Cumulonimbus clouds efficiently warm the upper troposphere: when present, they bring the CAPE back to very low values in a matter of hours. However, they are short lived (about 30 min) and are present only as long as the triggering of new elements continues. It is then tempting to suppose, following Neelin et al. (2008) and Stechmann and Neelin (2011), that the main reason why deep convection departs from QE is that there are lapses of time where triggering of new convective cells does not occur and where the upper troposphere may drift freely away from QE.

Subcloud lifting processes and convective inhibition (CIN) are known to exert a strong control on deep convection onset and intensity, modulating the entropy flux

---

*Corresponding author address:* Nicolas Rochetin, Laboratoire de Météorologie Dynamique, Boite 99, 4, place Jussieu, F-75252 Paris CEDEX 05, France.  
E-mail: nicolas.rochetin@lmd.jussieu.fr

from the boundary layer to the free troposphere (Emanuel et al. 1994). Mapes (2000) assumes that deep convective triggering occurs when turbulent kinetic energy in the boundary layer (the triggering energy) is sufficient to overcome CIN. With this picture in mind, the question of the occurrence and variability of moist convection in the tropics is strongly dependent on the departure of the troposphere from QE states and thus on the action of boundary layer processes on deep convection triggering. The present series of papers pursues of these ideas further and addresses the questions of deep convection triggering and its representation in climate models.

The QE hypothesis plays an important role in deep convection parameterizations since it makes it possible to express deep convection processes as a function of large-scale conditions. However, since departure from QE is a key factor for climate variability, convective parameterizations should not be bound too strongly by the QE hypothesis.

According to Jones and Randall (2011) (see also Xu et al. 1992), several methods have been used to drive the local atmospheric system away from QE. In a first approach (the superparameterization technique) a CRM is embedded within each general circulation model (GCM) grid cell, and the variability around QE is provided by the CRM internal variability [e.g., Plant and Craig (2008) emphasize the variability provided by CRMs for given large-scale conditions].

In a second approach (Palmer 2012) the tendencies computed by the physical parameterizations are perturbed randomly; the system is no longer driven toward QE but toward a target moving randomly around QE. We shall follow Neelin et al. (2008) and assume that movement away from QE occurs mainly when deep convection is not active. Consequently, determining the period of activity of deep convection is a key issue for representing climate variability.

In observations and in high-resolution simulations of moist convection, the triggering (or onset) of deep convection is the time when cumulus clouds reach the highest levels of the troposphere (i.e., congestus and cumulonimbus). Prior to this sharp transition, the convective boundary layer enters a transient regime (transition stage), during which cumulus clouds become gradually wider and deeper, but still remain in the low troposphere (Chaboureau et al. 2004; Guichard et al. 2004; Grabowski et al. 2006; Khairoutdinov and Randall 2006). Chaboureau et al. (2004) show that, during the transition phase, the updraft vertical velocities at cloud base are large enough to overcome the convective inhibition but that entrainment of exceedingly dry air limits the cloud vertical development. It is only when the lower free troposphere is moist enough that the sharp

transition to deep convection occurs. Thus, they propose a two-step trigger in which stability and moisture are the two critical variables controlling the transition.

The objective is therefore to design multiple-step triggering that accounts for the evolving properties of the strongest boundary layer thermals applicable to any GCM that treats boundary layer structures independently from deep convection, for example, through the eddy diffusivity–mass flux approach. This type of scheme combines a diffusivity scheme, representing the small-scale turbulence, with a mass flux scheme representing the organized structures of the boundary layer (including the cumulus clouds).

The question of deep convection triggering is of particular interest over lands, where the boundary layer is, on average, higher than over the ocean (Medeiros et al. 2005) and is capped by a stronger inhibition layer. Thus, over land, the lifted parcel cannot reach its level of free convection (LFC) without some dynamical forcing, and the shallow and deep regimes are thus more distinct in space and time.

Actually, most current GCMs miss this transition phase and consequently represent the diurnal cycle of deep convection over land rather poorly (Grabowski et al. 2006; Yang and Slingo 2001; Guichard et al. 2004; Bechtold et al. 2004). According to Guichard et al. (2004), this is because the gradual moistening of the low free troposphere due to the detrainment at the top of cumulus clouds is not well represented in GCM parameterizations, so current GCMs cannot capture the succession of dry, shallow, and deep convection regimes.

Observations have shown that shallow cumulus clouds are the saturated part of thermals initiated at the surface and driven by buoyancy (LeMone and Pennell 1976). Here we define a thermal as a coherent structure rising from the surface to the top of the dry or cloudy boundary layer and carrying out most of the vertical transport of heat, moisture, and momentum. The thermal then divides into a subcloud layer and a cloudy layer. Deeper congestus and cumulonimbus clouds burst locally, overcoming an inhibition barrier, and are associated with precipitation and cold pools driven by the evaporation of rain under the cloud base.

In the current version of the Laboratoire de Météorologie Dynamique–Zoom (LMDZ) model, we treat shallow and deep convection separately. On the one hand, shallow convection is handled in a unified way with the boundary layer turbulence. This is done by combining a diffusive approach with a mass flux approach representing both dry and shallow convection. The so-called thermal plume model (Rio and Hourdin 2008) idealizes the effect of all dry and cloudy thermals contained in a model grid cell by considering a mean

ascending dry or cumulus-topped thermal covering a fraction  $\alpha$  of the grid cell and compensated by subsidence in the surrounding environment. In this way, shallow convection occurs when the ascending thermal condenses and no triggering criterion for shallow convection is required. On the other hand, deep convection and the associated precipitation and downdrafts are handled by the Emanuel episodic mixing and buoyancy sorting scheme (Emanuel 1991), coupled with a parameterization of cold pools driven by the evaporation of deep convective rain (Grandpeix and Lafore 2010).

Deep convection is first initiated if the dynamical lifting provided by boundary layer thermals is sufficient to overcome the convective inhibition. Once activated, deep convection is sustained by cold pools that provide an additional source of lifting. Then, the deep convection scheme is coupled with local lifting processes through two variables: the available lifting energy (ALE, expressed in joules per kilogram) and the available lifting power (ALP, expressed in watts per square meter). Convection triggering and closure are expressed in terms of ALE (convection is triggered when  $ALE > |CIN|$ ) and ALP (cloud-base mass flux is proportional to ALP). In the LMDZ5B model, the lifting processes considered are (i) the boundary layer thermals (subscript  $_{BL}$ ) and (ii) the cold pools (subscript  $_{WK}$  for wake) fed by unsaturated downdrafts resulting from the reevaporation of rain below cumulonimbus clouds. The ALE is the maximum of the lifting energies [ $ALE = \max(ALE_{BL}; ALE_{WK})$ ] and ALP is the sum of the two lifting powers ( $ALP = ALP_{BL} + ALP_{WK}$ ). The present paper is only concerned with deep convection triggering, that is, only with the ALE variable. Moreover, since we are specifically interested in convection initiation, only the lifting energy due to boundary layer thermals has to be considered (cold pools only act to maintain deep convection, after its onset). In the current version of the LMDZ5 GCM (LMDZ5B), the lifting energy is deduced from the maximum vertical velocity within the thermal:  $ALE_{BL} = 0.5\{\max_z[w'_u(z)]\}^2$ . Note that in most cases,  $\max_z[w'_u(z)]$  is located in the cloudy part of the thermal in the model. However, as will be revealed in section 6, this is quite insufficient to describe the transition regime.

The aim of this paper is to revisit the definition of  $ALE_{BL}$  by identifying the key factors controlling the transition from shallow to deep convection. The final goal is to define a triggering criterion for deep convection from the properties of thermals associated with shallow convection.

Several studies using cloud-resolving models have been used to characterize this complex transition from shallow to deep convection and provide some insights into the variables that control deep convection triggering.

While Chaboureau et al. (2004) proposes that deep convection starts when a variable called the normalized saturation deficit (NSD) at the cloud base reaches its minimum (as NSD is strongly linked to the cloud cover, triggering occurs when the cloud cover reaches a critical value), Wu et al. (2009) shows that the virtual temperature profile of the average cloud is a key factor, and Khairoutdinov and Randall (2006) and Grabowski et al. (2006) stress the importance of horizontal cloud size. Thus, several parameters seem to play key roles in deep convection triggering: at cloud base, the humidity of the troposphere, the cloud cover, and the size of individual clouds are significant, and above cloud base, the thermodynamic properties of cumulus clouds are important.

Here we tackle the problem of how deep convection triggering is represented in climate models. Using large-eddy simulation (LES) data in a continental case of transition from shallow to deep convection, we extract the statistical properties of the thermals at cloud base and propose a new computation of  $ALE_{BL}$ . The goal is to propose a simple formulation of the triggering process, easily integrable in a GCM. This new formulation describes the whole transition process and in particular the episodic nature of the triggering.

Section 2 describes the theoretical framework, and section 3 describes the method. The cross-sectional spectrum of the thermals inside the domain is studied in section 4, and the vertical velocity spectrum inside the thermals is examined in section 5. The  $ALE_{BL}$  computation is described in section 6. The triggering formulation is proposed in section 7, and some final comments are given in section 8.

## 2. Single versus spectral thermal approaches

In a typical GCM grid ( $L \geq 100$  km), the expected number of thermals can be very large. The “bulk thermal” approach may then be useful to predict their collective effect on heat and moisture transport. This approach considers a single (or average, or bulk) thermal of cross section  $S_{tot}$ , covering a fractional area  $\alpha_{tot}$  (see Fig. 1). The vertical profiles of vertical velocity inside (i.e., in the ascending zone) and outside (i.e., in the surrounding environment) the bulk thermal are  $w'_u(z)$  and  $\bar{w}_e(z)$ , respectively (see Fig. 1). This approach is the basis of the mass flux scheme used to represent dry and moist thermals in LMDZ.

However, this approach is no longer useful when considering the size of the thermals. In such a case, a cross-sectional distribution has to be taken into consideration and requires a spectral approach. Figure 1 illustrates the differences between the single thermal and the statistical approaches.

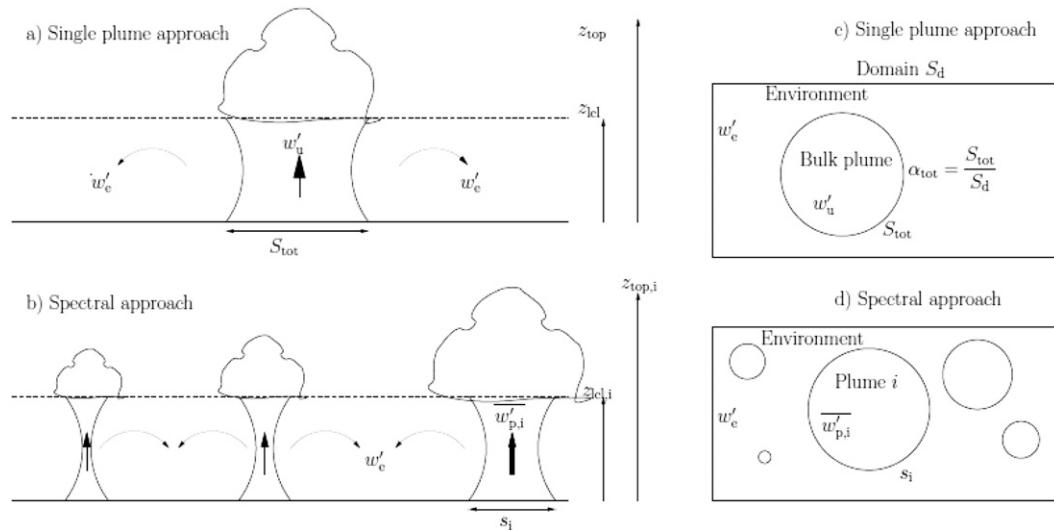


FIG. 1. (left) Side view and (right) top view of the (a),(c) single thermal vs the (b),(d) spectral approach to thermal modeling.

Concerning shallow cumulus (topping boundary layer thermals), Neggers et al. (2003) and Rodts et al. (2003) studied the properties of the cloud field using aircraft measurements, satellite data, and large-eddy simulations. They showed that many distribution laws are possible fits for the cloud cross-sectional (size) spectrum over the domain; among them were the exponential law, the log-normal law, and some other power laws.

