Quarterly Journal of the
Royal Meteorological Society

Large-scale length and time scales for use with stochastic convective parameterization

Journal:	*QJRMS*
Manuscript ID:	QJ-11-0113.R1
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Keane, Richard; Meteorologisches Institut, Physics Plant, Robert; University of Reading, Meteorology
Keywords:	stochastic parameterization, convection, ensemble averaging

Quarterly Journal of the Royal Meteorological Society *Q. J. R. Meteorol. Soc.* **00: 1–20 (0000)**

Large-scale length and time scales for use with stochastic convective parameterization

R. J. Keane,[a]* R. S. Plant[a]

[a]*Department of Meteorology, University of Reading, Earley Gate, PO Box 243, Reading, RG6 6BB, UK*

*Correspondence to: Meteorologishes Institut, Ludwig-Maximilians-Universität München, Fakultät für Physik, Theresienstr. 37, 80333 München, Germany. E-mail: richard.keane@lmu.de

Many numerical models for weather prediction and climate studies are run at resolutions that are too coarse to resolve convection explicitly, but too fine to justify the local equilibrium assumed by conventional convective parameterizations. The Plant-Craig (PC) stochastic convective parameterization scheme, developed in this paper, solves this problem by removing the assumption that a given grid-scale situation must always produce the same sub-grid-scale convective response. Instead, for each timestep and gridpoint, one of the many possible convective responses consistent with the large-scale situation is randomly selected. The scheme requires as input the large-scale state as opposed to the instantaneous grid-scale state, but must nonetheless be able to account for genuine variations in the large-scale situation. Here we investigate the behaviour of the PC scheme in three-dimensional simulations of radiative-convective equilibrium, demonstrating in particular that the necessary space-time averaging required to produce a good representation of the input large-scale state is not in conflict with the requirement to capture large-scale variations. The resulting equilibrium profiles agree well with those obtained from established deterministic schemes, and with corresponding cloud-resolving model simulations. Unlike the conventional schemes the statistics for mass flux and rainfall variability from the PC scheme also agree well with relevant theory and vary appropriately with spatial scale. The scheme is further shown to adapt automatically to changes in grid length and in forcing strength. Copyright © 0000 Royal Meteorological Society

Key Words: stochastic parameterization; convection; ensemble averaging

Received ¡Article history¿

Citation: . . .

1. Introduction

Forecast errors in numerical simulations of the atmosphere arise from uncertainties in the initial and boundary conditions, in the external forcings, and also in the model formulation itself. Structural model uncertainties primarily relate to unresolved or poorly-resolved physical processes that must be parameterized. These uncertainties are important because small-scale fluctuations can interact with grid-scale dynamics, leading to upscale growth of

errors (Zhang *et al.* 2003, for example). This contributes to the spreading of ensemble weather forecasts, and as a consequence the ensemble spreads produced by systems accounting for initial-condition uncertainty alone are often insufficient to cover the full range of possible flows (Buizza 1997; Buizza *et al.* 2005). The variability can be increased in such systems, so that it is more representative of the real atmosphere, but only at the cost of carefully inflating the initial condition perturbations (Bowler *et al.* 2008, for example).

Over the last decade, various methods have been introduced in order to provide explicit representations of some of the uncertainties arising from the parameterizations in numerical weather prediction (NWP) and general circulation models (GCMs). Although some of the methods are somewhat ad hoc, they are nonetheless beneficial and can provide more satisfactory treatments of model uncertainty than simple inflation of initial-condition perturbations. For example, Buizza *et al.* (1999) add a random perturbation to the tendencies obtained from the parameterization schemes. Other authors have included stochastic elements directly in a parameterization scheme itself. As examples, Bright and Mullen (2002) introduced a stochastic component to the trigger function of a convection scheme, Lin and Neelin (2003) added random perturbations to the CAPE closure and, separately, to the vertical heating profile, and Shutts (2005) developed a stochastic kinetic energy backscatter scheme, where a fraction of the energy dissipated by the model grid truncation is reintroduced near the model grid scale. A good overview of current methods can be found in a recent book (Williams and Palmer 2009).

The Plant and Craig (2008, hereinafter PC) stochastic convection parameterization scheme is designed to produce physically-realistic convective variability and to adapt automatically to changes in the resolution of the parent model, down to grid lengths of the order of $10\,\mathrm{km}$. For large grid lengths, where the variability is suitably low, the scheme should agree with results from conventional deterministic convection schemes. Plant and Craig (2008) demonstrated these features in some single-column radiative-convective equilibrium (RCE) experiments, while Ball and Plant (2008) demonstrated that for grid lengths of $\sim 50\mathrm{km}$ then the convective fluctuations it produces become of comparable importance to the variability produced by some of the generic methods in use for representing structural model uncertainty. Thus, the PC scheme would appear to be particularly well suited for use at grid lengths of ~ 10 to $50\mathrm{km}$, and even on variable resolution grids. However, the experiments just cited used single-column models. The implementation of any stochastic parameterization within a full, three-dimensional atmospheric model raises important scientific issues about the spatial and temporal scales associated with the parameterization and their relationship to the scales of the numerical model. The present article will establish the extent to which the input profile to the scheme must be averaged in order to reproduce the correct convective variability and to adapt appropriately to model resolution.

The PC scheme is based on the statistical mechanics theory of Craig and Cohen (2006) for non-interacting clouds at equilibrium, and is supported by the results from cloud-resolving models (CRMs) in RCE (Cohen and Craig 2006; Davoudi *et al.* 2010). Conventional convective parameterizations are deterministic, so that the same sub-grid-scale convective response is always output for a given grid-scale input. However, CRMs clearly demonstrate that a wide range of convective states are consistent with a given grid-scale situation, for the grid sizes currently used in NWP and GCMs (Xu *et al.* 1992; Cohen and Craig 2006; Shutts and Palmer 2007; Jones and Randall 2011). The Craig and Cohen (2006) theory predicts analytical formulae for convective statistics which can be compared with the results produced by a convective parameterization. We will perform such tests at different grid lengths, in order to prove that the PC scheme does indeed operate in a fully self-consistent way, independent of resolution. Results will also

be compared with those of deterministic parameterization schemes, and with CRM results.

