

# Statistical properties of cloud lifecycles in cloud-resolving models

R. S. Plant

Department of Meteorology, University of Reading, Earley Gate, Reading, P.O. Box 243, RG6 6BB, UK

Received: 15 August 2008 – Published in Atmos. Chem. Phys. Discuss.: 9 December 2008

Revised: 24 February 2009 – Accepted: 17 March 2009 – Published: 24 March 2009

**Abstract.** A new technique is described for the analysis of cloud-resolving model simulations, which allows one to investigate the statistics of the lifecycles of cumulus clouds. Clouds are tracked from timestep to timestep within the model run. This allows for a very simple method of tracking, but one which is both comprehensive and robust. An approach for handling cloud splits and mergers is described which allows clouds with simple and complicated time histories to be compared within a single framework. This is found to be important for the analysis of an idealized simulation of radiative-convective equilibrium, in which the moist, buoyant updrafts (i.e., the convective cores) were tracked. Around half of all such cores were subject to splits and mergers during their lifecycles. For cores without any such events, the average lifetime is 30 min, but events can lengthen the typical lifetime considerably.

## 1 Introduction

In recent years Cloud Resolving Models (CRMs) have become an increasingly important tool for the study of convective phenomena. CRMs should not be regarded as providing surrogates for observations; rather, they allow idealized but realistic simulations to be produced which provide a laboratory for the careful diagnostic analysis of generic convective systems. Such analysis is a distinctive methodology that is necessary to improve our understanding of the basic phenomena and to develop improved parameterization methods for larger-scale models.

This paper describes and illustrates the use of a novel analysis technique for CRM data, which allows one to investigate statistical properties of the lifecycles of clouds produced during CRM simulations.

Current analyses of CRM data often focus on determining and understanding the spatial and temporal average properties of the full ensemble of convective clouds that are produced in the model in response to some specified external forcing (e.g. Petch et al., 2007). Rather less attention has been devoted to the lifecycle behaviour of individual clouds. There are currently many simulations (whether labelled as CRM or otherwise) which are being performed with convection represented explicitly but at rather coarse resolution ( $\sim 1$  to 5 km) (e.g. Petch et al., 2002; Done et al., 2004; Khairoutdinov et al., 2005). In such simulations a deep convective cloud may occupy only a small number of model gridpoints. Thus, although the results may provide genuine value relative to their lower-resolution counterparts with parameterized convection (e.g. Roberts and Lean, 2008), it is far from clear that the simulations will provide a good representation of individual clouds. A statistical investigation into cloud lifecycles could therefore be valuable in order to reveal which aspects of the lifecycles are well or poorly captured at these model resolutions. Even assuming a high-resolution simulation, however, statistical information on the cloud lifecycle would be useful to test the realism of the model clouds, and to allow one to examine the detailed effects of model parameterizations, such as the microphysics. A good recent study of cloud lifecycles in a CRM simulation is that of Zhao and Austin (2005a,b). However, practical constraints limited that study to an investigation of six clouds, making it difficult to assess whether the results are generic.

Statistical investigations into the lifecycles of cumulus clouds could also allow improvements to be made to existing



Correspondence to: R. S. Plant  
(r.s.plant@reading.ac.uk)

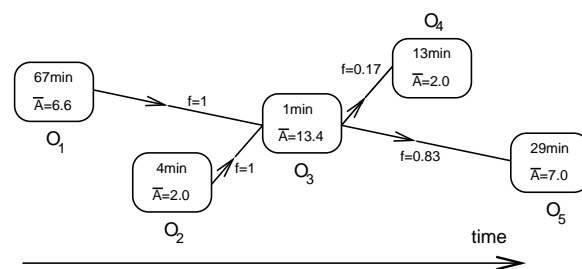
convective parameterizations. Cho (1977) considered the effects of incorporating a cloud lifecycle into a cumulus-ensemble mass-flux framework, and showed that the effects on the apparent heating were negligible. However, an additional contribution arises in the apparent moisture sink compared to a steady-state cloud model, due to mixing of air from the decayed cloud with its environment. Another example comes from the popular Kain and Fritsch parameterization (Kain, 2004) for mesoscale models. A rudimentary lifecycle is included by assigning to the convective plumes a (somewhat arbitrary) lifetime which extends over multiple timesteps. The parameterization is a mass-flux scheme which considers a single plume to be representative of all convection occurring within a model grid box. Based on the pioneering study of Arakawa and Schubert (1974), some other parameterizations consider a spectrum of convective plumes (Plant and Craig, 2008, is a recent example). Future parameterizations might seek to combine these two features: multiple cloud types and a simple cloud lifecycle. However, this is not possible at present, essentially because there is a lack of available information about how the cloud lifecycle varies with cloud type (and forcing regime).

An important feature of many observed cumulus clouds is that they may evolve through a sequence of pulse-like events (e.g. Scorer and Ludlam, 1953; Blyth et al., 2005). The existence of such pulses may complicate the careful tracking of cumulus clouds because identification criteria that pick-out individual thermals are liable to pick-out objects that are subject to various interactions. Those interactions may be difficult to describe even qualitatively (Westcott, 1984) but both cell-merging (Wiggert et al., 1981; Weusthoff and Hauf, 2008a) and cell-splitting (Fujita et al., 1975) have been observed to be common phenomena. One of the goals here then is to develop a tracking system that is robust but detailed enough to deal with situations in which interactions between the tracked objects are commonplace.

An automated method is presented which first identifies and then tracks the development of individual clouds in a CRM simulation. Its most important characteristic is that it is run online, at every timestep, alongside the model simulation. By exploiting the high temporal resolution available in a model, it is possible to devise a tracking method that is at once simple, comprehensive and robust. The method is fully described in Sect. 2 and results obtained from tracking moist, buoyant updrafts in a CRM simulation are discussed in Sect. 3. Conclusions are drawn in Sect. 4.

## 2 Methodology

The purpose of the tracking algorithm is to capture the complete time evolution of each cloud produced in a numerical model simulation. The algorithm can be divided into three main parts, which will be described in turn below. Before proceeding to the details of the algorithm, we present Fig. 1,



**Fig. 1.** Example of the evolution of and inter-relationships between cloud objects. Each object is labelled  $O_1$ ,  $O_2$  etc. It persists for the time indicated in min and with a time-mean area  $\bar{A}$  expressed in units of the grid box area. The values of  $f$  denote “fractions” that are defined in Sect. 2.3. They are used to characterise the combination or break-up of objects.

which provides an example of an evolution that one would wish to describe in the tracking.