For deeper (precipitating) clouds, Craig and Cohen (2006) proposed (using CRM simulations) an exponential probability distribution function (PDF)  $\mathcal{P}_m$  to represent the individual cloud mass flux spectrum [ $\mathcal{P}_m(m) = (1/\langle m \rangle) \exp(-m/\langle m \rangle)$ , where the angle brackets represent the mean over the cloud population], and they implemented this statistical model in a convective parameterization (Plant and Craig 2008). Furthermore, Plant and Craig (2008) assumed that, in the vicinity of the cloud base, vertical velocities in deep convective clouds were independent of the cloud-base size. Then, the PDF  $\mathcal{P}_s(s)$  of deep convective cloud sizes was also exponential [since  $\mathcal{P}_m(m)\rho\overline{w'_p} = \mathcal{P}_s(s)$ ]:

$$\mathcal{P}_s(s) = \frac{1}{\langle s \rangle} \exp\left(\frac{-s}{\langle s \rangle}\right). \quad (1)$$

The uniformity of the mean vertical velocity at cloud base for precipitating clouds has also been reported in observational studies such as Warner (1970) and is sometimes assumed in convective parameterizations (Donner 1993; Cohen and Craig 2006). Thus, the exponential spectrum appears as a likely property of deep cloud sizes near the cloud base. In section 5e, it is shown that this property can also be extended to the large cumulus

clouds topping the thermal structures of the convective boundary layer.

The internal fluctuations of vertical velocity in clouds may also be considered. Emanuel (1991) recalls that pioneering aircraft measurements have shown that in-cloud fluctuations exhibit a typical length scale of 100 m. From dual-Doppler cloud radar data analysis, Damiani et al. (2006) suggests a typical length scale of  $L = 200\text{--}600$  m. Craig and Dörnbrack (2008) also give physical arguments supporting an  $L = 200\text{--}300$ -m length scale for variability. Malkus (1954) and Warner (1970) revealed that vertical velocity fluctuations were, at least, as large as the mean value across the cloud section.

Thus, studies suggest that both intrathermal (vertical velocity) and interthermal (cross section) fluctuations are important. Our aim is now to propose a corresponding theoretical representation of the boundary layer thermal plume field.

### 3. Data and methodology

#### a. Case description

The case investigated here is the African Monsoon Multidisciplinary Analyses (AMMA) case of 10 July 2006, where a small, short-lived convective cell developed over Niamey (Lothon et al. 2011). The whole transition was recorded by several ground-based instruments (radar, wind profiler, and atmospheric soundings) and completed by satellite data. This case study concerned a typical case of transition from shallow to deep convection over semiarid land with a high Bowen ratio ( $Bo \approx 10$ ) and associated with an elevated cloud

base ( $z_{\text{icl}} \approx 2.5$  km). The structure of the boundary layer clouds evolved gradually from a “cloud street” organization (from morning to noon) to an isotropic structure composed of larger but more heterogeneous cells (from noon to midafternoon). Around 1540 LT, deep convective cells developed with associated cold pools. It was noted by Lothon et al. (2011) that the first convective cells developed over the largest horizontal cloud structures, which supports the relevance of the cloud-base cross section in describing the transition process and reinforces the hypotheses made in section 1. A modeling setup was developed to represent this case and a large-eddy simulation, able to represent the main observed features, was run (Couvreur et al. 2012).

### b. The large-eddy simulation

The simulation uses the LES version of the Meso-NH nonhydrostatic model developed by Lafore et al. (1998). The domain is  $100 \times 100 \times 20$  km<sup>3</sup>, with a horizontal resolution of 200 m, a stretched grid on the vertical (from 50 to 250 m), and periodic lateral boundary conditions. The simulation lasts from 0600 to 1800 LT, at which time the cold pool generated by deep convection became too large relative to the domain. The lower boundary condition consists of imposed homogeneous surface latent and sensible heat fluxes. However, the observations showed a large positive surface temperature anomaly (around 5 K), over which the first cell developed (at 1540 LT). This heterogeneity is suspected of playing an important role in the triggering of deep convection (enhancing mesoscale circulation and breeze convergence over the hot spot; see Taylor et al. 2011). To simulate a similar onset of deep convection, a low-level moisture convergence is applied in the morning, linked to the monsoon flow, and a low-level ascent of  $1.5 \text{ cm s}^{-1}$  during the afternoon. With these conditions, the LES’s first cumulus appeared around 1100 LT and deep convection was triggered around 1630 LT. This simulation was evaluated against observations in Couvreur et al. (2012).

### c. Data, definitions, and notations

#### 1) CLOUDS AND THERMALS

The material used in the present study comprises various fields extracted from the simulation every hour from 1200 to 1800 LT. In the LES, a column  $i$  is defined as cloudy if the liquid water content  $r_c(i) \geq 10^{-6} \text{ kg kg}^{-1}$  in any vertical level  $k > k_{\text{min}}$  of a 200-m layer above  $k_{\text{min}}$ , where  $k_{\text{min}}$  is the lowest vertical level where the threshold  $r_c = 10^{-6} \text{ kg kg}^{-1}$  is reached in the LES domain. Within a cloudy column, the cloudy levels are those that verify  $r_c(i, k) + r_i(i, k) \geq 10^{-6} \text{ kg kg}^{-1}$ , where  $r_i(i, k)$  is the ice water content.

Adjacent cloudy columns (at cloudy levels) are then grouped to form individual clouds, described by (i) their cloud-base cross section  $s_i$ , (ii) their horizontal mean cloud-base altitude  $z_{\text{icl},i}$ , and (iii) their horizontal mean cloud-top altitude  $z_{\text{top},i}$ .

We also assume that each cloudy column corresponds to an individual draft that is grouped similar to the cloud to form a thermal originating from the surface and extending up to the cloud top  $z_{\text{top},i}$ . Then, in the LES, for each cloud, we define a corresponding thermal, which is the ensemble composed of the subcloud and the cloudy part of the adjacent cloudy columns. The whole analysis is founded on cloud-base characteristics. Therefore, we discard thermals that will not reach the LCL, that is, the dry thermals.

#### 2) THERMAL FIELD

The study domain corresponds to the extent of the horizontal area of the LES ( $S_d = 10^4 \text{ km}^2$ ), in which  $N_{\text{tot}}$  thermals (and corresponding clouds) are present, covering an area  $S_{\text{tot}}$  and a fractional area  $\alpha_{\text{tot}}$  ( $\alpha_{\text{tot}} = S_{\text{tot}}/S_d$ ) at the cloud-base level [or lifting condensation level (LCL)]. At a given level, the domain is divided into several regions: (i) the individual thermals ( $p_i, i = 1, N_{\text{tot}}$ ) and (ii) the thermal environment ( $e$ ). Generally, the overbar denotes the average over a horizontal region, which may be the whole domain or the thermal environment or a given thermal (e.g.,  $\bar{w}$ ,  $\bar{w}_e$ ,  $\bar{w}_{p,i}$ , and  $\bar{w}_p$  are the large-scale vertical velocity and the mean vertical velocities over the thermal environment, over thermal  $i$ , and over all thermals, respectively), while the angle brackets denote the arithmetic average over the population of thermals.

#### 3) THERMAL GEOMETRY

The geometry of a given thermal  $i$  is characterized by the altitudes  $z_{\text{icl},i}$  and  $z_{\text{top},i}$  of its cloud base and cloud top, respectively, and by its cross section  $s_i$  at cloud base. Since the LES horizontal resolution is 200 m, we arbitrarily assume, for simplicity, that the cross section of the elementary drafts is  $\bar{s} = 4 \times 10^4 \text{ m}^2$ . First, this length scale is consistent with the observational, high-resolution, and theoretical studies mentioned in section 2. Second, we will show in section 6 that this arbitrary parameter is of secondary importance.

A thermal  $i$  is then composed of  $n_i$  adjacent drafts of cross section  $\bar{s}$  underlying a cloud. The number of elementary drafts in a thermal  $i$  is noted  $n_i = s_i/\bar{s}$ . In the following, this quantity is named either the dimensionless cross section (i.e., the number of drafts per thermal) or the thermal size.

#### 4) VERTICAL VELOCITIES

For an air parcel located at horizontal position  $(x, y)$  and at altitude  $z$  inside thermal  $i$ , two decompositions of

the vertical velocity will be used. First, the usual decomposition in the domain average  $\bar{w}$  and a fluctuation  $w'_{p,i}$  yields  $w_{p,i}(x, y, z) = \bar{w} + w'_{p,i}(x, y, z)$ . Then, the fluctuation  $w'_{p,i}(x, y, z)$ , which is further decomposed into a thermal average  $\overline{w'_{p,i}}$  and a second-order fluctuation  $w''_{p,i}$ , yields

$$w_{p,i}(x, y, z) = \bar{w} + \overline{w'_{p,i}} + \overline{w''_{p,i}}(x, y, z). \quad (2)$$

A similar development gives, for any parcel located in the subsiding environment,

$$w_e(x, y, z) = \bar{w} + w'_e + w''_e(x, y, z). \quad (3)$$

### 5) MEAN PROPERTIES

From the individual thermal average vertical velocities, the mean vertical velocity of the whole thermal field can be computed as  $\overline{w'_p} = (1/S_{\text{tot}}) \sum_{i=1}^{N_{\text{tot}}} s_i \overline{w'_{p,i}}$ . Similarly, the mean second- and third-order noncentered moments are defined by

$$\overline{w'^2_p} = \frac{1}{S_{\text{tot}}} \sum_{i=1}^{N_{\text{tot}}} s_i \overline{w'^2_{p,i}} \quad \text{and} \quad \overline{w'^3_p} = \frac{1}{S_{\text{tot}}} \sum_{i=1}^{N_{\text{tot}}} s_i \overline{w'^3_{p,i}},$$

respectively.

For each thermal  $i$ , the vertical velocity standard deviation and skewness are

$$\Gamma_{w'_{p,i}} = \sqrt{\overline{w'^2_{p,i}} - \overline{w'_{p,i}}^2} \quad \text{and} \quad \Phi_{w'_{p,i}} = \frac{\overline{w'^3_{p,i}} - 3\overline{w'_{p,i}}\overline{w'^2_{p,i}} + 2\overline{w'_{p,i}}^3}{\Gamma^3_{w'_{p,i}}},$$

respectively.

The arithmetic-mean cross section over the thermal population gives  $\langle s \rangle = (1/N_{\text{tot}}) \sum_{i=1}^{N_{\text{tot}}} s_i$  and, similarly, the mean cloud base  $\langle z_{\text{icl}} \rangle = (1/N_{\text{tot}}) \sum_{i=1}^{N_{\text{tot}}} z_{\text{icl},i}$  and cloud-top altitudes  $\langle z_{\text{top}} \rangle = (1/N_{\text{tot}}) \sum_{i=1}^{N_{\text{tot}}} z_{\text{top},i}$ .

Finally, the arithmetic-mean thermal average velocity over the thermal population is defined as  $\langle \overline{w'_p} \rangle = (1/N_{\text{tot}}) \sum_{i=1}^{N_{\text{tot}}} \overline{w'_{p,i}}$  and, similarly, for the second- and third-order moments  $\langle \overline{w'^2_p} \rangle = (1/N_{\text{tot}}) \sum_{i=1}^{N_{\text{tot}}} \overline{w'^2_{p,i}}$  and  $\langle \overline{w'^3_p} \rangle = (1/N_{\text{tot}}) \sum_{i=1}^{N_{\text{tot}}} \overline{w'^3_{p,i}}$ .

### 6) DATA

For every thermal  $i$  we extract the cloud base  $z_{\text{icl},i}$  and the cloud-top altitudes  $z_{\text{top},i}$ . We then extract the following variables at cloud base: the thermal's dimensionless cross section  $n_i$  (or size in the following); the thermal's cross-sectional average of vertical velocity  $\overline{w'_{p,i}}$ ,

of its square  $\overline{w'^2_{p,i}}$ , and of its cube  $\overline{w'^3_{p,i}}$ ; and finally, the thermal's maximum vertical velocity  $w'_{\text{max},i}$ .

### d. Method

Our final goal is to propose a new formulation of ALE<sub>BL</sub>, or in other words, to compute a maximum kinetic energy provided by the thermals, which has to be compared with CIN. Thus, the following LES analysis is aimed at finding the maximum value distribution for the thermal cross sections and for the thermal vertical velocities, so that ALE<sub>BL</sub> can be computed.

Three different types of errors are computed all along the study. The first category (see Fig. 2) represents the systematic error on the PDF  $[N_n(\Delta n)]$  estimate in each bin ( $\Delta n$ ). The second category (see Figs. 3, 4) represents the fitting function error when using the least  $\chi^2$  method. The last category (see Figs. 4, 6) gathers systematic errors, either on the arithmetical mean computation or on the systematic error on a function computed from several independent mean variables.

Our starting hypotheses are (i) two-step triggering [as suggested by Chaboureau et al. (2004)] and (ii) that the cloud-base cross section plays a crucial role in controlling deep convection triggering (see Lothon et al. 2011).

## 4. LES analysis: Distribution of maximum cross section at the cloud base

### a. Cross-sectional spectrum: $\mathcal{P}(s)$

Mapes (2000), Khairoutdinov and Randall (2006), Rio et al. (2009), Grandpeix et al. (2010) and Del Genio and Wu (2010) have suggested that the subcloud layer processes play a key role in producing the dynamical forcing, which lifts the parcel from the surface layer to its LFC. In a conditionally unstable atmosphere, the LCL nearly corresponds to the top of the boundary layer and to the bottom of the CIN. We shall consider it as the most relevant level at which to represent the couplings between boundary layer processes and deep convection. Consequently, the present study focuses on the thermal properties at cloud base.