Three-dimensional RCE simulations are performed with a specified tropospheric cooling rate and parameterized convection over a uniform sea surface. The setup is described in Section 2, alongside key points from the theory against which the simulated convective variability is to be compared. The comparison itself is presented in Section 3, and used to determine parameter settings and stratgeies for use of the PC scheme in three dimensions. The theory is extended in Section 4 to study fluctuations in the rainfall rate, enabling the PC scheme to be compared directly to other convection schemes. Finally, Section 5 discusses the implications of the results.

2. Methodology

2.1. *Theory of convective variability*

In order to make the present article self-contained, we provide here an overview of the theory for convective variability in equilibrium, and in the following subsection an overview of its implementation within the PC scheme.

An equilibrium condition supports an ensemble of possible states for the cumulus cloud field. Here, a 'cloud' is defined as an updraft (or updraft-downdraft pair) with mass flux due to a vertical velocity above some threshold. The possible cloud states are described by the mass fluxes $m(z)$ for each of the variable number of clouds present. Following normal parameterization practice a description of the mass flux evolution over the lifetime of each cloud is not considered, and so the mass flux should be regarded as a lifetime-averaged value. Using statistical mechanics theory for non-interacting clouds, Craig and Cohen (2006) showed that the probability distribution function (PDF) of mass flux for the individual clouds is given by

$$p(m)\mathrm{d}m = \frac{1}{\langle m\rangle}\mathrm{e}^{-m/\langle m\rangle}\mathrm{d}m, \qquad (1)$$

where the angled brackets denote an ensemble average. The distribution has been verified in CRM experiments (Cohen and Craig 2006; Davies 2008; Davoudi *et al.* 2010) and is robust for different heights and large-scale environments.

Following Plant and Craig (2008), for the experiments in this paper we take $\langle m\rangle$ at the lifting condensation level (LCL) to be a constant, $\langle m\rangle = 2\times10^7\,\mathrm{kgs}^{-1}$. Various CRM studies (e.g. Robe and Emanuel 1996; Shutts and Gray 1999; Cohen 2001; Parodi and Emanuel 2009) have shown that the strength of the forcing seems to have only a weak effect on the mean mass flux of individual clouds: rather, a change in forcing is associated mainly with a change in the mean number of clouds $\langle N\rangle$. A recent CRM investigation by Davies and Jakob (2011) shows that the vertical profile of $\langle m\rangle$ may depend on some rather subtle changes to the character of the forcing, but nonetheless it finds very little sensitivity in $\langle m\rangle$ in the lower atmosphere. In the observational literature, differences in typical convective core strengths between different regions and over different surface types have been identified and discussed (e.g. LeMone and Zipser 1980; Lucas *et al.* 1994; May and Rajopadhyaya 1999) but we are not aware of studies indicating systematic variations of $\langle m\rangle$ at the LCL that might usefully be incorporated into the parameterization.

The total mass flux will be denoted M, and its ensemble-mean value $\langle M\rangle$ can be taken to be known because of the equilibrium assumption. Given this constraint, a PDF for the total mass flux can be calculated, following Craig and Cohen (2006), as

$$p(M) = \delta(M)\mathrm{e}^{-\frac{\langle M\rangle}{\langle m\rangle}} + \frac{1}{\langle m\rangle}\sqrt{\frac{\langle M\rangle}{M}}\mathrm{e}^{-\frac{M+\langle M\rangle}{\langle m\rangle}} I_1\left(\frac{2}{\langle m\rangle}\sqrt{M\langle M\rangle}\right), \qquad (2)$$

where I_1 is the modified Bessel function of order 1. This distribution has also been verified in CRM experiments (Cohen and Craig 2006; Davoudi *et al.* 2010). In section 3 we will investigate whether it can be reproduced in

three-dimensional RCE experiments with parameterized convection. It will be convenient there to summarize some of the results by comparing with the theoretical prediction for the variance of M, which is

$$\langle (M - \langle M \rangle)^2 \rangle = 2 \langle M \rangle \langle m \rangle. \quad (3)$$

2.2. *Implementation as a stochastic parameterization*

The PC scheme is based on the theory outlined in the previous subsection. Full details are available in Plant and Craig (2008) but we present here a summary of the main aspects that are relevant for the present study. The first step is that the ensemble-mean mass flux at the LCL, $\langle M \rangle$, must be computed from the closure conditions of the scheme. Postponing our discussion of this crucial issue for the moment, the next step is then to determine the mass fluxes of individual clouds that constitute a possible, particular realization of the convective state at the grid box and time in question.

To model each individual cloud, the plume model from the Kain-Fritsch parameterization (Kain and Fritsch 1990; Kain 2004, hereinafter KF) is used, adapted to handle a spectrum of cloud types as described in Plant and Craig (2008). The KF scheme entraining/detraining plume model uses a buoyancy sorting approach in which it considers various possible mixtures of updraft and environmental air and retains each mixture in the updraft or else detrains it to the environment according to whether the mixture is positively or negatively buoyant, respectively. The calculations require a maximum entrainment rate to be specified, which is taken to be inversely proportional to the updraft radius,

$$\epsilon = \frac{-0.03 \delta p}{r} \quad (4)$$

where ϵ is the maximum entrainment per unit of mass flux within a pressure interval δp and for an updraft radius of r. Full details are given in Kain and Fritsch (1990). In the PC scheme the updraft radius is related to cloud mass flux by assuming that close to the LCL the mass flux varies only with the horizontal area of the cloud,

$$m = \frac{\langle m \rangle}{\langle r^2 \rangle} r^2. \quad (5)$$

Using the above equation, the PDF of cloud mass fluxes in equation 1 can be transformed into a PDF $p(r)$ of cloud radii (and, therefore, of entrainment rates):

$$p(r) \mathrm{d}r = \frac{2r}{\langle r^2 \rangle} \exp\left(\frac{-r^2}{\langle r^2 \rangle}\right) \mathrm{d}r. \quad (6)$$