Figure 1 shows the development of and inter-relationships between five “cloud objects”, each object being a connected group of “cloudy” grid boxes (as defined in Sect. 2.1). The object  $O_1$  can be recognized as a coherent and persistent structure (Sect. 2.2) for 67 min. It occupies an average of 6.6 grid boxes, growing from 2 connected grid boxes when first identified into an object of area 10 grid boxes by the time that it combines with another object,  $O_2$ . This second object has occupied 2 grid boxes since it was first identified, 4 min before the combination. The combined object is denoted  $O_3$  and retains coherence for 1 min before breaking-up into two distinct groupings: the small object  $O_4$  and the larger and longer-lived object  $O_5$ . We shall refer back to Fig. 1 on several occasions below in order to illustrate how the general tracking algorithm operates for this particular case.

### 2.1 Identify cloud objects at a given timestep

Cloud identification requires, first, a determination of the grid boxes that are considered cloudy, and second, connecting such boxes together into distinct structures that we will refer to as cloud objects. A wide variety of criteria have been used in the literature for the identification of cloudy boxes. Analyses of satellite observations often employ a brightness-temperature threshold (e.g. Kuo et al., 1993; Carvalho and Jones, 2001; Machado and Laurent, 2004). Other methods are based on radar echoes (e.g. Foote and Mohr, 1979; Dixon and Wiener, 1993; Theusner and Hauf, 2004) and even the visual inspection of photographs (e.g. Plank, 1969; Hozumi et al., 1982).

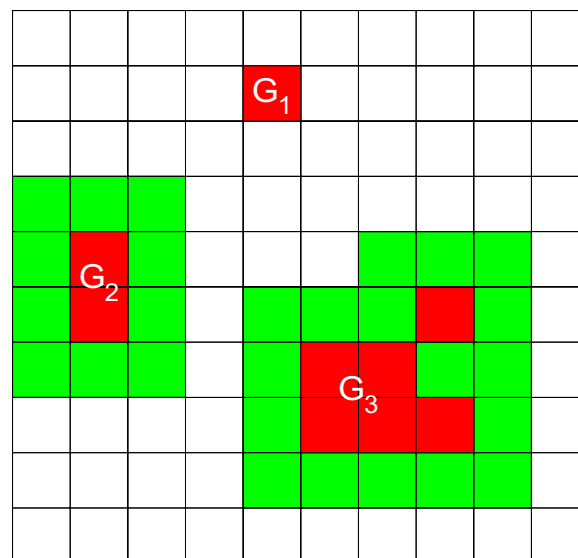
In model simulations, the identification of cloudy grid boxes is less constrained by the character of the data and it is possible to define thresholds for model variables that are arguably more directly related to the presence of cloud. One popular choice (e.g. Xu and Randall, 2001; Cohen and Craig,

2006) is to use a vertical velocity criterion ( $w > 1 \text{ ms}^{-1}$  anywhere in the column) in order to pick out strong updrafts. This approach has its origin in analyses of aircraft observations (LeMone and Zipser, 1980; Zipser and LeMone, 1980). Other methods are simply to use model variables for cloud water and/or ice content (e.g. Cohen and Craig, 2006), to consider the convective transport of boundary layer air by means of a passive tracer (Zhao and Austin, 2005a), or even visual inspection of data in a virtual reality environment (Heus, 2008). Siebesma and Cuijpers (1995) compared three identification methods for simulated shallow cumulus, which they referred to as the cloud decomposition (positive cloud water), the updraft decomposition (positive cloud water and vertical velocity) and the cloud-core decomposition (positive cloud water, vertical velocity and buoyancy). The cloud-core decomposition produced the best agreement between the mass flux representation of turbulent fluxes (assumed by many parameterizations) and the actual model fluxes. For related discussions, see also Swann (2001); Siebesma et al. (2003); Yano et al. (2004).

It would be wrong to view any particular cloud definition as intrinsically correct. Rather the different definitions allow one to focus attention on different aspects of the cloud field. In Sect. 3 we will use a “cloud-core” definition, but it would be straightforward to implement other choices. Specifically, a grid column is taken to be cloudy in this study if a small positive threshold ( $10^{-5}$ ) is exceeded for all three of the following variables on the same model level: the cloud water (in  $\text{kg kg}^{-1}$ ), the vertical velocity (in  $\text{ms}^{-1}$ ) and the buoyancy (actually  $\theta'_v$  in K).

Once the “cloudy” grid boxes have been determined, it remains to connect adjacent boxes together into cloud objects. As discussed by Kuo et al. (1993) for example, either a four-segmented or eight-segmented method can be used, the former considering only those adjacent grid boxes which share a gridbox edge, whereas the latter also allows connections to neighbouring grid boxes along a diagonal. In Sect. 3 an eight-connected method will be used. For example then, the group of cloudy grid boxes labelled  $G_3$  in Fig. 2 would be considered as an object containing six model grid boxes. Although Kuo et al. (1993) obtained similar results from the two methods, some differences occur in the numbers of one-point and two-point cloud objects in coarse-resolution simulations (Lennard, 2004).

For a cloud object to be included in the tracking process presented below, it is required to contain at least two cloudy grid boxes. Thus the very smallest clouds, such as the group  $G_1$  in Fig. 2, are ignored. One would not expect these to be well represented by the model. A further requirement for a cloud lifecycle to be included in the statistics is that tracking should be possible for at least 5 min. The combination of these two conditions helps to ensure that the final statistics should not be overly sensitive to the precise definition of cloudy grid boxes, since any isolated, short-lived fluctuations above a threshold are excluded.



**Fig. 2.** Schematic diagram showing a portion of the horizontal domain used by a numerical model. Grid boxes identified as cloudy are shown in red, whilst green indicates the halo of grid boxes to be considered when determining relationships with the cloud objects present at the previous timestep.  $G_1$ ,  $G_2$  and  $G_3$  label groupings of cloudy grid boxes.

## 2.2 Relationships with cloud objects at the previous timestep

The purpose of the second part of the algorithm is to establish the relationships between cloud objects present at the current timestep and those present at the previous one. Many tracking methods have been developed for determining the evolution of features in data of relatively low temporal resolution (e.g. Dixon and Wiener, 1993; Carvalho and Jones, 2001; Machado and Laurent, 2004, and references therein). Given two time slices, each of which contains one or more features of interest, the aim is to establish the features that are in common between the two slices, essentially satisfying oneself that a feature in the first slice is highly likely to have evolved into some feature(s) in the later slice. Often the method will involve forming some estimate of the propagation speed of the feature. On occasion, the relationships between features in the two time slices may not be entirely clear. (Data errors, such as radar clutter, can also produce some tracking errors, as noted by Weusthoff and Hauf (2008b) for instance.)

Here the tracking algorithm is applied online, as a diagnostic component of the numerical model simulation. Because data is available to the tracking algorithm with very high temporal resolution, it is possible to establish the relationships between cloud objects at adjacent timesteps using a method that is both simple and comprehensive. It is assumed that in a single timestep all motion is less than a single horizontal gridlength. This is a numerical stability requirement

for many of the advection schemes used by CRMs, including the LEM used in Sect. 3. For there to be a relationship between two cloud objects at adjacent timesteps, it follows that either the areas of the two objects must overlap, or else that the object at the current timestep must be no further than one grid box from that at the previous timestep. Thus, all of the relationships required can be found by looking for cloud objects present (at least in some part) at the previous timestep within a halo region for each current cloud object. Halos are illustrated in Fig. 2, and comprise grid boxes that either overlap or are adjacent to the cloud object of interest.