Figure 2 displays the  $N$ -normalized dimensionless cross-section spectrum ( $\mathcal{N}_n = N_{\text{tot}} \mathcal{P}$ , where  $\mathcal{P}$  is the PDF) at two different times using logarithmic coordinates. The range of cross sections is divided into bins of sizes that vary so that the number in each bin is sufficiently large for the statistical treatment to be applied. The spectra are fitted with double exponential PDFs using a least  $\chi^2$  method accounting for the statistical errors on the bin contents:

$$\mathcal{N}_n(n) = \frac{N_1}{n_1} \exp\left(\frac{-n}{n_1}\right) + \frac{N_2}{n_2} \exp\left(\frac{-n}{n_2}\right), \quad (4)$$

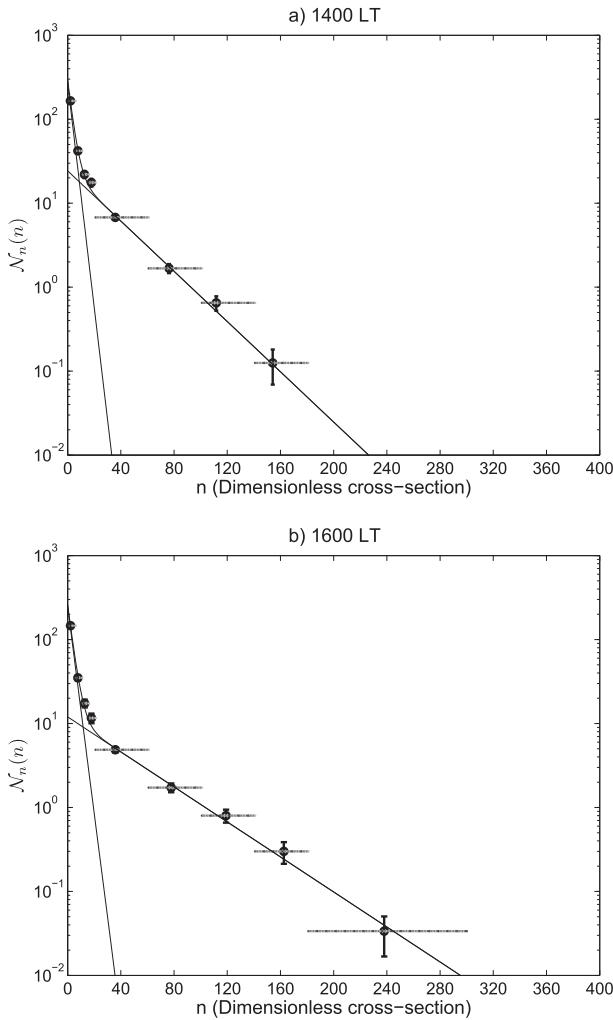


FIG. 2. The  $N$ -normalized dimensionless cross-sectional distribution [ $\mathcal{N}_n(n)$ , see Eq. (4)] of the thermals at the LCL for (a) 1400 and (b) 1600 LT. Horizontal lines indicate the dimensionless cross-sectional bins. Vertical lines are error bars  $\Delta\mathcal{N}_n(\Delta n)$  representing the systematic error on the  $\mathcal{N}_n(\Delta n)$  estimate in each bin  $\Delta n$ , which arises from the use of a Poisson distribution for the  $N(\Delta n)$  elements inside each bin  $\Delta n$ :  $\Delta\mathcal{N}_n(n) = \sqrt{\mathcal{N}(n)/\Delta n}$ .

where  $n_1 = S_1/\bar{s}$  and  $n_2 = S_2/\bar{s}$  are the average, dimensionless cross sections of each type of thermals and  $N_1$  and  $N_2$  represent their total population (with  $S_1$  and  $S_2$  as their average cross sections, in square meters).

If we consider the cross section  $s = n \times \bar{s}$  (instead of the dimensionless cross section  $n$ ), the  $N$ -normalized distribution becomes

$$\mathcal{N}_s(s) = \frac{N_1}{S_1} \exp\left(\frac{-s}{S_1}\right) + \frac{N_2}{S_2} \exp\left(\frac{-s}{S_2}\right). \quad (5)$$

This suggests the presence of two distinct thermal categories, whose distributions are exponential:

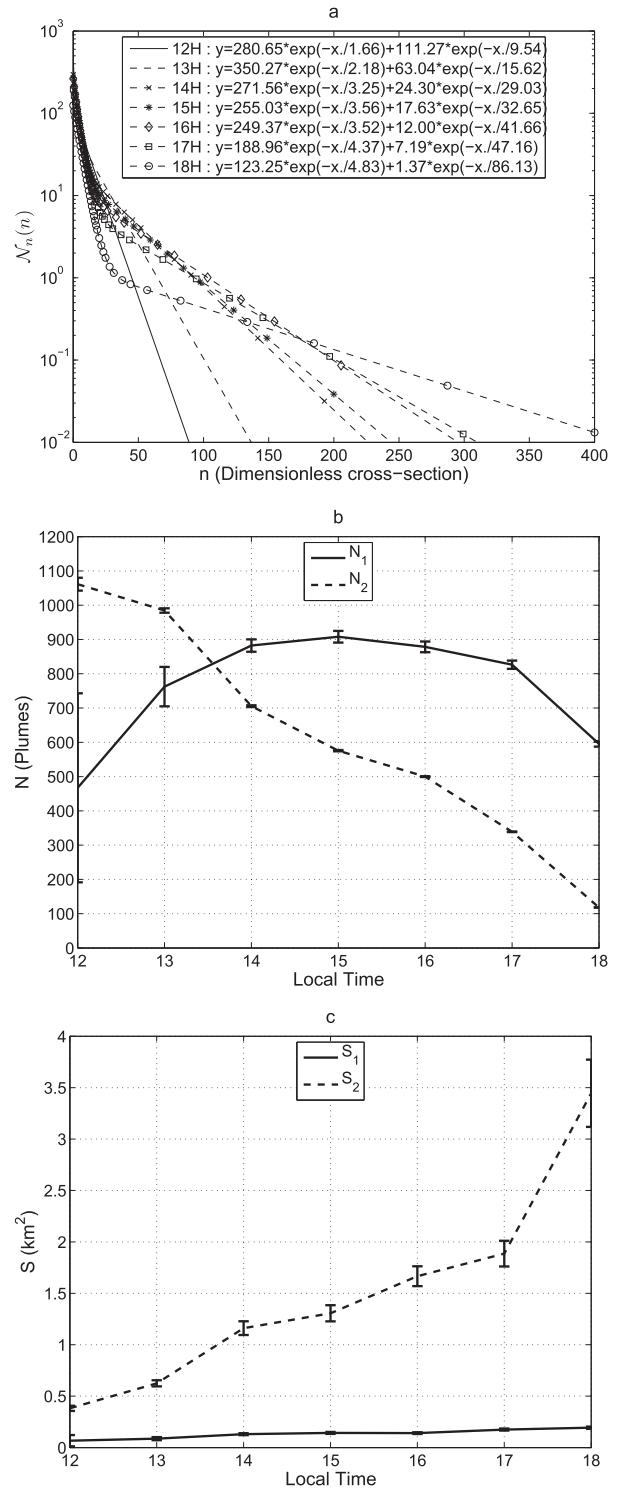


FIG. 3. (a) Time evolution of the  $N$ -normalized cross-sectional distribution  $\mathcal{N}_n(\Delta n)$  fitting function at LCL. (b) The  $N_1$  and  $N_2$  time series. (c) The  $S_1$  and  $S_2$  time series. Error bars represent the error on the estimate of  $N_1$ ,  $N_2$ ,  $S_1$ , and  $S_2$ , relative to the exponential fitting function.

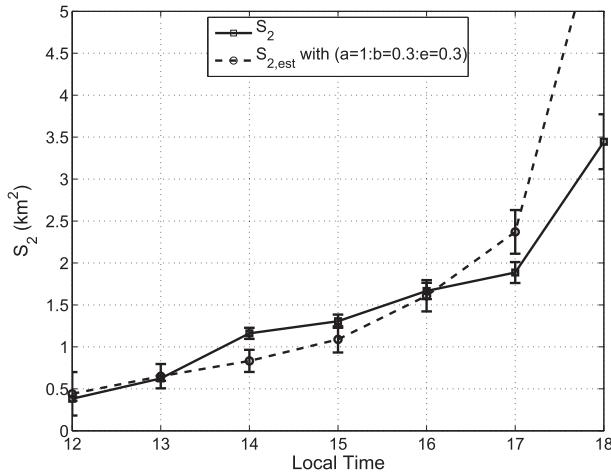


FIG. 4. Time evolution of the average cross section of type-2 thermals at cloud base  $S_2$  (km<sup>2</sup>) from LES (solid) and from  $S_2$  calculated following Eq. (8) with parameters:  $a = 1$  and  $b = 0.3$  (dashed). Vertical lines are error bars:  $\Delta S_2$  are the errors relative to the exponential fitting function and  $\Delta S_2$  are the uncertainty on the  $S_2$  calculations. According to Eq. (8),  $\Delta S_2(\langle z_{icl} \rangle, \langle z_{top} \rangle) = \{[(\partial S_2 / \partial \langle z_{icl} \rangle) \Delta \langle z_{icl} \rangle]^2 + [(\partial S_2 / \partial \langle z_{top} \rangle) \Delta \langle z_{top} \rangle]^2\}^{1/2}$ , where  $|\Delta \langle z_{icl} \rangle|$  and  $|\Delta \langle z_{top} \rangle|$  are the systematic errors on  $\langle z_{icl} \rangle$  and  $\langle z_{top} \rangle$ , respectively.

$$P_1(s) = \frac{1}{S_1} \exp\left(\frac{-s}{S_1}\right) \quad (6)$$

and

$$P_2(s) = \frac{1}{S_2} \exp\left(\frac{-s}{S_2}\right). \quad (7)$$

Category 1 gathers together a very large population of thermals topped by small cumulus clouds with cloud-base sizes (average value  $n = 3$ ) essentially ranging from  $n = 1$ –40 drafts (only one type-1 thermal is expected to have a size larger than 40; see Fig. 2). Their depths fluctuate between 50 and 500 m (not shown).

Category 2 concerns small and intermediate thermals accounting for the distribution tail (i.e., the right branch of the  $N$  PDF plotted in Fig. 2) with a cloud-base area ranging from  $n = 1$ –160 drafts (see Fig. 2) and depths fluctuating between 50 and 2000 m (not shown). The remaining class of clouds (not shown) is not represented by the fitting function given in Eq. (5) and concerns deep convective clouds (appearing after 1630 LT in the LES).

### b. Cross-sectional spectrum evolution

Figure 3a represents the  $N$ -normalized PDF evolution [defined in Eq. (5)] fitting the afternoon hours of the simulation (1200–1800 LT). The slope of the exponential distribution of type-2 thermals decreases with time,

while it does not seem to vary appreciably for type-1 thermals.

Figures 3b and 3c give further details on the evolution of each cloud population. Figure 3b shows that  $N_2$  decreases throughout the transition period. It is less trivial to extract a trend for population 1, as the error bars are very large at 1200 (only small clouds are present) and 1300 LT. At those times, populations 1 and 2 more or less overlap. On the other hand, according to Fig. 3c,  $S_2$  increases from 1200 up to 1800 LT. In other words, the transition from shallow to deep convection gives rise to fewer but larger thermals, suggesting that the gradual drying and deepening of the boundary layer (Lothon et al. 2011; Couvreux et al. 2012) is associated with larger cloud bases and deeper cumulus. Since  $N_2$  and  $S_2$  tendencies are of opposite signs, the fractional coverage  $\alpha_{tot}$  [as suggested by Chaboureau et al. (2004) through the NSD] is a priori not the best proxy for describing the transition process. The average cross section seems more pertinent for the transition. This result reinforces the relevance of considering spectral thermals rather than a bulk thermal and of treating both the thermal population and the cloud-base mean cross section independently. We also noted that the ratio between the surface covered by type-1 thermals and the total surface covered by thermals  $\epsilon = S_{tot,1}/S_{tot} = N_1 S_1 / \alpha_{tot} S_d$  was nearly constant and close to  $\epsilon = 0.3$ .

Since observations (Lothon et al. 2011) have suggested that the largest thermals are the key elements of the transition, we will study the statistical properties of the type-2 thermals only. However, since the cloud-base cross section is a variable that is absent from boundary layer parameterizations using the single thermal approach, we first need to establish empirical relationships between cloud-base cross section and vertical cloud development (that can be retrieved from any boundary layer parameterization).