This equation is used to determine how many clouds to initiate, and of what sizes, within a given grid box in the PC scheme. Because it applies to a single cloud, it is rescaled to account for the fact that a number of clouds can be present in a grid box, by multiplying by $\langle N \rangle$. This is obtained from the large-scale state by using $\langle N \rangle = \langle M \rangle / \langle m \rangle$. It is also multiplied by a factor $\mathrm{d}t/T_\mathrm{L}$ to allow for the finite lifetime T_L of the clouds ($\mathrm{d}t$ is the model timestep): proportionally fewer clouds are initiated to allow for the fact that they persist for multiple timesteps, as in the KF scheme. The initiation probability, then, within a model timestep for a plume with radius r to $r + \mathrm{d}r$ is as follows:

$$\frac{\langle M \rangle}{\langle m \rangle} p(r) \mathrm{d}r \frac{\mathrm{d}t}{T_\mathrm{L}} = \frac{\langle M \rangle}{\langle m \rangle} \frac{2r}{\langle r^2 \rangle} \exp\left(\frac{-r^2}{\langle r^2 \rangle}\right) \mathrm{d}r \frac{\mathrm{d}t}{T_\mathrm{L}}. \quad (7)$$

Clouds are initiated at random from that PDF by binning the cloud spectrum into finite bins of width $\mathrm{d}r$. The bin width is chosen so that the chance of two clouds occuring in the same bin is suitably negligible. For each bin, a random number uniformly distributed between zero and unity is generated and compared to the initiation probability. If the random number is lower then a cloud is initiated of that size. This paper follows Plant and Craig (2008) by setting $T_\mathrm{L} = 45\,\mathrm{min}$, a constant. It is easy to check that the normalization of Eq. 7 is such that the average number of clouds present is $\langle N \rangle = \langle M \rangle / \langle m \rangle$.

The persistence of clouds across multiple timesteps introduces in a very natural way an element of temporal coherence to the stochastic variability. Further temporal

and spatial coherence is inherent to the scheme, in that it is explicitly designed to operate with nonlocal inputs, as we discuss shortly. The extent of the nonlocality is likely to be important for the upscale impact of the stochastic fluctuations (e.g. Buizza *et al.* 1999). A source of noise that was entirely independent for each model grid point and timestep may be largely washed-out through numerical diffusion, however physical the noise source might be. The experiments described here will establish the impact of coherency that arises from intrinsic, physical correlation scales of deep convection.

At each timestep clouds that have been present for time $T_{\rm L}$ are removed, and new clouds are initiated. The KF plume model is used to compute vertical profiles of convective tendencies for each cloud and these are summed over all the clouds to feed back to the model dynamics. If the size of the grid box is large enough then it will contain a representative sample of the full spectrum of clouds, and the tendencies will tend to those from a deterministic scheme; in this case then the grid-box state will provide a good approximation to the large-scale environmental state and can be used directly in order to compute $\langle M \rangle$.

More generally, time and space averaging across multiple grid boxes is necessary in order to obtain a large-scale state suitable for use in the computation of the closure. By definition, the cloud-base mass flux to be computed as the closure of any mass-flux parameterization is a function of the large-scale forcing (Plant 2010). Conventionally, the local, instantaneous grid-box state is used to approximate the large-scale environment. However, the grid-box state is subject to fluctuations as traditional deterministic parameterizations often produce on-off behaviour and strong timestep-to-timestep variability (Willett and Milton 2006; Stiller 2009). In order to remove these fluctuations, and so obtain a large-scale environmental state from which $\langle M \rangle$ can properly be calculated, the grid-scale atmospheric state should be averaged in space and in time, over neighbouring grid points and recent time values. This averaging must be over a sufficient number of points that the resulting $\langle M \rangle$ can indeed be identified with the ensemble-mean total mass flux for the large-scale environment, but not over so many points that a significant fraction of them are not representative of the large-scale environment represented by the grid point at which the convection is being calculated (it should be emphasized that the averaging is not intended to transmit information about the environment at distant grid boxes to the grid box in question, rather to use distant grid boxes to provide more information about the environment at the grid box in question). In other words, the necessary averaging should not obscure genuine variations in the large-scale forcing. It is not at all obvious that both conditions can be satisfied simultaneously to produce a suitable slowly-varying $\langle M \rangle$, a point to be tested explicitly here.

For stochastic parameterizations such as PC, in which fluctuations are predicted and controlled, then the importance of the distinction between the grid-box state and the large-scale state is readily apparent. However, we stress that the distinction is potentially an important one for the representation of any parameterized process which does not have a clear scale separation between the intrinsic scales of the process and the grid scale. This point has actually been recognized for some time (e.g. Lander and Hoskins 1997; Kuo *et al.* 1997, p477) but practitioners have so far shied away from the natural consequence that aspects of parameterization should be non-local.

In the PC scheme, the basis for the closure is the removal of CAPE, following KF and several other parameterizations (e.g. Zhang and McFarlane 1995; Bechtold *et al.* 2008). Specifically $\langle M \rangle$ is defined as the total mass flux required to remove 90% of CAPE through the action of convection within a closure time $T_{\rm c}$. An appropriate value for $T_{\rm c}$ is discussed in Section 2.4.

2.3. *Averaging scales*

Studies by Ricciardulli and Sardeshmukh (2002) and Smith *et al.* (2005) suggest that the spatial correlation length of tropical convection on a 3-hourly time scale is of the

order of 120 km (we interpret the term "correlation length", here, to mean lengths over which variations in the large-scale environment are not significant). The number varies somewhat over different regions (and is generally slightly higher over the oceans than over land). These studies suggest that variations of the large-scale environment for convection are rather modest on scales of around 100 km, and so we might consider this to be a suitable upper limit for spatial averaging (in each direction) of the input to a closure calculation, in order to ensure that variations of the large-scale environment within the averaging area do not adversely affect the calculation. Another study which lends support to this proposal is that of Moron *et al.* (2007), who found that the correlation scale for tropical rain amounts on wet days, on daily timescales, is never more than about 100 km.

On the other hand, the spatial averaging must also be sufficient to smooth out local fluctuations that we would not wish to consider as part of the large-scale environment. The CRM study by Shutts and Palmer (2007) found that coarse-graining at a scale of 120 km or less was required in order to obtain the full range of convective behaviour, in an idealized experiment relevant to the tropics. Domain sizes of a similar order of magnitude have been successfully used in many CRM experiments (e.g. Robe and Emanuel 1996; Shutts and Gray 1999; Cohen and Craig 2006; Davies 2008; Davoudi *et al.* 2010), suggesting that it should indeed be possible to obtain sufficient statistical averaging on a scale of around 100 km.