From the set of all relationships, the character of the relationships between previous and current cloud objects can be determined. This proceeds from the construction of the maximum possible number of subsets of relationships, subject to the constraint that each cloud object at the current and previous timestep appears in one and only one subset. Figure 1 can be used to provide various examples. There is initially a single subset, containing nothing from the previous timestep and  $O_1$  from the current timestep. Let us denote this as  $(0, O_1)$  with the comma serving to separate the previous from the current timestep. Thereafter there is again a single subset,  $(O_1, O_1)$ . On identification of  $O_2$ , there will then be two subsets,  $(O_1, O_1)$  and  $(0, O_2)$ . Thereafter the two subsets are  $(O_1, O_1)$  and  $(O_2, O_2)$  until the timestep at which the two objects combine. At that time, a single subset will be identified, specifically  $(O_1 O_2, O_3)$ . After the combination, we will be dealing with a single, simple subset again,  $(O_3, O_3)$ .

A useful property of each such subset can be denoted by  $p \rightarrow c$ , where  $p$  and  $c$  are the total number of cloud objects in the subset from the previous and current timesteps respectively. This property allows the following characteristic relationships to be distinguished.

- $0 \rightarrow 1$  (i.e., there is no relationship to a cloud object at the current timestep from any of the objects present at the previous timestep) signifies the birth of a new cloud object.
- $1 \rightarrow 0$  signifies the death of a cloud object.
- $1 \rightarrow 1$  (by far the most common occurrence in practice) signifies a straightforward continuation of a pre-existing cloud object.
- $1 \rightarrow 2+$  signifies the splitting up of a pre-existing cloud object.
- $2+ \rightarrow 1$  signifies a merger of pre-existing cloud objects to form a single object.
- $2+ \rightarrow 2+$  signifies more complicated relationships, which might occur, for example, if a pre-existing cloud object simultaneously both breaks-up and absorbs another pre-existing cloud object. Such happenings are extremely rare, but nonetheless must be accounted for.

Note that  $2+$  has been used to denote two or more cloud objects. It is convenient to be able to distinguish between the births, deaths and straightforward continuations on the one hand, and the splits, mergers and complicated relationships on the other. In order to do so, we will henceforth refer to the latter types of relationship as “events”.

### 2.3 Compile timeseries data for each cloud

We consider a cloud lifecycle to be terminated by the death of a cloud object, and to have begun at the birth of the first cloud object that can be linked to the dead object through the tracking process. For each cloud object, a timeseries is stored of relevant data, including such properties as the object size (number of grid boxes), the precipitation rate and mass fluxes. The procedure for updating and organizing the timeseries depends upon the character of relationships to the cloud objects present at the previous timestep, as we now explain. Births, deaths and straightforward continuations are easy to deal with, signalling respectively the start of a new timeseries, the output of a completed cloud lifecycle, and the addition of a new entry to a pre-existing timeseries.

For any event, the timeseries of all pre-existing cloud objects contributing to the event are closed and archived into a library. New timeseries are begun for all of the cloud objects from the current timestep that are involved in the event. The full time history for current cloud objects can thus be reconstructed by means of references to the library. If a current cloud object has been subject to a single event in reaching its current state then we describe it as being a second-generation cloud object. Higher orders of generation are also possible: for example, in Fig. 1,  $O_4$  results from a merger of two cloud objects, with the combined object then splitting up. Higher orders can be incorporated by extending the above procedure to allow inter-library references. In the example of  $O_4$  then, this object is linked to the data for  $O_3$  that is held in the library. But the library also contains the data for  $O_1$  and  $O_2$ , and information is retained to link  $O_3$  back to this  $O_1$  and  $O_2$  data. The route from  $O_4$  back to the birth of  $O_1$  or  $O_2$  requires a sequence of three cloud objects, so that  $O_4$  may be considered a third-generation object. In this way, each current cloud can be followed back through all of its contributing elements.

As well as the character of each event, it is useful also to save parameters which estimate the relative contributions of the various cloud objects involved. Specifically, we calculate the quantities  $f_i^c$  which represent the fraction of a cloud object  $i$  from the previous timestep that can be linked to the current cloud object  $c$ . (The fraction should be interpreted as zero if there is no relationship between  $i$  and  $c$ .) For multi-generational clouds, this is generalized to a fractional association  $a_n^c$  with some cloud object  $n$  held in the library. The association is given by

$$a_n^c = \sum_{ml \dots ji} f_n^m f_m^l \dots f_j^i f_i^c \quad (1)$$

where the summation extends over all possible combinations of library objects that lead from  $n$  to  $c$ . Applying these concepts to Fig. 1, the merger of  $O_1$  and  $O_2$  into  $O_3$  is described by two fractions, one for the link between  $O_1$  and  $O_3$  and the other for that between  $O_2$  and  $O_3$ . Similarly there are two non-trivial fractions associated with the splitting-up of  $O_3$  into  $O_4$  and  $O_5$ . If we now consider  $O_4$ , we see that its complete description requires a fraction linking it to  $O_3$  and associations linking it with  $O_1$  and  $O_2$ . The association to  $O_1$  for example, is given by the product of the fraction linking  $O_4$  and  $O_3$  with the fraction linking  $O_3$  and  $O_1$ .

The determination of the fraction  $f_i^c$  makes use of the areas occupied by the cloud objects concerned. In a  $2 \rightarrow 1$  merger of objects  $i$  and  $j$  to produce object  $c$ , the fractions are trivially

$$f_i^c = f_j^c = 1 \quad (2)$$

while for a  $1 \rightarrow 2$  split of object  $i$  into objects  $c$  and  $d$  we have

$$f_i^c = \frac{A^c}{A^c + A^d}; \quad f_i^d = \frac{A^d}{A^c + A^d} \quad (3)$$

where  $A$  denotes the cloud object area. For example, just after the split in Fig. 1,  $O_4$  and  $O_5$  occupied areas of 2 and 10 grid boxes respectively, resulting in the fractions marked on that figure.