### c. Vertical versus horizontal scale of type-2 clouds

The mean horizontal length scale of type-2 clouds at cloud base is  $\sqrt{S_2}$ , and their vertical length scale is given by their arithmetic-mean cloud-base  $\langle z_{icl} \rangle$  and cloud-top  $\langle z_{top} \rangle$  altitudes over the population  $N_2$ . Craig and Dörnbrack (2008) indicate that past theoretical studies aimed at representing a rising bubble in a neutrally stratified fluid yielded a linear solution between radius and height for thermals. Rio and Hourdin (2008) also point out that, in many LESs, boundary layer coherent structures have a typical aspect ratio of 2. Thus, deeper boundary layers should be associated with larger horizontal cloud size. Following this idea, we assume a linear relationship between  $\sqrt{S_2}$  and  $\langle z_{icl} \rangle$ . Moreover, from the LES data analysis performed on the AMMA case and

from two other (not shown) additional cases [i.e., the Barbados Oceanographic and Meteorological Experiment (BOMEX) case, which is an oceanic, steady, trade wind cumulus case in a subsiding atmosphere, and the Atmospheric Radiation Measurement (ARM) shallow case, which is a continental, shallow cumulus case], we conclude that  $\sqrt{S_2}$  is also sensitive to the mean cloud depth  $\langle z_{\text{top}} \rangle - \langle z_{\text{icl}} \rangle$ .

Consequently, we assume a linear relationship between the horizontal length scale, the cloud-base altitude, and the cloud-top altitude, giving an estimated  $\sqrt{S_2}$ :

$$\sqrt{S_2} = a(\langle z_{\text{top}} \rangle - \langle z_{\text{icl}} \rangle) + b\langle z_{\text{icl}} \rangle. \quad (8)$$

Coefficients  $a$  and  $b$  are determined by fitting  $\sqrt{S_2}(t)$ ,  $\langle z_{\text{top}} \rangle(t)$ , and  $\langle z_{\text{icl}} \rangle(t)$  at times  $t$  in the range 1200–1600 LT (i.e., before deep convection triggers) in the AMMA case. The resulting pair  $[a:b]$  can be approximated by

$$a = 1.5 \pm 0.8 \quad \text{and} \quad b = 0.25 - 0.1(a - 1.5).$$

The quality of the fit with parameters  $a = 1$  and  $b = 0.3$  is shown in Fig. 4, where the time evolution of both  $S_2$  and its estimation,  $S_2$  from Eq. (8), are displayed. The difference between the two variables is within two standard deviations during the whole transition period (from 1200 to 1700 LT). Another important result is that neither  $a$  nor  $b$  is compatible with zero (at two standard deviations for  $a$  and more than three for  $b$ ). Hence, both the dependence on the boundary layer height and the cloud thickness are necessary to determine the cloud-base cross section.

The dependency of the thermal width at cloud base on the cloud-base altitude is consistent with the common hypothesis of a nearly constant aspect ratio (around 2, according to Rio and Hourdin 2008) for the boundary layer thermal structures. However, the issue of the potential mechanism(s) explaining the matching horizontal and vertical cumulus growth is trickier to address. The increase of cumulus buoyancy along the transition reported by Wu et al. (2009) may be associated with the decreasing lateral entrainment rate due to the increasing cloud width. Then, the entrainment process provides an explanation of how the cloud width increase causes the cloud-depth increase, but it does not reveal anything about how cloud depth feeds back onto cloud width. We suggest two potential mechanisms here, based on two diabatic processes, which may explain how cloud height influences cloud width.

The first one is diabatic cooling by rain evaporation. High-resolution simulations (Khairoutdinov and Randall 2006; Matheou et al. 2011; Boing et al. 2010) have

shown that density currents induced by cumulus rain re-evaporation often appear before deep convection onset and play a key role in the transition from shallow to deep convection. They suppress convection in their core and favor it on their edges by lifting the surrounding unstable air, in particular where colliding density currents result in sparser but stronger updrafts. They tend to feed deeper and broader clouds as cold pools grow. The second mechanism is diabatic heating by condensation. Clark et al. (1986) assert that midlevel cumulus cloud heating can trigger gravity waves, which reflect on the tropopause and feed back onto the lower levels, selecting eddies having horizontal length scales comparable with the gravity wave spacing. Such a mechanism would operate a scale selection on thermal eddies and favor sparser and larger horizontal structures during the transition.

#### d. Maximum cross-sectional distribution: $\mathcal{P}_{\text{max}}(S_{\text{max}})$

Since we assume that the triggering occurs over the largest cloud of the domain, we now look at the maximum cross-sectional distribution of type-2 thermals. As shown in the appendix, the cross section  $S_{\text{max}}$  of the largest thermal is a random variable with complementary cumulative distribution function (CCDF)  $\mathcal{F}_{\text{max}}(S_{\text{max}})$  given in Eqs. (A1) and (A6), from which we can derive a PDF  $\mathcal{P}_{\text{max}}(S_{\text{max}})$  that verifies Eq. (A2):

$$\mathcal{P}_{\text{max}}(S_{\text{max}}) = \frac{-d\mathcal{F}_{\text{max}}(S_{\text{max}})}{dS_{\text{max}}}. \quad (9)$$

The median  $S_{\text{max}}$  of the  $S_{\text{max}}$  distribution is given by the approximate formula of Eq. (A8) with  $\pi_t = \ln(2)$ :

$$S_{\text{max}} = S_2 \ln\left(\frac{N_2}{\pi_t}\right). \quad (10)$$

The maximum value PDF  $\mathcal{P}_{\text{max}}(S)$  and  $S_{\text{max}}$  values estimated at various times are plotted in Fig. 5. The graph of  $\mathcal{P}_{\text{max}}(S)$  in Fig. 5a, confirms that the distribution tail increases while transition occurs. Larger structures appear in the domain but still coexist with numerous small ones. As a result, the cross-sectional spectrum widens and goes toward higher  $S$ . During the early afternoon (between 1200 and 1300 LT), the graph of  $\mathcal{P}_{\text{max}}(S)$  is relatively peaked, and accordingly,  $S_{\text{max}}$  fits the simulated values ( $S_{\text{max}}$ ) well (see Fig. 5b). Then, later on, the error  $|S_{\text{max}} - S_{\text{max}}|$  increases with time, as predicted by the spectrum widening (except at 1500 LT for which, by chance, the estimated value is almost equal to the simulated one). This result shows that the exponential law  $\mathcal{P}_2(s)$  provides, at least, a good approximation of the maximum cross-sectional distribution of the domain. However, this is no longer true after the onset of deep convection;

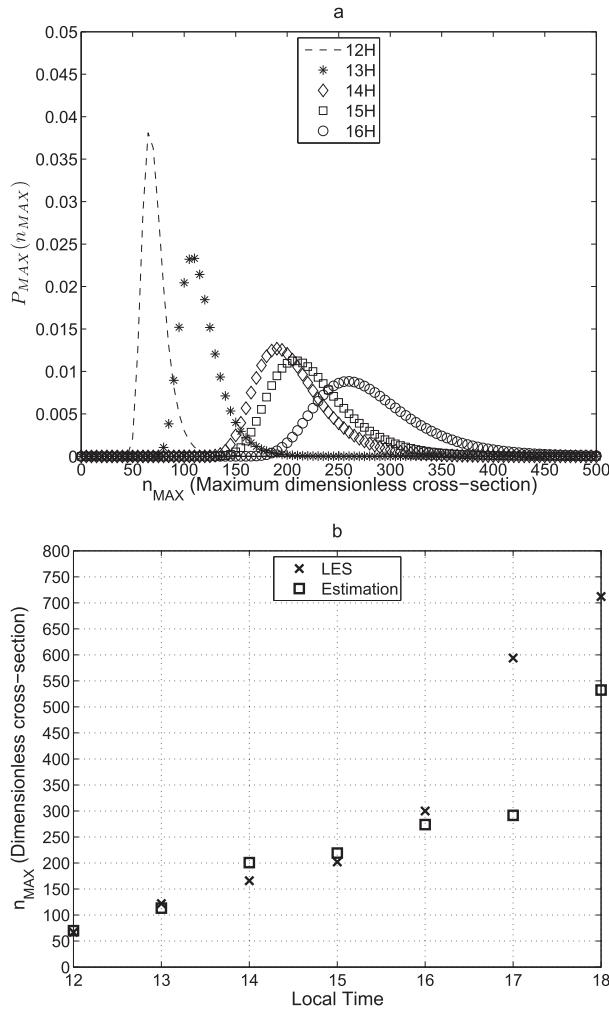


FIG. 5. (a) The  $P_{max}(S_{max})$  time evolution from 1200 to 1600 LT. (b) Time series of the estimated maximum cross section  $S_{max}$  (squares) and simulated  $S_{max}$  (crosses) from 1200 to 1800 LT.

and the error  $|S_{max} - S_{max}|$  increases considerably from 1700 LT (see Fig. 5b).

Therefore, from the PDF  $\mathcal{P}_2(s)$ , we have extracted an estimator  $S_{max}$  of the maximum cross section of the

thermal population, over which the first cumulonimbus has the greatest probability of appearing. Since we are trying to compute  $ALE_{BL}$ , which is the maximum kinetic energy over the thermal population, our goal is then to estimate the statistical maximum vertical velocity corresponding to the widest thermal of the field.

### 5. LES analysis: Distribution of maximum vertical velocities of type-2 thermal

#### a. Method

In this section, the objective is to compute a statistical maximum velocity for the type-2 thermals. For that purpose, the whole LES simulation (i.e., from 1200 to 1800 LT) is grouped together in a single dataset of 9500 thermals. Then, only the thermals with sizes exceeding  $n = 40$  drafts (i.e., diameters exceeding 1500 m) are kept to give a final dataset of 900 thermals, almost exclusively type 2. Finally, this dataset is divided into 10 samples sorted by increasing cross section.

For each sample  $k$ , characterized by its  $n$  range and composed of  $N_{tot,k} = 90$  clouds, Table 1 shows the arithmetic means  $\langle \cdot \rangle_k$  over the clouds of various fields defined at cloud base in section 5c: (i) the average vertical velocity  $w'_{p,i}$ , (ii) the second- and third-order non-centered moments  $w'^2_{p,i}$  and  $w'^3_{p,i}$ , (iii) the maximum velocity  $w'_{max,i}$ , and (iv) the standard deviation  $\Gamma_{w'_{p,i}}$  and the skewness  $\Phi_{w'_{p,i}}$ .

#### b. Vertical velocity moments

In an attempt to characterize the vertical velocity distribution inside the thermals, we first look at the sensitivity of the vertical velocity moments to the mean cross section of each sample. Figure 6 displays the pairs  $[\langle n \rangle_k; \langle w'_p \rangle_k]$ , and  $[\langle n \rangle_k; \langle w'^3_p \rangle_k]$ . From Fig. 6 and Table 1, it seems that the sample-mean velocities  $\langle w'_p \rangle_k$ ,  $\langle w'^2_p \rangle_k$ , and  $\langle w'^3_p \rangle_k$  do not vary with the sample-mean dimensionless cross section  $\langle n \rangle_k$ . Hence, whatever the sample  $k$ ,  $\langle w'_p \rangle_k = \langle w'_p \rangle$ ,  $\langle w'^2_p \rangle_k = \langle w'^2_p \rangle$ , and  $\langle w'^3_p \rangle_k = \langle w'^3_p \rangle$ , where

TABLE 1. Mean dynamical characteristics of the 10 thermal samples of category 2.