The studies of Ricciardulli and Sardeshmukh (2002) and Holloway and Neelin (2010) estimate typical durations of "wet events" and autocorrelation timescales for tropical precipitation to be on the order of a few hours. Following similar reasoning as for the spatial averaging, these correlation times provide a suitable upper limit for the temporal averaging that might reasonably be used for the input to the closure calculations of a parameterization. Variations of the large-scale environment should be rather modest on scales of less than a few hours, but one would not wish to take an average over longer timescales on which the diurnal variation of land surface temperature (for example) would become a significant aspect of the large-scale environmental forcing for convection.

Subject to the above upper limits on averaging scales, a major purpose of this paper is to determine practical lower limits for averaging scales that are required in order to obtain an accurate estimate of $\langle M \rangle$ in the CAPE closure calculations, as judged by the ability of the PC scheme to yield the correct theoretical equilibrium PDF for M (i.e., that which is self-consistent with its design principles). The setup used for doing this is described in the next subsection.

2.4. *Radiative-convective equilibrium setup*

The setup used to investigate the statistics of parameterized convection is an idealized configuration of the UK Met Office Unified Model (Davies *et al.* 2005, MetUM), running at version 6.1. The boundary layer and large-scale cloud parameterizations are used with the standard settings of the MetUM, as descibed in Lock *et al.* (2000) and Wilson and Ballard (1999) respectively.

We use the term radiative-convective equilibrium in its generic sense, as being the outcome from integrating a model of convection for a long period with a time-invariant forcing being imposed. By forcing we refer to any process that would act to destabilize the atmosphere in the absence of convective activity. The forcing imposed is extremely simple here, a fixed tropospheric cooling rate being applied as follows:

$$\left.\frac{\partial T}{\partial t}\right|_{\text{forced}} = \begin{cases} -\dot{T}_0 & p > 2p_0 \\ \dot{T}_0 (p_0 - p)/p_0 & 2p_0 > p > p_0 \\ 0 & p_0 > p \end{cases} \quad (8)$$

where $p_0 = 200\,\text{hPa}$ and $\dot{T}_0$ is a constant defining the strength of the forcing. The domain is a homogeneous sea surface with a constant sea surface temperature of 300 K. The Coriolis parameter is set to zero. There is no wind

profile imposed in the simulations, and so all winds that occur are induced by the convection.

The domain is three-dimensional and of size $512\,\mathrm{km}$ in each horizontal direction. Bicyclic boundary conditions are applied. The default choices for the grid length and the cooling rate are $32\,\mathrm{km}$ and $\dot{T}_0 = 8\,\mathrm{Kday}^{-1}$ respectively, although other choices, namely $16\,\mathrm{km}$ and $51.2\,\mathrm{km}$ for the grid length and $10\,\mathrm{Kday}^{-1}$ and $12\,\mathrm{Kday}^{-1}$ for the cooling rate, are also explored. The model has 49 levels in the vertical. The initial conditions were a horizontally homogeneous atmosphere at rest, with vertical profiles of temperature and moisture being taken from the equilibrium state of an equivalent CRM experiment by Cohen and Craig (2006), which is forced in the same way and which has a tropospheric cooling rate of $\dot{T}_0 = 8\,\mathrm{Kday}^{-1}$. The CRM profile is plotted in figures 1 and 2. Since the initial conditions were from a state of RCE, albeit from a different model, the MetUM was able to reach its own RCE state rather quickly, and all MetUM simulations had reliably reached equilibrium after 5 days. The simulations were each run for 25 days.

It is somewhat unusual, but by no means unprecedented, to perform idealized RCE experiments at resolutions which require convection to be parameterized. Given that almost all convective parameterizations are based on quasi-equilibrium thinking (Emanuel 2000), it is perhaps surprising that such tests are not a standard part of parameterization development, and given that the tests are far from trivial we echo the remarks of Held *et al.* (2007) that further studies with such configurations would be beneficial. In the present context, we assert that a convective parameterization must behave in accordance with its own design principles in an RCE configuration, or else it is scarcely likely to behave adequately in more complex situations. Alongside runs with the PC scheme, we have also performed comparison experiments with the KF scheme, and with the standard convection scheme of the MetUM, as described in Martin *et al.* (2006) and based on the scheme of Gregory and Rowntree (1990, hereinafter GR). Previous studies using a convection parameterization in a similar RCE setup have focussed on the organized structures that are established in large domains with interactive radiation (e.g. Held *et al.* 2007; Held and Zhao 2008). The study of Larson and Hartmann (2003) also employed interactive radiation, and demonstrated that the KF parameterization is capable of producing a realistic large-scale mean state in such experiments, at least in terms of its response to changes in sea-surface temperature.

As discussed by Held *et al.* (2007), RCE experiments with parameterized convection can exhibit "gridpoint storms" if an instability is not removed by the convection scheme, but is instead manifest as intense localized precipitation produced by the large-scale cloud parameterization. Preliminary experiments with our setup showed that the simplest way to eliminate such storms was to choose a sufficiently short CAPE closure timescale T_c. However, it was also found that with too short a T_c then the convection scheme responds too strongly and "overcompensates", leading to oscillations in the rainfall, even if averaged over the whole domain. The simulations to be presented here use a constant $T_\mathrm{c} = 75\,\mathrm{min}$, except for the case with a $51.2\,\mathrm{km}$ grid length, where a value of $120\,\mathrm{min}$ is used. These values are compatible with the range of values found in the convective parameterization literature. Moreover, a systematic reducing in the closure timescale with reducing gridlength has been found to be beneficial for the partitioning between convective and large-scale processes in the ECMWF model (Bechtold *et al.* 2008), a point that seems likely to be related to the systematic increases in the vertical velocities of explicitly resolved motions at reduced gridlengths (e.g. Paulius and Garner 2006).