A generalization of the approach to encompass other events is given by the equation below, allowing fractions to be determined for potentially complicated events involving multiple cloud objects from both the previous and current timesteps. Specifically:

$$f_i^c = N_i \left( r^c + \frac{A_i}{l_i} \right) \quad (4)$$

where  $l$  is the total number of relationships from a particular (subscripted) cloud object at the previous timestep to all objects at the current timestep. The reduced area  $r^c$  may be positive or negative and is intended to provide an indication of any portion of the current cloud object  $c$  that is not linked to objects from the previous timestep. (In a merger, for example, a positive value would indicate an object at the current timestep that is larger than the sum of areas of its constituent objects from the previous timestep.) It is defined by

$$r^c = A^c - \sum_i^l \frac{A_i}{l_i}, \quad (5)$$

where the primed summation extends over those objects  $i$  from the previous timestep that have a relationship to the current cloud object  $c$ . To complete the specification of the fraction, it remains to define the quantity  $N_i$ . This is a normalization factor, chosen to ensure that all objects from the previous timestep that are involved in events are fully linked to current objects. Hence,  $N_i$  is such that

$$\sum_c f_i^c = 1. \quad (6)$$

As a check on the formula in Eq. (4), it is easy to confirm that for a  $2 \rightarrow 1$  merger and a  $1 \rightarrow 2$  split, the fractions reduce to those given in Eqs. (2) and (3) respectively. Taking the split in Fig. 1 as an example, the reduced area for  $O_4$  is the difference between its own area just after the split and the area of  $O_3$  just before the split. Eq. (4) then indicates that the fraction linking  $O_3$  and  $O_4$  is proportional to the area of  $O_4$ , and finally Eq. (6) provides the constant of proportionality as the reciprocal of the sum of the areas of  $O_4$  and  $O_5$ .

We wish to be able to compare cloud lifecycles with events during their time history alongside simple lifecycles without any events. In order to do so, it is necessary to construct a single timeseries for each cloud lifecycle, even for lifecycles that encompass multiple events. We define the lifetime of a cloud lifecycle as its complete duration, extending backwards from the death of a cloud object to the birth of its first contributing object. For an extensive cloud property  $E$ , the lifecycle timeseries is obtained from

$$E^c = \sum_n a_n^c E_n \quad (7)$$

the sum extending over all contributing cloud objects  $n$ , with the understanding that this includes the terminating object  $c$  itself and that  $a_c^c = 1$ . For an intensive cloud property  $I$ , the product of the association with the area of a cloud object is considered to provide a weighting factor, so that

$$I^c = \frac{\sum_n a_n^c A_n I_n}{\sum_n a_n^c A_n}. \quad (8)$$

As an example, consider the timeseries of total precipitation (an extensive variable) for the cloud lifecycle that concludes with the cloud object  $O_4$ . The full timeseries covers 85 min, the first 63 min capturing 17% of the precipitation from  $O_1$ , the next 4 min capturing 17% of the precipitation from  $O_1$  and 17% of that from  $O_2$ , the next 1 min capturing 17% of the precipitation from  $O_3$  and the final 14 min capturing the entirety of  $O_4$ .

Finally, we note two restrictions on the multi-generational cloud-object library that are imposed for purely practical reasons. If the library becomes very large, or if lifecycles extend through many generations, then searching through the library can become a time-consuming operation, considerably slowing the model simulation. Therefore, we remove from the library any archived cloud objects  $n$  with associations  $a_n^c$  that are less than 0.05 for all of the current cloud objects  $c$ . Moreover, we do not allow a lifecycle to extend backwards for more than 10 generations. Some diagnostics characterizing the removed cloud objects are output in order to allow various checks that the removals do not have significant adverse effects on the final lifecycle statistics. For example, in the simulation results to be presented in Sect. 3 the removed objects did not persist for long: under 2 min on average, which compares with an average of 14 min for the cloud objects retained in the library. Various test runs with different values for the removal criteria have also been performed to check explicitly for any effects of the removals.

**Table 1.** Statistics of the convective-core lifecycles tracked during a CRM simulation of radiative-convective equilibrium. Grid boxes containing moist, buoyant updrafts are referred to as cloudy, while two-or-more connected cloudy grid boxes constitute a cloud object (Sect. 2.1). The proportions of births, deaths, splits, mergers and complicated events are expressed as fractions relative to the number of straightforward continuations (of which there were  $3.1 \times 10^7$ ). All quantities are computed from data at each model timestep between 19.5 and 36 days.

Quantity	Mean	Standard deviation
Cloudy grid boxes	52.4	6.9
Cloudy boxes not part of cloud objects	7.0	2.8
Number of cloud objects	10.0	2.0
Proportion of births and deaths	$3.0 \times 10^{-4}$	–
Proportion of splits	$2.4 \times 10^{-4}$	–
Proportion of mergers	$1.7 \times 10^{-4}$	–
Proportion of complicated events	$4.2 \times 10^{-6}$	–

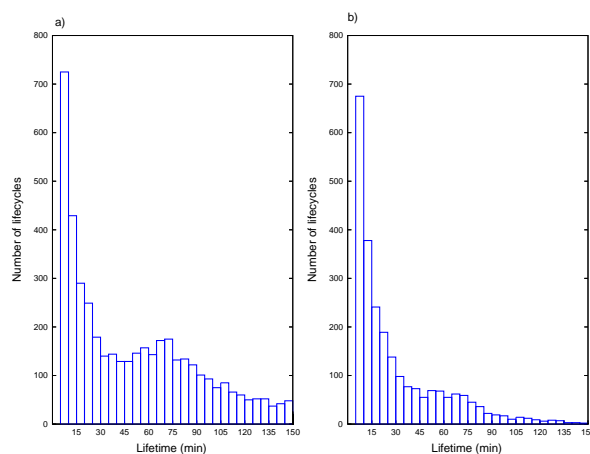
### 3 Results from a CRM simulation

The tracking algorithm described in Sect. 2 has been tested in both a cloud-resolving model and in artificial dynamical systems of cellular automata (based on variations of the game-of-life rules). The advantage of the artificial system is that its rules can be altered to test various aspects of the algorithm: for example, allowing events to be extremely rare or else frequent and complex. Explicit timestep-to-timestep validations have been performed to check that the algorithm is robust and functions as designed.

We present results for a simulation of radiative-convective equilibrium performed with the Met. Office Large Eddy Model (LEM) (Petch and Gray, 2001). The setup is not dissimilar to simulations that have previously been studied by Cohen and Craig (2004, 2006). The convection is forced by cooling the troposphere at  $4 \text{ K day}^{-1}$ , over a sea-surface which has its temperature held fixed at 300 K. The simulation domain is a doubly-periodic grid of size  $64 \times 64 \times 20 \text{ km}^3$  with a horizontal gridlength of 2 km and 76 staggered vertical levels. The Coriolis parameter is set to zero and no mean wind is imposed, so that the convection is not expected to be organized by the large-scale state. In fact, limited self-organization does occur in such conditions, as discussed by Cohen and Craig (2006); Davies (2008).

The LEM has a variable timestep, which can change during the simulation. This is to ensure good behaviour of the subgrid model (Mason, 1989; Brown et al., 1994), and also that the CFL stability condition for advection remains satisfied throughout. The timestep ranges from 0.30 to 0.65 s, with a mean value of 0.51 s.