Sample $k$	$n$ range (drafts)	$\langle n \rangle_k$ (drafts)	$\langle w'_p \rangle_k$ (m s <sup>-1</sup> )	$\langle w'^2_p \rangle_k$ (m s <sup>-1</sup> )	$\langle w'^3_p \rangle_k$ (m s <sup>-1</sup> )	$\Gamma_{w'_p}$ (m s <sup>-1</sup> )	$\Phi_{w'_p}$	$\langle w'_{p,max} \rangle_k$ (m s <sup>-1</sup> )
1	40:43	41.0	1.04 ± 0.06	2.67 ± 0.21	7.74 ± 0.93	1.08 ± 0.04	0.26 ± 0.05	3.35 ± 0.15
2	43:47	44.7	0.91 ± 0.05	2.23 ± 0.18	5.98 ± 0.78	1.01 ± 0.04	0.23 ± 0.05	3.14 ± 0.13
3	47:51	49.0	1.11 ± 0.06	2.89 ± 0.21	8.53 ± 0.96	1.11 ± 0.04	0.14 ± 0.06	3.66 ± 0.14
4	51:57	53.8	1.04 ± 0.05	2.63 ± 0.17	7.42 ± 0.75	1.12 ± 0.04	0.22 ± 0.04	3.79 ± 0.15
5	57:63	60.1	1.07 ± 0.05	2.60 ± 0.18	7.02 ± 0.75	1.06 ± 0.04	0.13 ± 0.04	3.59 ± 0.14
6	63:71	66.8	1.07 ± 0.04	2.44 ± 0.15	6.25 ± 0.58	1.04 ± 0.03	0.19 ± 0.04	3.53 ± 0.12
7	71:84	77.2	1.07 ± 0.05	2.59 ± 0.16	6.65 ± 0.63	1.08 ± 0.03	0.16 ± 0.04	3.81 ± 0.11
8	85:103	99.5	1.11 ± 0.04	2.80 ± 0.15	7.85 ± 0.70	1.16 ± 0.03	0.24 ± 0.05	4.08 ± 0.13
9	103:137	116.9	1.02 ± 0.04	2.57 ± 0.13	6.55 ± 0.57	1.14 ± 0.03	0.21 ± 0.04	3.97 ± 0.12
10	138:430	215.9	1.08 ± 0.04	2.72 ± 0.12	6.64 ± 0.48	1.17 ± 0.03	0.06 ± 0.05	4.23 ± 0.10

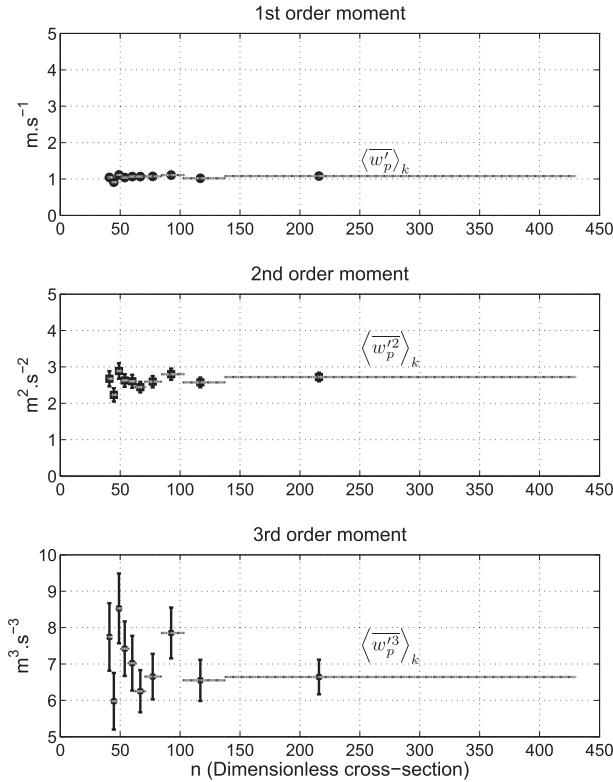


FIG. 6. (top to bottom) Scatterplots of sample mean  $\langle w'_p \rangle_k$ , second-order  $\langle w'^2_p \rangle_k$ , and third-order noncentered moments  $\langle w'^3_p \rangle_k$  of cloud-base vertical velocity, as a function of the dimensionless cross section ( $n$ ). Horizontal lines mark the  $n$  bins and vertical lines indicate systematic errors  $|\Delta \langle w'_p \rangle_k|$ ,  $|\Delta \langle w'^2_p \rangle_k|$ , and  $|\Delta \langle w'^3_p \rangle_k|$  (they are very small and hardly visible for the first- and second-order moments).

$\langle w'_p \rangle$  is the arithmetic average over the 10 samples taken together. Extending this result to the individual thermal scale shows that the thermal averages  $w'_{p,i}$ ,  $w'^2_{p,i}$ , and  $w'^3_{p,i}$  are also insensitive to the cross section  $s_i$  of the thermal considered, which also means that all the thermals of the domain have the same mean vertical velocity  $w'_{p,i}$ , standard deviation  $\Gamma_{w'_{p,i}}$ , and skewness  $\Phi_{w'_{p,i}}$ .

Hence, whatever the thermal  $i$ ,

$$\overline{w'_{p,i}} = \overline{w'_p}, \quad (11)$$

$$\overline{w'^2_{p,i}} = \overline{w'^2_p}, \quad (12)$$

and

$$\overline{w'^3_{p,i}} = \overline{w'^3_p}. \quad (13)$$

In other words, the vertical velocity spectrum is uniform over the thermal field. This means that, at a given time, all the elementary drafts underlying the type-2 clouds of the domain exhibit the same vertical velocity distribution  $\mathcal{P}(w'_p)$ .

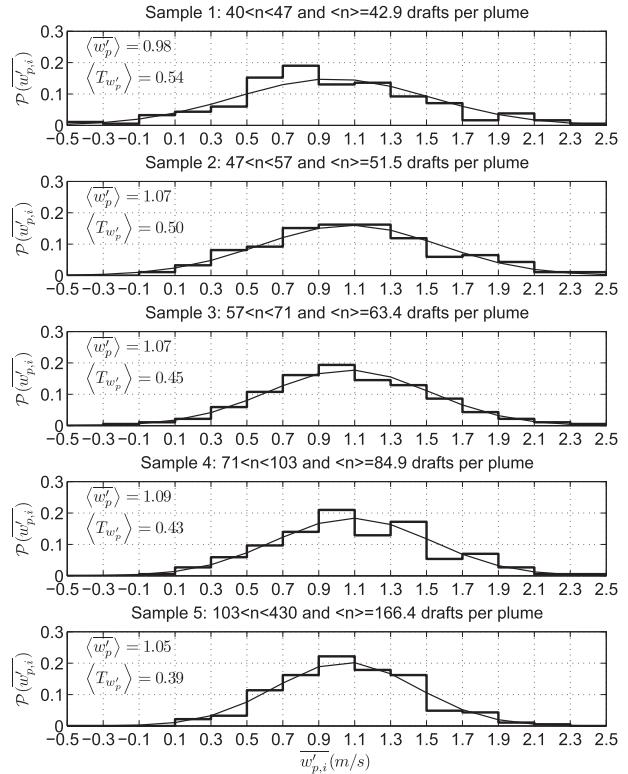


FIG. 7. Normalized histogram of  $w'_{p,i}$  and fitting PDF  $\mathcal{P}(w'_{p,i})$  for (top to bottom) samples 1–5. The sample mean cross-sectional-averaged vertical velocity  $\langle w'_p \rangle$  and standard deviation  $(\Gamma_{w'_p})$  are noted in the upper left corner of each panel.

### c. PDF of draft vertical velocities $\mathcal{P}(w'_{p,i})$

To increase the statistical significance, we decide to temporarily divide (only in this subsection) the dataset into only five samples of 180 thermals each, sorted by increasing cross sections. Figure 7 displays the five corresponding histograms of the thermal mean vertical velocities at cloud base  $w'_{p,i}$ . The distributions  $\mathcal{P}(w'_{p,i})$  in the various samples look very close to Gaussians (Table 1 shows very low skewnesses  $\langle \Phi_{w'_p} \rangle_k$ , with widths  $\Gamma_{w'_{p,i}}$  roughly proportional to  $4/\sqrt{n}$ ). Hence, the hypothesis that the elementary drafts are independent Gaussian random variables can be taken, and their velocity distribution is then

$$\mathcal{P}(w'_{p,i}) = \frac{1}{\sqrt{2\pi}\Gamma_{w'_{p,i}}} \exp \left[ -\frac{(w'_{p,i} - \overline{w'_{p,i}})^2}{2\Gamma_{w'_{p,i}}^2} \right]. \quad (14)$$

### COMPLEMENTARY COMMENTS

#### (i) Reference cross section $\tilde{s}$ of the drafts

The fact that  $\Gamma_{w'_{p,i}} \approx 4/\sqrt{n}$  suggests that independent drafts may have a typical cross section  $\tilde{s}' = 4\tilde{s} = 1.6 \times 10^5 \text{ m}^2$ , that is, a length scale of  $l = 400 \text{ m}$ . This result is in

good agreement with earlier studies mentioned in section 2, but twice as large as the arbitrary  $l = 200$  m chosen in this study, in accordance with the LES grid spacing. However, considering a length scale ranging from  $l = 200$  to  $l = 400$  m does not affect the course of this study much (not shown).

(ii) Vertical velocity mean and standard deviation

From Table 1 (columns 4 and 7), it can be seen that the mean and the standard deviation of the cloud-base vertical velocity  $\langle w'_p \rangle_k$  and  $\langle \Gamma_{w'_p} \rangle_k$  are very similar. Therefore, we shall assume that the thermal mean vertical velocity and standard deviation are equivalent at cloud base:

$$\overline{w'_p} = \Gamma_{w'_p}. \tag{15}$$

This assumption has already been used by Grandpeix and Lafore (2010) and Grandpeix et al. (2010) and will also be used in the stochastic parameterization for deep convection triggering presented in Rochetin et al. (2014, hereafter Part II).

d. Maximum vertical velocity distribution

$$\mathcal{P}_{\max}(w'_{\max,i})$$

Since we have characterized the draft velocity spectrum  $\mathcal{P}(w'_{p,i})$ , the next step is to look for the maximum-value distribution of  $w'_{p,i}$ , given a thermal  $i$  made up of  $n_i$  independent drafts (of cross section  $s_i = n_i \delta$ ). Following the same method as in section 4d, and according to the appendix, it is possible to retrieve a distribution law for the maximum values  $\mathcal{P}_{\max}(w'_{\max,i})$  [see Eq. (A2)] from the vertical velocity PDF and to compute an estimator  $W'_{\max,i}$  [see Eq. (A11)] of the maximum velocity at cloud base such as the median value. Introducing this information into Eq. (A11), taking into account the uniformity of  $\overline{w'_p}$  and  $\Gamma_{w'_p}$  [Eqs. (11), (12)], averaging over each sample  $k$ , and assuming that, in every sample  $k$ , we have  $n_i \approx \langle n \rangle_k$  yields the following median maximum velocity for every sample  $k$ :

$$\langle W'_{p,\max} \rangle_k = \overline{w'_p} + \Gamma_{w'_p} \sqrt{\ln \left( \frac{\langle n \rangle_k^2}{2\pi\pi_i^2} \right) - \ln \left[ \ln \left( \frac{\langle n \rangle_k^2}{2\pi\pi_i^2} \right) \right]}. \tag{16}$$

Figure 8a displays the sensitivity of the maximum-value PDF  $[\mathcal{P}_{\max}(w'_{\max,i})]$  of a thermal  $i$  to its size  $s_i$ . The PDF is relatively peaked and thin in all cases, becoming slightly narrower as the cloud-base area increases. Consequently, the estimator  $W'_{\max,i}$  is expected to give a good approximation of the simulated  $w'_{\max,i}$ . Figure 8b compares

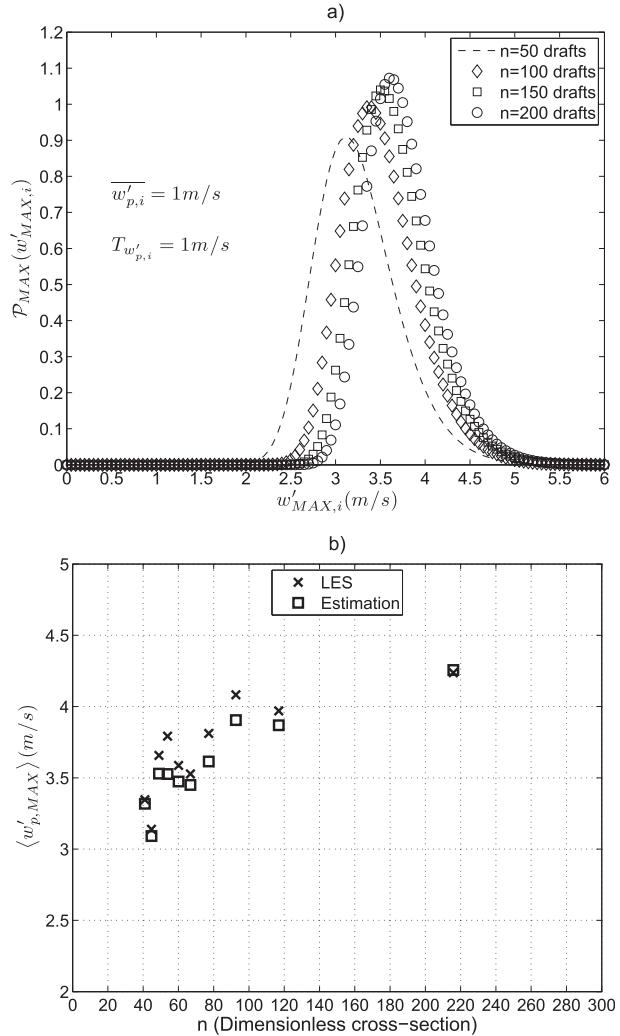


FIG. 8. (a) The  $\mathcal{P}_{\max}(w'_{\max,i})$  sensitivity to the thermal  $i$  cross section  $s_i$  with  $w'_{p,i} = 1 \text{ m s}^{-1}$  and  $\Gamma_{w'_{p,i}} = 1 \text{ m s}^{-1}$ . (b) Scatterplot of the estimated  $\langle W'_{p,\max} \rangle_k$  (squares) and simulated  $\langle w'_{p,\max} \rangle_k$  (crosses) maximum velocity at the cloud base vs cross section.