The numerical experiments will determine whether the PC scheme can reproduce the theoretically-expected PDF of total mass flux M (equation 2) under large-scale forcing conditions for which the theory holds. The idealized setup used here provides just such conditions, the fixed cooling ensuring that once the system has reached equilibrium then the total mass flux $\langle M \rangle$ required to balance the forcing

is constant. In these conditions, the upper limit on an acceptable averaging area is removed since at all times the entire domain is representative of the same large-scale environment. This allows us to determine the lower limits for temporal and spatial averaging of the input to the closure calculations that are required in order to obtain a sufficiently accurate representation of the large-scale environment at each grid box and timestep, so that $\langle M \rangle$ as computed by the closure is sufficiently steady to allow the PC scheme to reproduce equation 2. As an additional test on the theory, and its implementation in the PC scheme, we will also investigate whether the PDF of mass flux per cloud, equation 1, can be reproduced. Although this PDF is imposed by the PC scheme at the LCL, that is not the case at other heights.

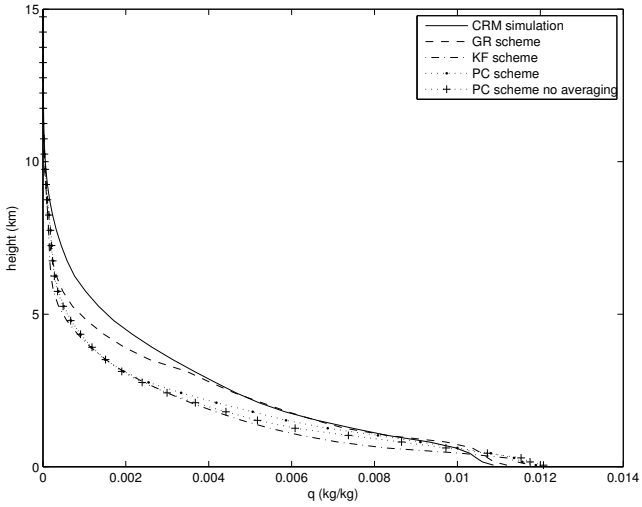

Figure 1. Mean vertical profiles of water vapour q in radiative-convective equilibrium experiments with $8\,\mathrm{K day}^{-1}$ imposed cooling. Results are shown for a cloud-resolving model study (CRM, solid line) and for MetUM experiments using the Gregory-Rowntree scheme (GR, dashed line), the Kain-Fritsch scheme (KF, dash-dotted line) and the Plant-Craig scheme (PC, dotted lines). Results from the PC scheme are shown for standard averaging (dotted line with points) and for no averaging (dotted line with pluses).

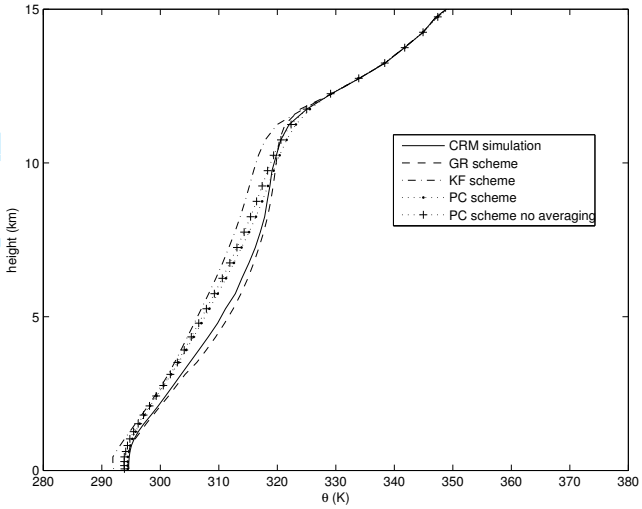

Figure 2. As in figure 1 but showing mean vertical profiles of potential temperature θ.

3. Results for equilibrium state

3.1. *Mean profiles*

Before considering the mass-flux fluctuations produced by the stochastic parameterization scheme, we first test its ability to produce appropriate mean profiles. For the idealized RCE configuration studied here, a suitable comparison is provided by the results from the corresponding CRM experiment by Cohen and Craig (2006). Vertical profiles of water vapour q and potential temperature θ are shown in figures 1 and 2 respectively, for the CRM study and for the three convection schemes used in the current MetUM experiments. The profiles have been averaged over the full model domain in each case, and also daily over the last 20 days of the simulations. The time variations amongst domain-mean profiles were found to be modest, and certainly no larger than the variation amongst the results for different convection schemes. The averaging strategy for the input to the closure calculations of the PC scheme that has been used here is the "standard averaging" to be fully specified in Section 3.3. Also shown are the results from an experiment with the PC scheme and grid-scale input only: i.e., with no space or time averaging of the input.

Also shown, in figures 3 and 4, are the differences between the various profiles produced by the MetUM experiments with parameterized convection and the corresponding CRM profiles. The magnitudes of these differences are similar to those found in corresponding single-column model (SCM) experiments with the MetUM (Plant and Craig 2008). In the current three-dimensional experiments the lower atmosphere is drier than that of

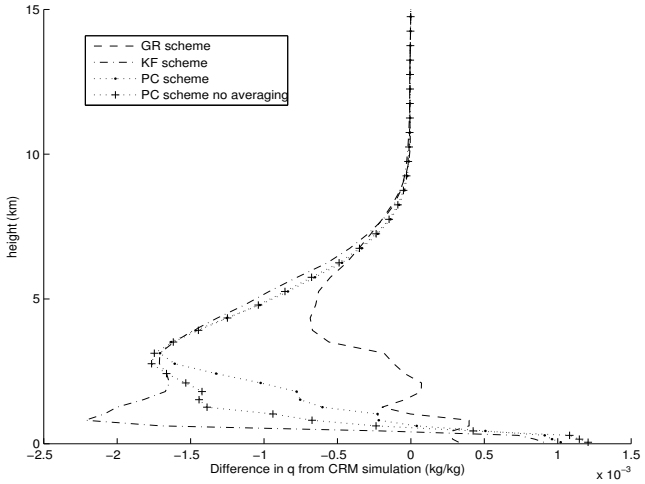

Figure 3. Vertical profiles of the difference between water vapour q from MetUM experiments with parameterized convection scheme and from a corresponding CRM experiment. Results are shown using the Gregory-Rowntree scheme (GR, dashed line), the Kain-Fritsch scheme (KF, dash-dotted line) and the Plant-Craig scheme (PC, dotted lines). Results from the PC scheme are shown for standard averaging (dotted line with points) and for no averaging (dotted line with pluses).