The simulation is run for 36 model days, of which the first 19.5 days are used to spin-up from a rather arbitrary initial condition to the equilibrium state. The domain-mean model



**Fig. 3.** Convective-core lifetime distribution for (a) all lifecycles, and (b) lifecycles which do not contain events in their time histories. The lifetimes are binned into intervals of 5 min.

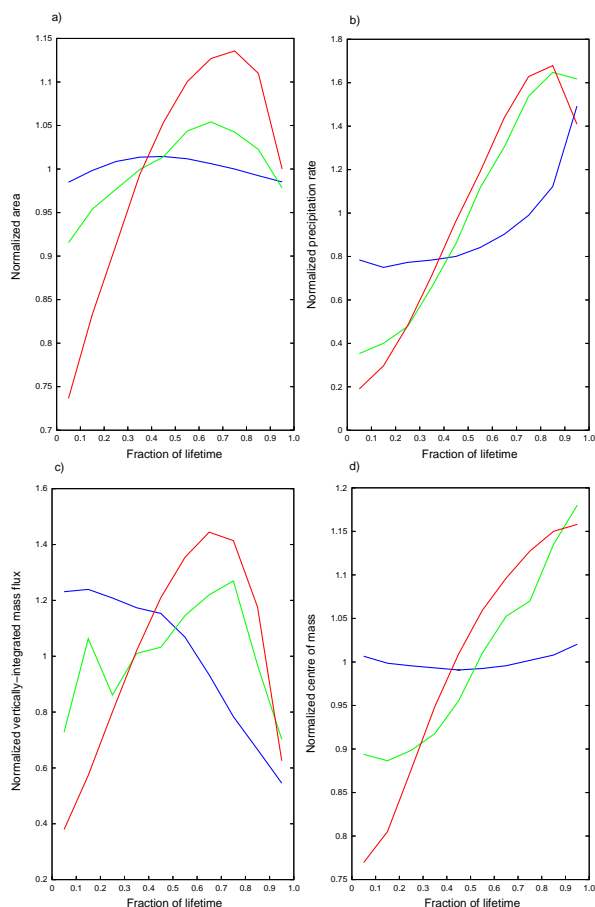
state does not vary in time once equilibrium is reached, apart from fluctuations attributable to the finite size of the domain (Cohen and Craig, 2006). Statistics are presented for 4617 lifecycles of convective cores that are tracked during the remainder of the simulation. Table 1 summarizes some basic statistics of interest. Although there are some isolated single cloudy grid boxes present, it is clear that the portion of the domain containing cloudy grid boxes (moist, buoyant updrafts) remains well captured when the grid boxes are combined into cloud objects. Complicated events are seen to be rare, as anticipated, but splits and mergers are not unusual.

The statistics for the proportions of various events are potentially very sensitive to the removal criteria applied to the cloud-object library (Sect. 2.3). However, in test runs allowing up to 40 generations and reducing the required associations to 0.01, the same proportions were produced to within 5%.

#### 3.1 Convective core lifetimes

Figure 3 shows the distribution of lifetimes (as defined in Sect. 2.3) for the convective-core lifecycles, both for all lifecycles (panel a), and for those whose time histories do not contain any events (panel b). 54.2% of the lifecycles do not contain any events, and these have a mean lifetime of 29.7 min. This is broadly consistent with observational studies on the lifetime of individual, isolated cells (e.g. Foote and Mohr, 1979; Westcott, 1984; Wilson et al., 1998). The lifetime distribution appears to be approximately exponential up to  $\sim 45$  min, with a small peak for around 60 to 75 min. The rapid fall-off for lifetimes in the range 10 to 60 min is again qualitatively consistent with observations (see, for example, the distributions of radar cell lifetimes in Fig. 3 of Foote and Mohr, 1979, Fig. 8 of Wiggert et al., 1981, Fig. 12 of López et al., 1984 and Fig. 17 of Weusthoff and Hauf, 2008b).

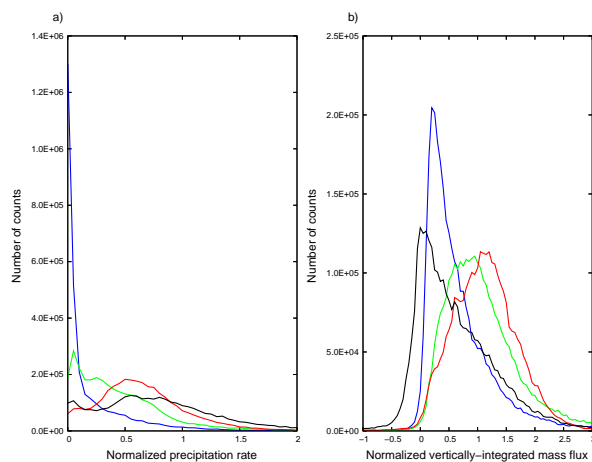




**Fig. 4.** Timeseries of composited lifecycles, showing the mean evolution across the lifecycle of (a) cloud area, (b) precipitation rate, (c) vertically-integrated mass flux, and (d) the centre of mass. The normalized lifecycle has been divided into bins of 0.1. Each panel shows three lines, each of which corresponds to a composite constructed from lifecycles having a particular range of lifetimes: 5 to 30 min (blue), 30 to 60 min (green) and longer than 60 min (red).

Including the lifecycles which do contain events enhances the proportion of the long-lasting lifecycles, raising the mean lifetime to 54.6 min. Again, this chimes with observations of radar cells, which show that merged echoes persist for significantly longer than unmerged echoes (Wiggert et al., 1981; Westcott, 1984, 1994; Wilson et al., 1998). In Sect. 3.2 we discuss further the effects of events on the convective core lifetime.

Examining timeseries for individual lifecycles shows that (as expected) there is strong lifecycle-to-lifecycle variability, with qualitative differences in the development. Nonetheless, it is possible to draw out some general properties of the simulated, convective-core lifecycles by normalizing the timeseries of each lifecycle and then compositing these to produce an averaged lifecycle. López et al. (1984) and Weusthoff and Hauf (2008b) have also attempted similar



**Fig. 5.** Frequency distribution functions at various stages of the convective-core lifecycle of the normalized (a) precipitation rate, and (b) vertically-integrated mass flux. Each panel shows four lines, each of which corresponds to a range of the normalized lifetime: 0 to 0.1 (blue), 0.3 to 0.4 (green), 0.6 to 0.7 (red) and 0.9 to 1.0 (black).

composites for radar cells. Here, the time is normalized using the lifetime, and each cloud property is normalized for each lifecycle by its time-mean value across the lifecycle. Figure 4 shows such composites for various cloud properties, and for lifecycles with different ranges of lifetime. Panel (d) shows the centre of mass,  $h$ , which is defined as the mass-weighted first moment of  $C$ , the mixing ratio of the total condensed water (here, the sum of rain, snow and cloud liquid water, graupel and ice).

$$h = \frac{\int C z \rho dz}{\int C \rho dz} \quad (9)$$

$z$  is the height and  $\rho$  the density. In order to demonstrate the variability between lifecycles, and to allow comparison with that across lifecycles, Fig. 5 shows frequency distribution functions of normalized cloud properties for selected lifecycle stages.