the pairs  $[\langle w'_{p,\max} \rangle_k : \langle n \rangle_k]$  and  $[\langle W'_{p,\max} \rangle_k : \langle n \rangle_k]$  for each sample  $k$ . The sample-mean estimator  $\langle W'_{p,\max} \rangle_k$  is very close to the sample-mean maximum  $\langle w'_{p,\max} \rangle_k$  extracted from LES data. First, this result strongly supports the hypothesis of considering elementary drafts (of  $l \approx 200$  m) to be independent and Gaussian, especially for the high velocities. Of course, it is worth noting that the independence of drafts is only valid when a reasonable minimal length scale  $l_{\min}$  is considered. Second, the maximum velocity encountered does not depend on anything other than the thermal size: the more drafts there are at the cloud base, the greater the probability is of finding a strong one. The analytical formula given in Eq. (16) does not take into account of entrainment–detrainment mixing or any other physical process and

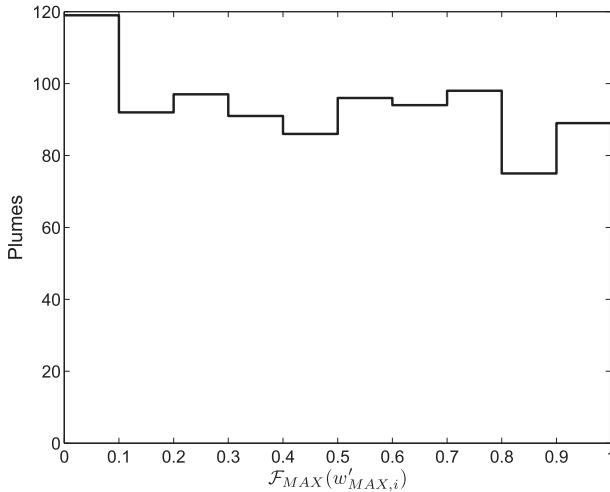


FIG. 9. Histogram of the CCDF of  $w'_{\max,i}$  [ $\mathcal{F}_{\max}(w'_{\max,i})$ ] for the cloud base of type 2. The horizontal axis represents the CCDF  $\mathcal{F}_{\max}(w'_{\max,i})$  for each thermal  $i$  and the vertical axis gives the number of thermals in each bin (0.1).

does not tell us anything about the location of the strongest draft inside the thermal.

Of course, one can expect this particular draft to have a greater chance of being close to the thermals core, in which the rising parcels have less chances of mixing with surrounding dry air via lateral entrainment. Nevertheless, the present result shows that we can easily compute the maximum velocity in a thermal at the cumulus base, just by knowing its width.

This concordance between simulated and calculated maximums finally shows that the tail of the Gaussian distribution of the 200-m draft velocities [ $\mathcal{P}(w'_{p,i})$ ] in each thermal  $i$  is relevant. Since our concern is the triggering of deep convection, we focus on high velocities, and we need to make sure that the hypothesis of independent Gaussian drafts is, at least, robust for the distribution tail. One way to do that is to plot the histogram of the CCDF of  $w'_{\max,i}$  [ $\mathcal{F}(w'_{\max,i})$ ], that is, the probability of having a larger value than  $w'_{\max,i}$  for each

thermal  $i$ ] given in Eq. (A1). To do that, for each thermal  $i$  (of type 2) we compute the CCDF  $\mathcal{F}_{\max}(w'_{p,i})$ , then we distribute them in bins of 0.1, and then we can plot the  $\mathcal{F}_{\max}(w'_{\max,i})$  histogram displayed in Fig. 9. The flat distribution shows that the simulated  $w'_{\max,i}$  of each thermal  $i$  is equally distributed on both sides of the PDF of the maximum velocities  $\mathcal{P}_{\max}(w'_{\max,i})$ . This proves that, at least for the tail of  $\mathcal{P}(w'_{p,i})$ , (i) the hypothesis of the independent drafts is robust and (ii) the Gaussian PDF is pertinent also.

### e. Summary

To sum up, the dynamical properties of the type-2 thermals are uniform over the thermal field. Since the cross-sectional spectrum for type-2 thermals is exponential, this result is somehow consistent with the exponential distribution for individual mass fluxes proposed by Plant and Craig (2008). Moreover, each thermal can be considered as being composed of independent drafts (i.e., with no spatial coherence) of typical length scale  $l \approx 200$  m, following a Gaussian distribution for the vertical velocities, in which the average is practically equivalent to the standard deviation. Finally, the Gaussian distribution effectively describes the maximum-value statistics, which mostly depend on the cloud-base cross section.

## 6. Statistical available lifting energy ( $\text{ALE}_{\text{BL,stat}}$ )

### a. $\text{ALE}_{\text{BL,stat}}$ computation

The statistical  $\text{ALE}_{\text{BL,stat}}$  corresponds to the maximum kinetic energy found over the thermal population. From sections 4 and 5, respectively, we have extracted a median value  $S_{\max}$  for the maximum cross section [Eq. (10)] and a median value  $\langle W'_{p,\max} \rangle$  for the maximum vertical velocity [Eq. (16)] of a thermal sample. Then, by combining Eq. (10) with Eq. (16) and introducing Eq. (15), we obtain a statistical maximum velocity inside the largest thermal:

$$W'_{\max} = \overline{w'_p} \left[ 1 + \sqrt{\ln \left\{ \frac{\left[ \frac{S_2}{\bar{s}} \ln \left( \frac{N_2}{\pi_i} \right) \right]^2}{2\pi\pi_i^2} \right\} - \ln \left( \ln \left\{ \frac{\left[ \frac{S_2}{\bar{s}} \ln \left( \frac{N_2}{\pi_i} \right) \right]^2}{2\pi\pi_i^2} \right\} \right)} \right]. \quad (17)$$

We recall that the arbitrary value  $\bar{s}$  had a limited influence on  $W'_{\max}$ . Hence, if  $N_2$ ,  $S_2$ , and  $\overline{w'_p}$  are known, we can finally compute the statistical maximum kinetic energy at cloud base:

$$\text{ALE}_{\text{BL,stat}} = \frac{1}{2} W'^2_{\max}. \quad (18)$$

Figure 10 shows the time evolution of  $\text{ALE}_{\text{BL,stat}}$ . It is maximum around 1300 LT and decreases later on.

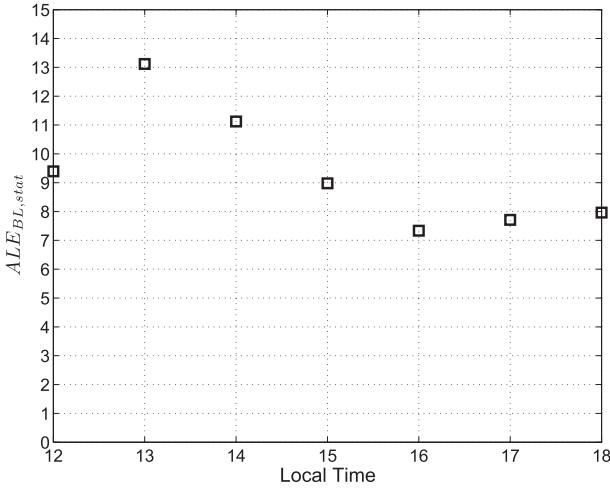


FIG. 10. Time series of  $ALE_{BL,stat}$  ( $m^2 s^{-2}$ ).

Actually,  $\mathcal{W}'_{max}$  is approximately in phase with  $\overline{w'_p}$  (not shown), which is itself correlated with the sensible heat flux (not shown). Although the maximum cross section  $\mathcal{S}_{max}$  is about twice as large at 1600 LT as at 1300 LT (see Fig. 5b), the surface heating is less and, consequently, the mean velocity of the thermal population  $\overline{w'_p}$  is around 30% lower (not shown). This correlation between  $\mathcal{W}'_{max}$  and  $\overline{w'_p}$  can be explained through Eq. (17), from which we can reasonably presume that the sensitivity of  $\mathcal{W}'_{max}$  to  $\overline{w'_p}$  variations is much greater than its sensitivity to  $\mathcal{S}_{max}$  variations:

$$\left(\frac{\partial \mathcal{W}'_{max}}{\partial \overline{w'_p}}\right) \mathcal{S}_{max} d\overline{w'_p} \gg \left(\frac{\partial \mathcal{W}'_{max}}{\partial \mathcal{S}_{max}}\right) \overline{w'_p} d\mathcal{S}_{max}$$

Thus, during the transition phase, the  $\overline{w'_p}$  decrease dominates the  $\mathcal{S}_{max}$  increase.

According to the LES, the morning-time large-scale inhibition is very high, and  $ALE_{BL,stat}$  reaches the CIN (not shown) around 1300 LT. Therefore, since both observational (Lothon et al. 2011) and LES (Couvreur et al. 2012) data show that deep convection is triggered near 1600 LT, the dynamical threshold  $ALE_{BL,stat} > |CIN|$  alone is not sufficient to describe the whole transition process.

*b. Toward a new triggering formulation*

Lothon et al. (2011) mentioned that, around 1200–1300 LT, the boundary layer moved from a regular, steady cloud-street organization to a more isotropic structure with bigger clouds. This could correspond to the beginning of the transition stage. Thus, although  $ALE_{BL} > |CIN|$  is apparently not a pertinent threshold for deep-convection triggering, it may be relevant for describing the threshold for moving from a shallow

cumulus regime to a transition regime. In the shallow cumulus regime, no clouds cross the inhibition layer. In the transition regime, many cumulus clouds have enough kinetic energy to overshoot the CIN, but are still too small to reach the high troposphere. So, we shall impose a complementary constraint on the size of the thermal to permit the triggering of deep convection.

**7. Deep convection triggering formulation**

In the current version of the LMDZ model, deep convection triggering by boundary layer thermals is exclusively based on the threshold condition  $ALE_{BL} > |CIN|$ . As the associated thermal representation is deterministic, either no thermals trigger, or all the thermals trigger, deep convection. However, since a thermal spectrum is considered here, we can, a priori, expect to have both passive boundary layer cumulus clouds and overshooting clouds in a given domain. As already mentioned, the thermal size appears to be of primary importance in the triggering process; Lothon et al. (2011) noticed that the first deep convective cells occurred over a zone covered by the largest horizontal structures of the observed domain. Chaboureau et al. (2004) also stressed the existence of two-step triggering, in which a transition phase clearly appeared.

Hence, the main idea of the triggering formulation is that the thermal field must require (i) at least one thermal with a maximum kinetic energy exceeding the CIN, which means  $ALE_{BL,stat} > |CIN|$  and (ii) a sufficient number of thermals having sizes that may potentially exceed a certain threshold value  $S_{trig}$ . This threshold corresponds to an arbitrary limit after which the associated cloud no longer corresponds to a cumulus but to a congestus or a cumulonimbus cloud, whose top nearly reaches the freezing level (see section 3b in Part II for more details). It might be expected that the largest thermal size should grow gradually up to the time when it reaches this threshold.

Let  $S_{trig}$  be the threshold value for deep convection triggering, and assume that  $t_0$  corresponds to the instant when  $ALE_{BL,stat} > |CIN|$ . The triggering probability  $P_\tau$  for one thermal scene of duration  $\tau$  (i.e., an average picture of the thermal field over the period  $\tau$ ), composed of  $N_2$  thermals, is the probability that  $S_{max} > S_{trig}$ , which is the CCDF  $\mathcal{F}_{max}(S_{trig})$  given in Eqs. (A1) and (A6):

$$P_\tau = \mathcal{F}_{max}(S_{trig}) = 1 - [1 - \hat{\mathcal{F}}(S_{trig})]^{N_2}.$$

The no-triggering probability is then

$$\hat{P}_\tau = \hat{\mathcal{F}}_{max}(S_{trig}) = 1 - P_\tau,$$

giving, for every independent cloud scene of duration  $\tau$ ,

$$\hat{P}_\tau = \left[ 1 - \exp\left(\frac{-S_{\text{trig}}}{S_2}\right) \right]^{N_2}. \quad (19)$$

The no-triggering probability definition  $\hat{P}_\tau$  can be generalized to every time period  $\Delta t = n \times \tau$ , composed of  $n$  independent scenes of duration  $\tau$ :

$$\hat{P}_{\Delta t} = \prod_{k=1}^n \hat{P}_\tau.$$

A continuous formulation (i.e., whatever  $\Delta t > \tau$ ) of the no-triggering probability  $\hat{P}_{\Delta t}$ , which verifies  $\hat{P}_{\Delta t} = \hat{P}_\tau$  if  $\Delta t = \tau$ , is

$$\hat{P}_{\Delta t} = (\hat{P}_\tau)^{\Delta t/\tau}.$$

It can be shown that this formula holds also for time periods  $\Delta t < \tau$ , provided one makes the strong assumption that the cumulus clouds that may potentially trigger are sparse enough in the field to be considered as independent, even in a time period shorter than  $\tau$ .