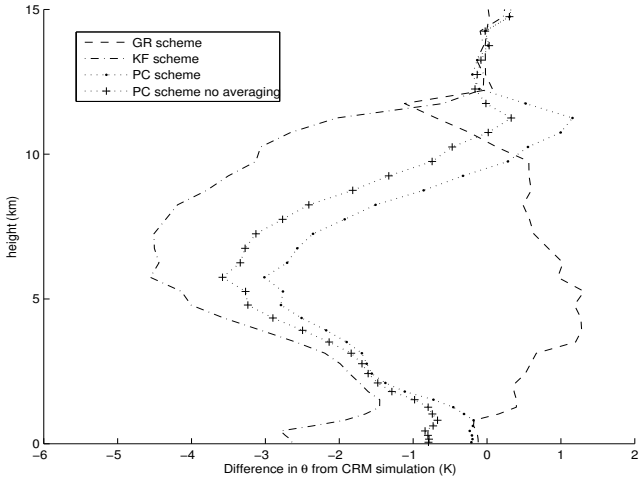

Figure 4. As in figure 3 but showing vertical profiles of the differences in potential temperature θ.

the CRM, whereas it was moister in the earlier SCM experiments. The difference is likely to be due to the different treatments of surface fluxes. In the current experiments the surface fluxes are computed from bulk aerodynamic formulae and so are dependent on the near-surface wind speeds that develop in the simulations. The relative dryness suggests that such wind speeds tend to be somewhat weaker than in the CRM. By contrast the surface fluxes in the comparison SCM were simply taken to be proportional to the moisture difference across the surface layer, with a fixed, pre-defined constant of proprtionality (Plant and Craig 2008).

The GR scheme yields 'better' profiles than the KF scheme (i.e., closer to the CRM results). The profiles from the PC scheme are intermediate between those of the GR and KF schemes, although it is closer to the KF result. This is scarcely a surprise given that it is based on the same plume model and so is essentially a spectral and stochastic generalization of the KF scheme. Averaging the input to the closure calculations of the PC scheme has little effect on the mean state produced, with a small improvement over the experiment with no averaging.

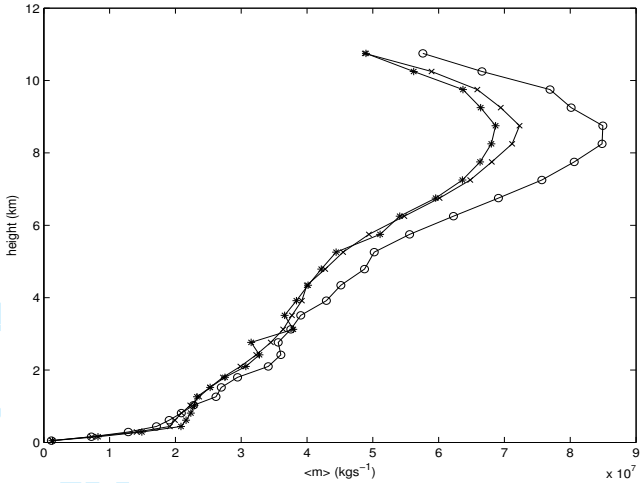

Figure 5. Vertical profile of the mean mass flux per cloud for three different averaging strategies applied to the input of the PC scheme. The strategies are described as no-averaging (open circles), standard-averaging and maximal-averaging (stars) and are detailed in Section 3.3. The values were obtained by averaging the mass flux of every convective plume present at each model level, starting from 5 days into the simulation and continuing until data for 1300 clouds had been recorded.

3.2. *Vertical profile of mean mass flux per cloud*

The vertical profile of mean mass flux per cloud is plotted in figure 5 in order to establish the context for the following results. The profile is plotted with different averaging strategies having been applied to the input of the closure calculations of the PC scheme. The strategies will be fully described in the following subsection. For the present purposes, it is sufficient to note simply that the averaging strategy chosen has little effect on this profile.

The profile can be compared with that of the PC scheme in the SCM (Plant and Craig 2008, their figure 8). The behaviour is broadly similar, although the current

experiments have a lower LCL (recall that $\langle m \rangle = 2 \times 10^7 \mathrm{kgs}^{-1}$ is imposed at the LCL) than in the SCM, and show a smoother increase with height of the mean mass flux per cloud. The peak value of $\langle m \rangle(z)$ is similar but occurs at just over $8\,\mathrm{km}$ in these three-dimensional experiments rather than at just below $10\,\mathrm{km}$ in the SCM.

3.3. *Development of averaging strategy*

We consider first an experiment with a grid length of $32\,\mathrm{km}$ and a forcing $\dot{T}_0 = 8\,\mathrm{Kday}^{-1}$, along with a "maximal" averaging strategy for the input to the PC scheme. Specifically, the closure calculations of the scheme were performed on profiles averaged temporally over $63\,\mathrm{min}$, and averaged spatially over the grid box in question and its neighbours up to seven grid boxes away. Thus, the averaging area was a square of side $480\,\mathrm{km}$, encompassing almost the entire model domain. All grid boxes within the averaging area and all timesteps within the averaging period were treated equally: i.e., no weighting functions were applied in constructing the averages.

The upper panel of figure 6 shows a PDF obtained at $1.52\,\mathrm{km}$ for the mass flux per cloud, m. This height was chosen as it is above the LCL but low enough such that almost all plumes launched by the PC scheme are still buoyant and so will contribute to the PDF. The PDF was constructed from values for the mass flux of each cloud present in the domain recorded at 6 hourly intervals for a period of 20 days. The PDF agrees well with the theoretical prediction from equation 1, as do the results obtained for other heights (not shown).