Clearly the longer-lived lifecycles have more pronounced variations across their lifecycles. This is consistent with the evolution of the composite lifecycles constructed by López et al. (1984); Weusthoff and Hauf (2008b). The short lifecycles (those with lifetimes less than 30 min) receive little support for time development, the vertically-integrated mass flux decreasing monotonically through their composite lifecycle (Fig. 4c). The decrease is rather strong and is most rapid during the latter part of the lifecycle, in qualitative agreement with some of Barnes et al.'s (1996) aircraft observations. The cloud area and centre of mass remain almost constant through the composite short lifecycle (Fig. 4a,d).

Measures of updraft strength (the area and mass flux) for the longer lifecycles (those with lifetimes larger than 30 min) exhibit clear peaks towards the later part of the composite lifecycles, and (as in López et al., 1984; Weusthoff and Hauf, 2008b) seem to have their strongest variations at the start and end of the lifecycles. The longer composite lifecycles increase their centre of mass throughout (Fig. 4d). This may occur due to vertical transport of the normalized condensate, or else because the production of condensate occurs at progressively higher levels through the course of the lifecycles.

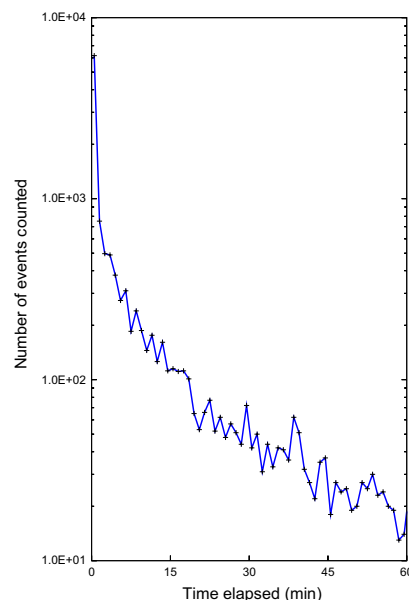
The mass flux frequency distributions (Fig. 5b) are consistent with the composite lifecycles, changing most strongly at the start and end of the lifecycles and with the largest mass fluxes tending to occur a little after the midpoint. The distributions have a broad spread, comparable to or perhaps somewhat larger than the variations seen in the composites across the lifecycle. This suggests that the composites do indeed have value in describing the lifecycle, but that caution should be used in interpreting them as generic lifecycles. Note also that the frequency distributions tend to become more spread as the lifecycle progresses.

The normalized precipitation rate shows a considerable increase across the composite lifecycles and the rates remain relatively large when the lifecycle is terminated. This is consistent with the notion that the lifetime of a convective updraft is similar to the time required for precipitation to develop (Rogers and Yau, 1989). It also highlights the importance of differences between convective cloud definitions. For example, the rain rate of the composite radar cells constructed by Weusthoff and Hauf (2008b) peaks midway through the lifecycle. Moreover, Fig. 5a shows that the use of a precipitation threshold would not capture many of the convective cores during the first 10% of their lifecycles. Certainly then, it would not be appropriate to compare too directly the lifecycle statistics obtained here to statistics obtained by, say, tracking radar echoes. Comparisons of a qualitative nature may nonetheless be reasonable and have been made above. We consider that such comparisons are valuable in order to demonstrate that these model-based statistics are physically plausible, but they are not intended as a detailed assessment of the accuracy of the simulation.

Observationally-based “cloud” identification and tracking methods (such as those cited in Sects. 2.1, 2.2) are strongly constrained by the nature of the available data. Systematic studies of the sensitivity of lifecycle statistics to cloud definition are needed, both to inform comparisons between clouds observed with different systems and to allow comparisons between modelled and observed clouds.

### 3.2 The role of events within the lifecycle

It was shown in Sect. 3.1 that lifecycles containing “events” in the time history had considerably longer lifetimes on average than those without any events. Here we consider the role of events in more detail.



**Fig. 6.** Distribution of the times separating consecutive events within the lifecycles. The vertical scale is logarithmic, and the separation bin size is 1 min.

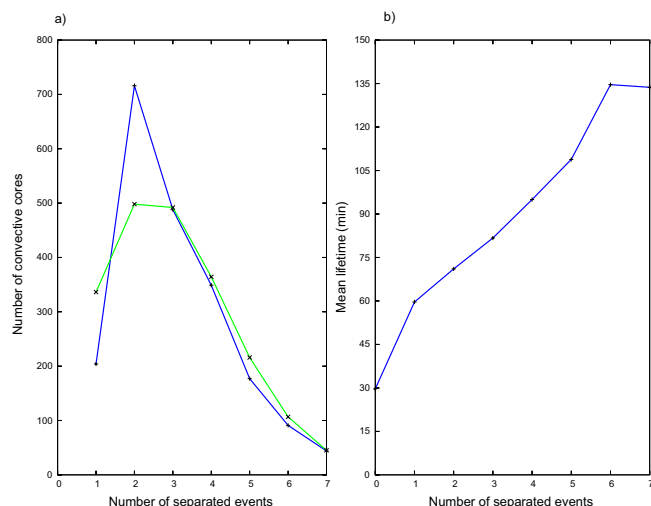
Figure 6 shows the distribution of times that separate consecutive events identified by the tracking algorithm. For separations larger than  $\sim 5$  min, the distribution is roughly exponential. However, many of the events picked-out by the algorithm are quickly followed by other events, often within tens of seconds. Indeed 49.1% of all event separations are less than 1 min. The interpretation is that the joining together or breaking up of cloud objects is rarely a clean process that happens once only at a single timestep. Rather, the joining (for example) of two cloud objects is more typically a somewhat messy affair, perhaps with some portion of the combined object becoming temporarily detached as the constituent parts coalesce to produce what may ultimately become a unified entity.

In order to examine the effects of cloud objects joining up or splitting, it would therefore not be appropriate to rely on the total number of events found by the tracking algorithm as a useful measure. The total reflects not only the number of incidents occurring in a lifecycle, but also how clean or messy those incidents are. Instead, we prefer to define “separated events” as those events satisfying the following criteria.