When combined with Eq. (19), this yields

$$\hat{P}_{\Delta t} = \left\{ \left[ 1 - \exp\left(\frac{-S_{\text{trig}}}{S_2}\right) \right]^{N_2} \right\}^{\Delta t/\tau}. \quad (20)$$

Thus, during every time period  $\Delta t$ , we can compute a no-triggering probability  $\hat{P}_{\Delta t}$ .

The point is now to determine what a reasonable estimation could be for  $\tau$ . During a period  $\Delta t$ , the mean cumulus lifetime directly influences the correlation between the two consecutive scenes at  $t$  and  $t + \Delta t$ ; thus, it could be reasonable, as a first approximation, to consider  $\tau$  as the mean cumulus lifetime. Several LES studies have investigated this issue in various contexts (oceanic trade wind cumulus case, continental cumulus, etc.). Considering the studies by Zhao and Austin (2005), Heus et al. (2009), and Seifert and Heus (2013), the cumulus lifetime basically ranges from 1000 to 2000 s.

However, here  $\tau$  represents a decorrelation time between two cumulus scenes only for type-2 thermals. We used 5-min interval snapshots from 1100 to 1530 LT in the LES to reach an estimation of this decorrelation time. We found an increase of the decorrelation time  $\tau$  along the simulation, starting from  $\tau = 1000$  s between 1100 and 1200 LT and going to  $\tau = 1800$  s between 1400 and 1530 LT. Therefore, we shall consider the range  $1000 < \tau < 2000$  s as a reference in Part II.

Looking back to Fig. 5, the distribution of  $S_{\text{max}}$  is broad, meaning that  $S_{\text{max}}$  may vary greatly around the

median value  $S_{\text{max}}$  [Eq. (10)], and the median value  $S_{\text{max}}$  does not represent the large fluctuations of  $S_{\text{max}}$ . In other words, a trigger based on a threshold  $S_{\text{max}} > S_{\text{trig}}$  cannot represent properly the triggering process  $S_{\text{max}} > S_{\text{trig}}$ . Therefore, we have to consider the triggering process  $S_{\text{max}} > S_{\text{trig}}$  to be stochastic. In a time period  $\Delta t$ , the probability that  $S_{\text{max}} > S_{\text{trig}}$  is equal to the probability that a random sample  $0 < \mathcal{R} < 1$  exceeds the no-triggering probability per unit time  $\hat{P}_{\Delta t}$ . By the same token, in a time period  $\Delta t$ , the stochastic triggering happens if  $\mathcal{R} > \hat{P}_{\Delta t}$ .

*Recap: The three steps of the transition process*

### 1) PRELIMINARY CONDITION

Only moist thermals are expected to trigger deep convection; consequently, the reference altitude for computing the thermals' lifting energy ( $ALE_{\text{BL}}$ ) is taken at cloud base. Thus, the first necessary condition is that the boundary layer must be cloudy.

### 2) THE DYNAMICAL THRESHOLD

This threshold governs the transition from a regime in which cumulus clouds cannot reach their LFC (i.e., they stay under the CIN) to a transient regime where at least some cumulus overshoot the CIN but do not reach the high troposphere in significant numbers. It occurs when the statistical maximum kinetic energy provided by type-2 thermals,  $ALE_{\text{BL,stat}}$ , exceeds the CIN:

$$ALE_{\text{BL,stat}} > |\text{CIN}|. \quad (21)$$

### 3) THE GEOMETRIC THRESHOLD

Once the dynamical criterion is reached, the convective boundary layer enters a transient regime, in which some cumulus overshoot the inhibition but do not reach the high atmosphere. The geometric criterion governs the abrupt transition from the transient regime to the deep convection regime. It considers the type-2 thermal population spectrum and states that every independent cloud scene of duration  $\tau$  can potentially trigger on the condition that a random sample  $\mathcal{R}$  exceeds the no-triggering probability  $\hat{P}_\tau$ :

$$\mathcal{R} > \hat{P}_\tau. \quad (22)$$

In a modeling framework, since time is divided into time steps of duration  $\Delta t$ , this criterion becomes  $\mathcal{R} > \hat{P}_{\Delta t}$  for every time step.

Figure 11 illustrates the conceptual view of this formulation, from the first cloud to the triggering of deep

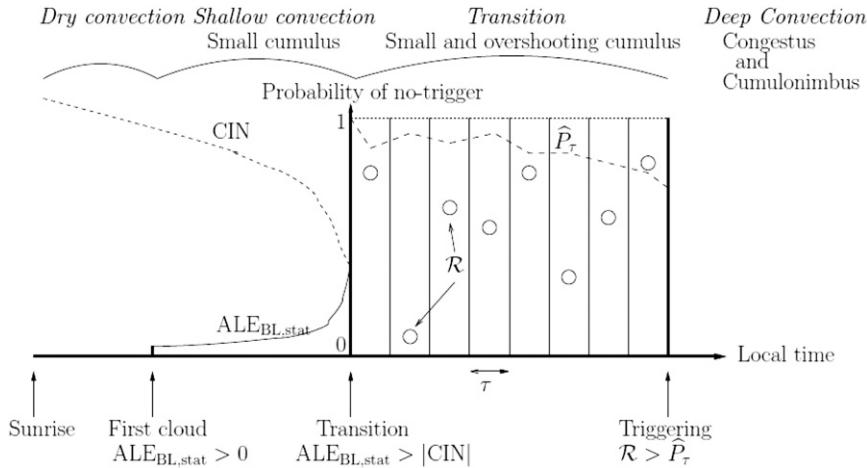


FIG. 11. Sketch of the transition from shallow to deep convection. Note that, in a numerical framework, a similar picture can be drawn by replacing  $\tau$  by  $\Delta t$ , where  $\Delta t$  is the model's time step.

convection. From this new formulation, a stochastic triggering parameterization is proposed in Part II.

**8. Discussion and conclusions**

Here the transition process is described considering a statistical ensemble of thermals with intrathermal velocity fluctuations and interthermal cloud-base cross-sectional fluctuations. Data from a LES case of transition from shallow to deep convection over land provide the geometric and dynamical properties of the cloudy thermals at the cloud-base level during the transition from shallow to deep convection.

The thermal population is fitted by a sum of two exponential distribution laws corresponding to two types. During the transition time, type-2 thermals are less numerous, become wider, and feed deeper clouds. A linear relationship between type-2 thermal width at cloud base, cloud depth, and cloud-base altitude has been proposed and verified on the LES case. Moreover, type-2 thermals can be described as a sum of independent drafts with length scales of several hundreds of meters, having a velocity distribution  $\mathcal{P}(w'_{p,i})$  that is nearly Gaussian and uniform over the thermal field. The Gaussian mean and standard deviation are similar, and a study of the distribution of maximum values  $\mathcal{P}_{\max}(w'_{\max,i})$  has shown that the Gaussian distribution is relevant for estimating the maximum velocities.

From the distribution of the maximum values of thermal sizes and draft velocities, a statistical maximum velocity  $\mathcal{W}'_{\max}$  for the whole thermal field has been computed to give a statistical estimate of  $ALE_{BL,stat}$ . In addition, the new triggering considers an arbitrary

threshold size  $S_{\text{trig}}$ , which has to be exceeded by the largest thermal  $S_{\max}$  to trigger deep convection. Knowing that the maximum size distribution  $\mathcal{P}_{\max}(S_{\max})$  is broad,  $S_{\max}$  fluctuations are very large, and it is pertinent to consider triggering (i.e.,  $S_{\max} > S_{\text{trig}}$ ) as a stochastic process, in which a random sample  $\mathcal{R}$  has to exceed a no-trigger probability  $\hat{P}_{\Delta t}$  to allow the onset of deep convection.

Thus, the present formulation proposes a three-step transition and consists of two consecutive thresholds, the first deterministic and the second stochastic. The first threshold is dynamic; it governs inhibition being exceeded by at least one thermal of the domain (i.e.,  $ALE_{BL,stat} > |CIN|$ ). It represents the moment when shallow clouds start to overshoot the inhibition layer and reach their LFC, that is, the transition phase. The second threshold is geometric and rules deep convection triggering. Since deep convection tends to trigger where the largest horizontal structures occur, there is a threshold cloud-base cross section which has a certain probability of being exceeded at every independent cloud scene.

To sum up, the new triggering formulation (i) suggests a thermal size distribution, in which only the largest elements control the triggering; (ii) proposes a new computation of the thermal available lifting energy at the cumulus cloud base; (iii) allows the existence of a transition stage between shallow and deep regimes through a multistep process; and (iv) includes a stochastic component, to better mimic the episodic aspect of the onset of deep convection.

However, integrating such a formulation in a parameterization of deep convection triggering by boundary layer thermals is still a difficult task. The main difficulty

is to retrieve a cross-sectional spectrum from the variables given by the boundary layer parameterization, which is single-thermal based in most cases. A triggering parameterization for the LMDZ model based on this formulation is proposed in Part II.

One may contest that this triggering formulation takes its inspiration from only one case study and so has little chance of being applicable in other situations. For this reason, the robustness of the formulation will be further investigated in Part II; the corresponding parameterization will be tested over various environmental conditions (continental and oceanic) and also in conditions favorable and unfavorable to triggering. It will be tested first in a single-column framework on different case studies and then in the global framework to estimate the added values in comparison to the standard approach in the full GCM.

*Acknowledgments.* The research leading to these results has been supported by both the French Department of Teaching and Research and the European Union, Seventh Framework Program (FP7/2011-2015) under Grant Agreement n282672. The authors thank Françoise Guichard and Jean-Philippe Lafore, from Météo-France (Toulouse), for their very helpful comments and for the numerous and enlightening discussions we had about deep convection issues. They also thank the three anonymous reviewers, whose meticulous analysis and enlightening suggestions made it possible to improve the quality of the manuscript.

## APPENDIX

### Maximum of a Large ( $\approx 100$ ) Number of Random Variables with Identical PDFs

We consider a set of  $N$  independent random variables  $(x_i)_{i=1,N}$  with identical PDF  $\mathcal{P}$ , CDF  $\hat{\mathcal{F}}$ , and CCDF  $\mathcal{F}$ . The CDF  $\hat{\mathcal{F}}$  (CCDF  $\mathcal{F}$ ) is defined by  $\hat{\mathcal{F}}(X) = \text{probability that } x_i < X (x_i > X)$ . The following relations hold

$$\mathcal{F}(X) = 1 - \hat{\mathcal{F}}(X); \quad \mathcal{P}(x) = \frac{d\hat{\mathcal{F}}}{dx} = -\frac{d\mathcal{F}}{dx}.$$

#### a. CCDF of the maximum

We seek the CCDF  $\mathcal{F}_{\max}$  of the maximum of the  $(x_i)_{i=1,N}$ . The probability that  $\max(x_i)$  exceeds a given value  $X$  is equal to the probability that at least one of the  $x_i$  exceeds  $X$ , which is equal to  $1 - (\text{probability that, for all } i, x_i < X)$ . Since the  $(x_i)_{i=1,N}$  are independent, the last

probability reads  $1 - [1 - \mathcal{F}(X)]^N$ . Thus, the CCDF  $\mathcal{F}_{\max}$  of the maximum of the  $(x_i)_{i=1,N}$  reads:

$$\mathcal{F}_{\max}(X) = 1 - [1 - \mathcal{F}(X)]^N, \quad (\text{A1})$$

which gives for the PDF of the maximum values:

$$\mathcal{P}_{\max}(X_{\max}) = \frac{-d\mathcal{F}_{\max}(X_{\max})}{dX_{\max}}. \quad (\text{A2})$$

#### b. Inverse formula

Given a probability  $P_t$ , we seek the corresponding threshold value  $X_t$  such that the probability that  $\max(x_i)_{i=1,N} > X_t$  is equal to  $P_t$ :

$$\mathcal{F}_{\max}(X_t) = P_t. \quad (\text{A3})$$

Note that we are interested in large values of the  $x_i$ , which implies that some upper bound be imposed upon  $P_t$ . As will appear later, an upper bound of  $P_t = 0.9$  is sufficient for the coming developments.