A PDF for M, the total convective mass flux, is shown in the lower panel of figure 6. The totals are for areas of $(64\,\mathrm{km})^2$ and were computed by partitioning the domain into 8×8 sub-domains (each containing 2×2 grid boxes) and summing over plumes within each sub-domain. The same time-sampling was used as above, giving a total of 5184 values of M for use in constructing the plot shown. Again, the numerical results agree well with the theory, and

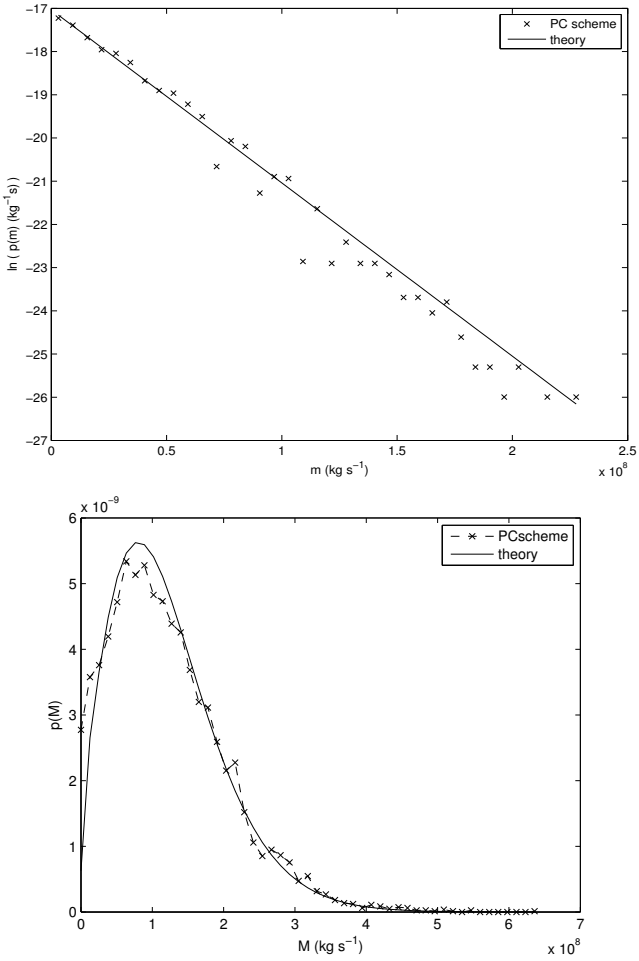

Figure 6. PDFs at a height of $1.52\,\mathrm{km}$ for (top) the mass flux per cloud, m, and (bottom) for the total mass flux M over the horizontal area $(64\,\mathrm{km})^2$. The crosses show results from a simulation with the PC scheme for a grid length of $32\,\mathrm{km}$, an imposed cooling of $\dot{T}_0 = 8\,\mathrm{Kday}^{-1}$ and the maximal-averaging strategy described in the main text. Also shown are the theoretical predictions (solid lines) given by equations 1 and 2, evaluated using values for $\langle m \rangle$ and $\langle M \rangle$ computed directly from the model data. The bin width for m, and all similar plots in this paper, is $\langle m \rangle / 4$ (i.e. dependent on the average value for that particular plot) and the bin width for M, and all similar plots in this paper, is such that there are 51 equally spaced bins from the minimum to the maximum value of M inclusive.

agreement also holds good for other heights and over other sizes of the horizontal area.

The agreement with theory for the maximal-averaging strategy is a valuable result which demonstrates that the PC scheme is providing a correct implementation of its underpinning theory when embedded within a full three-dimensional atmospheric model. We now examine whether this remains the case when the degree of averaging is reduced, thereby testing whether the scheme will be capable of accomodating variations in the large-scale forcing. Experiments have been performed for various combinations of temporal and spatial averaging. In each case the temporal

Table I. List of experiments performed in order to test averaging strategies for the input to the closure calculations of the PC scheme. In each case the model grid length was $32\,\mathrm{km}$ and the cooling rate was $\dot{T}_0 = 8\,\mathrm{Kday}^{-1}$. The first two columns show the length of the side of the averaging square, and the averaging time period respectively. The third column gives the correlation coefficient between the experimental PDF of M over an area of $(64\,\mathrm{km})^2$ at $1.52\,\mathrm{km}$ and the theoretically-expected PDF. The fourth column gives the normalized variance of total mass flux, $\langle (M - \langle M \rangle)^2 \rangle / (\langle M \rangle \langle m \rangle)$, calculated from the simulation results at the same height.

Space avg. (km)	Time avg. (min)	Correlation coefficient	Normalized variance
32	1	0.866	2.58
96	50	0.895	2.18
160	1	0.942	2.05
160	20	0.927	2.07
160	50	0.940	2.02
160	120	0.948	2.09
160	200	0.931	2.04
480	63	0.982	2.08

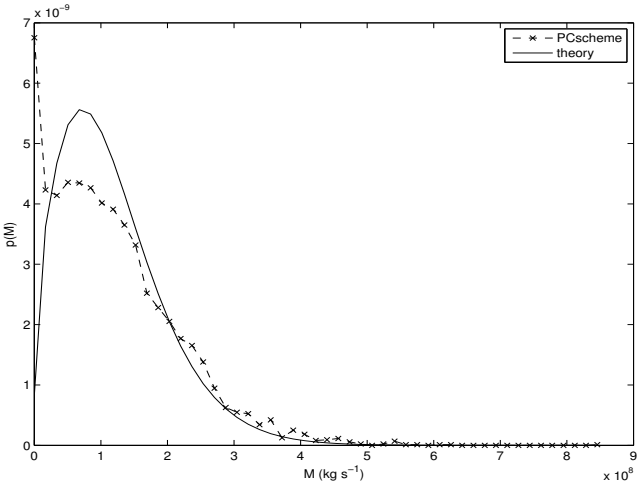

Figure 7. PDF at a height of $1.52\,\mathrm{km}$ for the total mass flux M over the horizontal area $(64\,\mathrm{km})^2$. The crosses show results from a simulation with the PC scheme for a grid length of $32\,\mathrm{km}$, a cooling of $\dot{T}_0 = 8\,\mathrm{Kday}^{-1}$ and no averaging in the input to the scheme. Also shown is the theoretical prediction (solid line) given by equation 2, evaluated using values for $\langle m \rangle$ and $\langle M \rangle$ computed directly from the model data.

averaging combines the current state of the atmosphere with its state at previous timesteps, while the spatial-averaging domain is composed of a square with the grid point of interest in the centre and a whole number of grid boxes along each side.

A summary of the experiments performed is shown in table I, alongside two measures of the fit to theory obtained for the PDF of M at $1.52\,\mathrm{km}$. The third column of the table lists the correlation coefficient between the model data for each bin and the theoretical value obtained from equation 2. The fourth column shows the normalized variance of M, which should be equal to 2 according to equation 3, when the PC scheme works correctly according to its underlying theory. It is clear from both the measures in the table that increased spatial averaging improves the agreement with theory. The time averaging is less important, and does not produce clear systematic differences in the two measures.