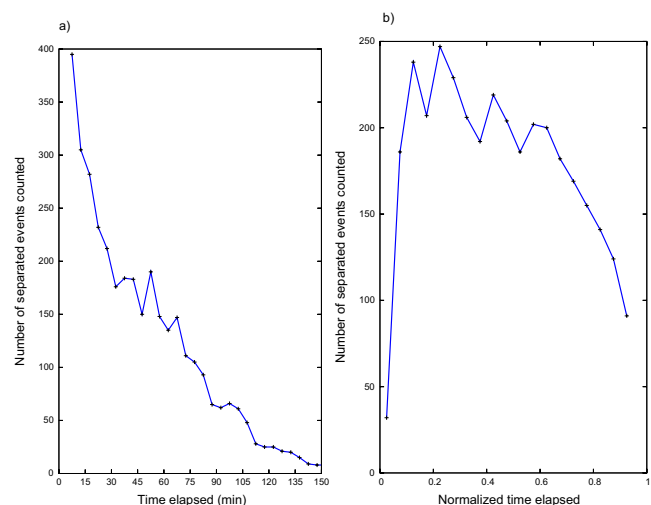
- The event must not take place within 5 min of the start or end of the lifecycle.
- The event must take place at least 5 min after the previous “separated event”.

Thus, the evolution shown in Fig. 1 for example, would be considered to contain one separated event. The choice of





**Fig. 7.** Panel (a) shows the number of lifecycles that contain a given number of separated events (blue). Also shown (green) is a zero-truncated Poisson distribution for the same mean number of separated events. Panel (b) shows the mean lifetime of lifecycles with a given number of separated events.



**Fig. 8.** Distribution of the timings of separated events, both as (a) absolute times after the start of a lifecycle, and as (b) relative times, with the timings being normalized by the lifetime. The bin size is 5 min in (a) and 0.05 in (b).

5 min is a reasonable but somewhat arbitrary one. It is equivalent to a relatively high time resolution that might be available in the data from current operational radar networks (e.g. Weusthoff and Hauf, 2008a). We have checked that our conclusions are not qualitatively affected by reasonable changes of this choice.

Some statistics based on separated events are shown in Figs. 7 and 8. There are 2113 lifecycles that contain sepa-

rated events, and the effects on lifetime are indeed considerable. Each separated event increases the mean lifetime by about 15 min, or around half the mean lifetime of a lifecycle that does not contain any separated events.

A simple-minded explanation for this increase can be provided if it is supposed that each convective core initiated has a mean lifetime of  $\approx 30$  min (irrespective of whether there is another core close by), and that if cores are to be initiated in the vicinity of a pre-existing core then the characteristics and lifecycle-stage of the pre-existing core have no effect on the initiation process. These are strong assumptions, and whether there is any truth in the explanation we leave as a topic for future work. Westcott's comment (1994, p789), based on observations of the merging of radar echoes, that "merging itself may be considered a passive process" is suggestive. More solidly though, there are at least hints in Figs. 7a and 8 that the idea may not be entirely unreasonable. Statistical independence of the cores would imply that the number of events satisfies a zero-truncated Poisson distribution: such a distribution is overlaid on Fig. 7a. Moreover, the timing of separated events within the lifecycles is shown in Fig. 8. Although such events are less likely to occur towards the beginning or end of a lifecycle, the likelihood of events for most of the lifecycle is fairly uniform.

## 4 Conclusions

This paper describes the design and implementation of a novel method for analysing the results of CRM simulations. The algorithm developed operates as an almost self-contained diagnostic suite that is plugged into the model simulation. The objective is to identify clouds from the CRM results and track their evolution. By examining the cloud field on a timestep-to-timestep basis it is possible to exploit the high temporal resolution data available to an online diagnostic system. In conjunction with a numerical stability condition that is satisfied by the advection schemes typically used in CRMs, a simple methodology is sufficient to provide comprehensive and robust tracking. The algorithm has been designed to be as generic as possible. Alternative identification criteria for cloudy grid boxes would be trivial to implement, and most of the decisions about how to examine the lifecycle data are deferred to the postprocessing. Thus, the methodology would be straightforward to adapt in order to track other features online in other models.

The use of the methodology was demonstrated for an idealized simulation of radiative-convective equilibrium. The statistics obtained allow one to quantify the behaviour of the simulated convective clouds in ways that are not accessible to other analysis methods. We chose to track moist, buoyant updrafts, which we referred to as convective cores. Around half of all the cores tracked were subject to merging and splitting during their lifecycles. While this fact complicates the analysis, we were able to demonstrate how cores with simple

and complicated time histories can be considered within a single framework and to demonstrate explicitly the considerable impact of “events” on the core lifetimes.

*Acknowledgement.* This work was supported by a Royal Society grant.