Substituting the expression of  $\mathcal{F}_{\max}$  [Eq. (A1)] in Eq. (A3) and solving for  $\mathcal{F}(X_t)$ , we obtain

$$\mathcal{F}(X_t) = 1 - (1 - P_t)^{1/N}. \quad (\text{A4})$$

This is an exact formula, but we shall use an approximate form taking account of the fact that  $N$  is large. To that end, we rewrite Eq. (A4) as

$$\mathcal{F}(X_t) = 1 - \exp\left[\frac{1}{N} \ln(1 - P_t)\right].$$

Introducing the new variable

$$\pi_t = -\ln(1 - P_t),$$

which verifies  $0 < \pi_t < 2.3$  when  $0 < P_t < 0.9$ , Eq. (A4) reads:

$$\mathcal{F}(X_t) = 1 - \exp\left(-\frac{\pi_t}{N}\right).$$

Since  $\pi_t/N$  is of the order of  $10^{-2}$ , the exponential may be replaced by a first-order expansion:

$$\mathcal{F}(X_t) = \frac{\pi_t}{N}. \quad (\text{A5})$$

Thanks to this equation, finding  $X_t$  amounts merely to inverting  $\mathcal{F}$ . In particular, the median  $X_{\text{med}}$  of the distribution of the maximum, which corresponds to  $P_t = 0.5$  and  $\pi_t \approx 0.7$ , is given by

$$\mathcal{F}(X_{\text{med}}) = \frac{\ln(2)}{N}.$$

c. Case of a thermal cross section  $s$

The CCDF is

$$\mathcal{F}(S) = \exp\left(-\frac{S}{\langle s \rangle}\right), \tag{A6}$$

and the number of random variables is the number  $N_{\text{tot}}$  of thermals in the grid cell. The threshold cross section  $S_t$  is given by Eq. (A5) where expression (A6) is substituted for  $\mathcal{F}(X_t)$ , that is,

$$\exp\left(-\frac{S_t}{\langle s \rangle}\right) = \frac{\pi_t}{N_{\text{tot}}}; \tag{A7}$$

thus,

$$S_t = \langle s \rangle \ln\left(\frac{N_{\text{tot}}}{\pi_t}\right). \tag{A8}$$

d. Case of vertical velocity  $w'_{p,i}$

The CCDF is

$$\mathcal{F}(W'_{p,i}) = \frac{1}{2} \text{Erfc}\left(\frac{W'_{p,i} - \overline{w'_{p,i}}}{\sqrt{2}\Gamma_{w'_{p,i}}}\right), \tag{A9}$$

and the number of random variables is the number  $n_i$  of elementary drafts in the thermal indexed  $i$ . The threshold vertical velocity  $W'_{t,i}$  verifies Eq. (A5) where expression (A9) is substituted for  $\mathcal{F}(X_t)$ , that is,

$$\frac{1}{2} \text{Erfc}\left(\frac{W'_{t,i} - \overline{w'_{p,i}}}{\sqrt{2}\Gamma_{w'_{p,i}}}\right) = \frac{\pi_t}{n_i}. \tag{A10}$$

Since  $\pi_t/n_i \ll 1$ , we use the asymptotic form of  $\text{Erfc}^{-1}(x)$  in the limit  $x \rightarrow 0$ ,

$$\text{Erfc}^{-1}(x) \approx \frac{1}{\sqrt{2}} \sqrt{\ln\left(\frac{2}{\pi x^2}\right) - \ln\left[\ln\left(\frac{2}{\pi x^2}\right)\right]}$$

(see the Wolfram Mathematica website at <http://www.wolframalpha.com> for reference), which yields

$$W'_{t,i} = \overline{w'_{p,i}} + \Gamma_{w'_{p,i}} \sqrt{\ln\left(\frac{n_i^2}{2\pi\pi_t^2}\right) - \ln\left[\ln\left(\frac{n_i^2}{2\pi\pi_t^2}\right)\right]}. \tag{A11}$$

REFERENCES

Bechtold, P., J. Chaboureau, A. Beljaars, A. Betts, M. Köhler, M. Miller, and J. Redelsperger, 2004: The simulation of the diurnal cycle of convective precipitation over land in a global model. *Quart. J. Roy. Meteor. Soc.*, **130**, 3119–3137.

Boing, S., H. Jonker, A. Siebesma, and W. Grabowski, 2010: Influence of subcloud-layer structures on the transition to deep convection. Preprints, *19th Symp. on Boundary Layers and Turbulence*, Keystone, CO, Amer. Meteor. Soc., P4.4. [Available online at <https://ams.confex.com/ams/pdfpapers/172640.pdf>.]

Chaboureau, J., F. Guichard, J. Redelsperger, and J. Lafore, 2004: The role of stability and moisture in the diurnal cycle of convection over land. *Quart. J. Roy. Meteor. Soc.*, **130**, 3105–3117.

Clark, T., T. Hauf, and J. Kuettner, 1986: Convectively forced internal gravity waves: Results from two-dimensional numerical experiments. *Quart. J. Roy. Meteor. Soc.*, **112**, 899–925.

Cohen, B. G., and G. C. Craig, 2006: Fluctuations in an equilibrium convective ensemble. Part II: Numerical experiments. *J. Atmos. Sci.*, **63**, 2005–2015.

Couvreux, F., C. Rio, F. Guichard, M. Lethien, G. Canut, D. Bouniol, and A. Gounou, 2012: Initiation of daytime local convection in a semi-arid region analysed with high-resolution simulations and AMMA observations. *Quart. J. Roy. Meteor. Soc.*, **138**, 56–71.

Craig, G. C., and B. Cohen, 2006: Fluctuations in an equilibrium convective ensemble. Part I: Theoretical formulation. *J. Atmos. Sci.*, **63**, 1996–2004.

—, and A. Dörnbrack, 2008: Entrainment in cumulus clouds: What resolution is cloud-resolving? *J. Atmos. Sci.*, **65**, 3978–3988.

Damiani, R., G. Vali, and S. Haimov, 2006: The structure of thermals in cumulus from airborne dual-Doppler radar observations. *J. Atmos. Sci.*, **63**, 1432–1450.

Del Genio, A., and J. Wu, 2010: The role of entrainment in the diurnal cycle of continental convection. *J. Climate*, **23**, 2722–2738.

Donner, L. J., 1993: A cumulus parameterization including mass fluxes, vertical momentum dynamics, and mesoscale effects. *J. Atmos. Sci.*, **50**, 889–906.

Emanuel, K. A., 1991: A scheme for representing cumulus convection in large-scale models. *J. Atmos. Sci.*, **48**, 2313–2335.

—, J. D. Neelin, and C. S. Bretherton, 1994: On large-scale circulations in convecting atmospheres. *Quart. J. Roy. Meteor. Soc.*, **120**, 1111–1143.

Grabowski, W., and Coauthors, 2006: Daytime convective development over land: A model intercomparison based on LBA observations. *Quart. J. Roy. Meteor. Soc.*, **132**, 317–344.

Grandpeix, J., and J. Lafore, 2010: A density current parameterization coupled with Emanuel’s convection scheme. Part I: The models. *J. Atmos. Sci.*, **67**, 881–897.

—, —, and F. Cheruy, 2010: A density current parameterization coupled with Emanuel’s convection scheme. Part II: 1d simulations. *J. Atmos. Sci.*, **67**, 898–922.

Guichard, F., and Coauthors, 2004: Modelling the diurnal cycle of deep precipitating convection over land with cloud-resolving models and single-column models. *Quart. J. Roy. Meteor. Soc.*, **130**, 3139–3172.

Heus, T., H. J. Jonker, H. E. Van den Akker, E. J. Griffith, M. Koutek, and F. H. Post, 2009: A statistical approach to the life cycle analysis of cumulus clouds selected in a virtual reality environment. *J. Geophys. Res.*, **114**, D06208, doi:10.1029/2008JD010917.

- Jones, T., and D. Randall, 2011: Quantifying the limits of convective parameterizations. *J. Geophys. Res.*, **116**, D08210, doi:10.1029/2010JD014913.
- Khairoutdinov, M., and D. Randall, 2006: High-resolution simulation of shallow-to-deep convection transition over land. *J. Atmos. Sci.*, **63**, 3421–3436.
- Lafore, J., and Coauthors, 1998: The Meso-NH atmospheric simulation system. Part I: Adiabatic formulation and control simulations. *Ann. Geophys.*, **16**, 90–109.
- LeMone, M. A., and W. T. Pennell, 1976: The relationship of trade wind cumulus distribution to subcloud layer fluxes and structure. *Mon. Wea. Rev.*, **104**, 524–539.
- Lothon, M., B. Campistron, M. Chong, F. Couvreux, F. Guichard, C. Rio, and E. Williams, 2011: Life cycle of a mesoscale circular gust front observed by a c-band doppler radar in west africa. *Mon. Wea. Rev.*, **139**, 1370–1388.
- Malkus, J., 1954: Some results of a trade-cumulus cloud investigation. *J. Meteor.*, **11**, 220–237.
- Mapes, B., 2000: Convective inhibition, subgrid-scale triggering energy, and stratiform instability in a toy tropical wave model. *J. Atmos. Sci.*, **57**, 1515–1535.
- Matheou, G., D. Chung, L. Nuijens, B. Stevens, and J. Teixeira, 2011: On the fidelity of large-eddy simulation of shallow precipitating cumulus convection. *Mon. Wea. Rev.*, **139**, 2918–2939.
- Medeiros, B., A. Hall, and B. Stevens, 2005: What controls the mean depth of the PBL? *J. Climate*, **18**, 3157–3172.
- Neelin, J., O. Peters, J. Lin, K. Hales, and C. Holloway, 2008: Rethinking convective quasi-equilibrium: Observational constraints for stochastic convective schemes in climate models. *Philos. Trans. Roy. Soc. London*, **A366**, 2579–2602.
- Neggers, R., H. Jonker, and A. Siebesma, 2003: Size statistics of cumulus cloud populations in large-eddy simulations. *J. Atmos. Sci.*, **60**, 1060–1074.
- Palmer, T. N., 2012: Towards the probabilistic earth-system simulator: a vision for the future of climate and weather prediction. *Quart. J. Roy. Meteor. Soc.*, **138**, 841–861.
- Plant, R., and G. Craig, 2008: A stochastic parameterization for deep convection based on equilibrium statistics. *J. Atmos. Sci.*, **65**, 87–105.
- Raymond, D., and M. Herman, 2011: Convective quasi-equilibrium reconsidered. *J. Adv. Model. Earth Syst.*, **3**, M08003, doi:10.1029/2011MS000079.
- Rio, C., and F. Hourdin, 2008: A thermal plume model for the convective boundary layer: Representation of cumulus clouds. *J. Atmos. Sci.*, **65**, 407–425.
- , —, J. Grandpeix, and J. Lafore, 2009: Shifting the diurnal cycle of parameterized deep convection over land. *Geophys. Res. Lett.*, **36**, L07809, doi:10.1029/2008GL036779.
- Rochetin, N., J.-Y. Grandpeix, C. Rio, and F. Couvreux, 2014: Deep convection triggering by boundary layer thermals. Part II: Stochastic triggering parameterization for the LMDZ GCM. *J. Atmos. Sci.*, **71**, 515–538.
- Rodts, S., P. Duynkerke, and H. Jonker, 2003: Size distributions and dynamical properties of shallow cumulus clouds from aircraft observations and satellite data. *J. Atmos. Sci.*, **60**, 1895–1912.
- Seifert, A., and T. Heus, 2013: Large-eddy simulation of organized precipitating trade wind cumulus clouds. *Atmos. Chem. Phys. Discuss.*, **13**, 1855–1889, doi:10.5194/acpd-13-1855-2013.
- Stechmann, S. N., and J. D. Neelin, 2011: A stochastic model for the transition to strong convection. *J. Atmos. Sci.*, **68**, 2955–2970.
- Taylor, C., A. Gounou, F. Guichard, P. Harris, R. Ellis, F. Couvreux, and M. De Kauwe, 2011: Frequency of sahelian storm initiation enhanced over mesoscale soil-moisture patterns. *Nat. Geosci.*, **4**, 430–433.
- Warner, J., 1970: The microstructure of cumulus cloud. Part III: The nature of the updraft. *J. Atmos. Sci.*, **27**, 682–688.
- Wu, C., B. Stevens, and A. Arakawa, 2009: What controls the transition from shallow to deep convection? *J. Atmos. Sci.*, **66**, 1793–1806.
- Xu, K., A. Arakawa, and S. Krueger, 1992: The macroscopic behavior of cumulus ensembles simulated by a cumulus ensemble model. *J. Atmos. Sci.*, **49**, 2402–2420.
- Yang, G., and J. Slingo, 2001: The diurnal cycle in the tropics. *Mon. Wea. Rev.*, **129**, 784–801.
- Zhao, M., and P. H. Austin, 2005: Life cycle of numerically simulated shallow cumulus clouds. Part I: Transport. *J. Atmos. Sci.*, **62**, 1269–1290.