Edited by: H. Wernli

## References

- Arakawa, A. and Schubert, W. H.: Interaction of a cumulus cloud ensemble with the large-scale environment. Part I, *J. Atmos. Sci.*, 31, 674–701, 1974.
- Barnes, G. M., Frankhauser, J. C., and Browning, W. D.: Evolution of the mass flux and diagnosed net lateral mixing in isolated convective clouds, *Mon. Weather Rev.*, 124, 2764–2784, 1996.
- Blyth, A. M., Lasher-Trapp, S. G., and Cooper, W. A.: A study of thermals in cumulus clouds, *Q. J. Roy. Meteor. Soc.*, 131, 1171–1190, 2005.
- Brown, A. R., Derbyshire, S. H., and Mason, P. J.: Large-eddy simulation of stable atmospheric boundary layers with a revised stochastic subgrid model, *Q. J. Roy. Meteor. Soc.*, 120, 1485–1512, 1994.
- Carvalho, L. M. V. and Jones, C.: A satellite method to identify structural properties of mesoscale convective systems based on the maximum spatial correlation tracking technique (MASCOTTE), *J. Appl. Meteorol.*, 40, 1683–1701, 2001.
- Cho, H.-R.: Contributions of cumulus cloud life-cycle effects to the large-scale heat and moisture budget equations, *J. Atmos. Sci.*, 34, 87–97, 1977.
- Cohen, B. G. and Craig, G. C.: The response time of a convective cloud ensemble to a change in forcing, *Q. J. Roy. Meteor. Soc.*, 130, 933–944, 2004.
- Cohen, B. G. and Craig, G. C.: Fluctuations in an equilibrium convective ensemble. Part II: Numerical experiments, *J. Atmos. Sci.*, 63, 2005–2015, 2006.
- Davies, L.: Self organisation of convection as a mechanism for memory, Ph.D. thesis, University of Reading, UK, 165 pp., 2008.
- Dixon, M. and Wiener, G.: TITAN: thunderstorm identification, tracking, analysis and nowcasting – a radar-based methodology, *J. Atmos. Ocean Tech.*, 10, 785–797, 1993.
- Done, J., Davis, C. A., and Weisman, M.: The next generation of NWP: Explicit forecasts of convection using the weather research and forecasting (WRF) model, *Atmos. Sci. Lett.*, 5, 110–117, 2004.
- Foote, G. B. and Mohr, C. G.: Results of a randomized hail suppression experiment in northeast Colorado. Part VI: Post hoc stratification by storm intensity and type, *J. Appl. Meteorol.*, 18, 1589–1600, 1979.
- Fujita, T. T., Fearl, E. W., and Shenk, W. E.: Satellite-tracked cumulus velocities, *J. Appl. Meteorol.*, 14, 407–413, 1975.
- Heus, T.: On the edge of a cloud, Ph.D. thesis, Delft University of Technology, The Netherlands, 142 pp., 2008.
- Hozumi, K., Harimaya, T., and Magono, C.: The size distribution of cumulus clouds as a function of cloud amount, *J. Meteorol. Soc. Jpn.*, 60, 691–699, 1982.
- Kain, J. S.: The Kain–Fritsch convective parameterization: An update, *J. Appl. Meteorol.*, 43, 170–181, 2004.
- Khairoutdinov, M., Randall, D., and DeMott, C.: Simulations of the atmospheric general circulation using a cloud-resolving model as a superparameterization of physical processes, *J. Atmos. Sci.*, 62, 2136–2154, 2005.
- Kuo, K.-S., Welch, R. M., and Weger, R. C.: The three-dimensional structure of cumulus clouds over the ocean. 1: Structural analysis, *J. Geophys. Res.*, 98, 20 865–20 711, 1993.
- LeMone, M. A. and Zipser, E. J.: Cumulonimbus vertical velocity events in GATE. Part I: Diameter, intensity and mass flux, *J. Atmos. Sci.*, 37, 2444–2457, 1980.
- Lennard, R.: The distribution of cumulus cloud sizes, MSc thesis, University of Reading, UK, 67 pp., 2004.
- López, R. E., Blanchard, D. O., Rosenfeld, D., Hiscox, W. L. and Casey, M. J.: Population characteristics, development processes and structure of radar echoes in South Florida, *Mon. Weather Rev.*, 112, 56–75, 1984.
- Machado, L. A. T. and Laurent, H.: The convective system area expansion over Amazonia and its relationships with convective system life duration and high-level wind divergence, *Mon. Weather Rev.*, 132, 714–725, 2004.
- Mason, P. J.: Large-eddy simulation of the convective atmospheric boundary layer, *J. Atmos. Sci.*, 46, 1492–1516, 1989.
- Petch, J. C. and Gray, M. E. B.: Sensitivity studies using a cloud-resolving model simulation of the tropical west Pacific, *Q. J. Roy. Meteor. Soc.*, 127, 2287–2306, 2001.
- Petch, J. C., Brown, A. R., and Gray, M. E. B.: The impact of horizontal resolution on the simulations of convective development over land, *Q. J. Roy. Meteor. Soc.*, 128, 2031–2044, 2002.
- Petch, J. C., Willett, M., Wong, R. Y., and Woolnough, S. J.: Modelling suppressed and active convection. Comparing a numerical weather prediction, cloud-resolving and single-column model, *Q. J. Roy. Meteor. Soc.*, 133, 1087–1100, 2007.
- Plank, V. G.: The size distribution of cumulus clouds in representative Florida populations, *J. Appl. Meteorol.*, 8, 46–67, 1969.
- Plant, R. S. and Craig, G. C.: A stochastic parameterization for deep convection based on equilibrium statistics, *J. Atmos. Sci.*, 65, 87–105, 2008.
- Roberts, N. M. and Lean, H. W.: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events, *Mon. Weather Rev.*, 136, 78–97, 2008.
- Rogers, R. R. and Yau, M. K.: *A Short Course in Cloud Physics*, 3rd Ed., Butterworth-Heinemann, Woburn, MA, USA, 304 pp., 1989.
- Scorer, R. S. and Ludlam, F. H.: Bubble theory of penetrative convection, *Q. J. Roy. Meteor. Soc.*, 79, 94–103, 1953.
- Siebesma, A. P. and Cuijpers, J.: Evaluation of parametric assumptions for shallow cumulus convection, *J. Atmos. Sci.*, 52, 650–666, 1995.
- Siebesma, A. P., Bretherton, C. S., Brown, A., Chlond, A., Cuxart, J., Duynkerke, P. G., Jiang, H., Khairoutdinov, M., Lewellen, D., Moeng, C. H., Sanchez, E., Stevens, B., and Stevens, D. E.: A large eddy simulation intercomparison study of shallow cumulus convection, *J. Atmos. Sci.*, 60, 1201–1219, 2003.
- Swann, H.: Evaluation of the mass-flux approach to parametrizing deep convection, *Q. J. Roy. Meteor. Soc.*, 127, 1239–1260, 2001.
- Theusner, M. and Hauf, T.: A study on the small scale precipitation structure over Germany using the radar network of the German weather service, *Meteorol. Z.*, 13, 311–322, 2004.
- Westcott, N.: A historical perspective on cloud mergers, *B. Am.*

- Meteor. Soc., 65, 220–226, 1984.
- Westcott, N.: Merging of convective clouds: Cloud initiation, bridging and subsequent growth, *Mon. Weather Rev.*, 122, 780–790, 1994.
- Weusthoff, T. and Hauf, T.: The life cycle of convective-shower clouds under post-frontal conditions, *Q. J. Roy. Meteor. Soc.*, 134, 841–857, 2008a.
- Weusthoff, T. and Hauf, T.: Basic characteristics of post-frontal shower precipitation rates, *Meteorol. Z.*, 17, 793–805, 2008b.
- Wiggert, V., Lockett, G. J., and Ostlund, S. S.: Radar rainshower growth histories and variations with wind speed, echo motion, location and merger status, *Mon. Weather Rev.*, 109, 1467–1494.
- Wilson, J. W., Crook, N. A., Mueller, C. K., Sun, J., and Dixon, M.: Nowcasting thunderstorms: A status report, *B. Am. Meteor. Soc.*, 79, 2079–2099, 1998.
- Xu, K.-M. and Randall, D. A.: Updraft and downdraft statistics of simulated tropical and midlatitude cumulus convection, *J. Atmos. Sci.*, 58, 1630–1649, 2001.
- Yano, J.-I., Guichard, F., Lafore, J.-P., Redelsperger, J.-L., and Bechtold, P.: Estimations of mass fluxes for cumulus parameterizations from high-resolution spatial data, *J. Atmos. Sci.*, 61, 829–842, 2004.
- Zipser, E. J. and LeMone, M. A.: Cumulonimbus vertical velocity events in GATE. Part II: Synthesis and model core structure, *J. Atmos. Sci.*, 37, 2458–2469, 1980.
- Zhao M. and Austin P. H.: Life cycle of numerically simulated shallow cumulus clouds. Part I: Transport, *J. Atmos. Sci.*, 62, 1269–1290, 2005a.
- Zhao M. and Austin P. H.: Life cycle of numerically simulated shallow cumulus clouds. Part II: Mixing dynamics, *J. Atmos. Sci.*, 62, 1291–1310, 2005b.