Note that the last line of table I corresponds to the maximal averaging strategy discussed earlier, while the first line corresponds to a case of no averaging, with the parameterization relying only on local, instantaneous input as in a traditional approach. As shown explicitly in figure 7, there are clear departures from theory if no averaging is performed. This arises because (by design for a stochastic scheme, but also in practice for a determinstic mass flux scheme), there are local fluctuations in the profiles. If local profiles are used as a basis for closure calculations then those calculations will be distorted by the fluctuations thereby introducing an artificial source of variance. We reiterate that the theoretical basis of a quasi-equilibrium closure requires that the closure depends on the large-scale environment for which the grid-scale state is not necessarily a good approximation.

Based on these results, an averaging domain of side $160\,\mathrm{km}$ and an averaging time of $50\,\mathrm{min}$ were chosen to define the "standard" averaging strategy for use in the remainder of this paper. Figure 8 shows the agreement with theory for this standard averaging, which is clearly improved over the case of no averaging (figure 7), and not too much worse than that for maximal averaging (figure 6). Thus the standard averaging is judged to be sufficient to obtain good agreement with the theory, but it nonetheless would allow for variations in the large-scale forcing to be captured, respecting the arguments based on observed correlation scales in section 2.3.

3.4. Implementation of averaging strategy for other test cases

The standard-averaging strategy developed in the previous subsection was also adapted for use in experiments with

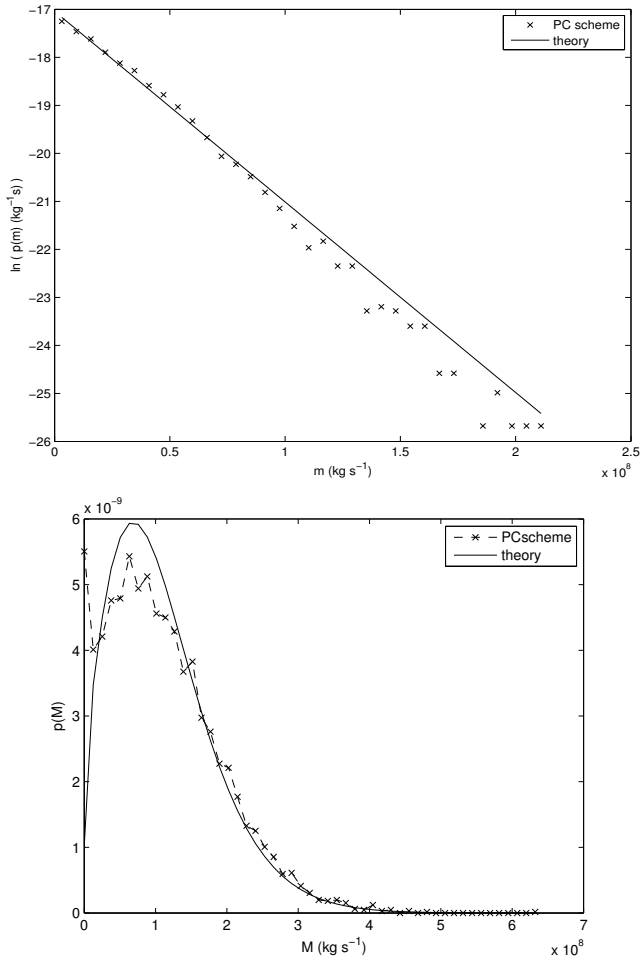

Figure 8. As figure 6, but for the standard-averaging strategy described in the main text.

different strengths of forcing and with different grid lengths. Specifically experiments have been performed using $\dot{T}_0 = 10\,\mathrm{Kday}^{-1}$ and $\dot{T}_0 = 12\,\mathrm{Kday}^{-1}$, as well as with the original amplitude of cooling but with grid lengths of $16\,\mathrm{km}$ and $51.2\,\mathrm{km}$.

The adpatations of the averaging strategy for these experiments are detailed in table II. The adaptations for different grid spacings (bottom two rows) are entirely for technical reasons, since the spatial averaging is performed over a square with a whole number of grid boxes on each side. The time averages for different grid spacings vary somewhat because different timesteps were used for different grid spacings. The adaptations for different forcings (top two rows), on the other hand, arise because, for stronger forcing, the number of clouds produced increases proportionally (given that the mass flux per cloud $\langle m \rangle$ remains constant). It is therefore possible to reduce the

Table II. Averaging strategies used for the experiments described in section 3.4. The second column shows the length of the side of the averaging square, and the third column shows the averaging time period.

Experiment	Space average (km)	Time average (min)
$\dot{T}_0 = 10\,\mathrm{Kday}^{-1}$	160	40
$\dot{T}_0 = 12\,\mathrm{Kday}^{-1}$	160	33
Grid spacing $16\,\mathrm{km}$	144	67
Grid spacing $51.2\,\mathrm{km}$	153.6	51

averaging area and/or period, and still average over an equivalent number of clouds. In practice, the temporal averaging was reduced proportionally to account for this point, the decision being dictated by the fact that the averaging square could only be reduced in size by rather large discrete steps.

The resulting PDFs from the experiment with a cooling rate of $\dot{T}_0 = 12\,\mathrm{Kday}^{-1}$ are shown in figure 9. There is good agreement with the theoretical curves, as was also found in the experiment with a cooling rate of $\dot{T}_0 = 10\,\mathrm{Kday}^{-1}$ (results not shown). Thus, the PC scheme is able to reproduce the correct mass-flux variability for different forcing conditions.

Results with grid lengths of $16\,\mathrm{km}$ and $51.2\,\mathrm{km}$ are shown in figures 10 and 11 respectively. The PDFs for M in these figures are for areas of $(64\,\mathrm{km})^2$ (a partition into 8×8 sub-domains containing 4×4 grid boxes) and $(102.4\,\mathrm{km})^2$ (a partition into 5×5 sub-domains containing 2×2 grid boxes) respectively. The agreement with theory is again good, demonstrating that the PC scheme adapts correctly to different resolutions.

3.5. *Results at other heights*

Mass flux statistics from the PC scheme have so far been presented for a height of $1.52\,\mathrm{km}$. The underpinning theory is not dependent on the height in question, so it is important to investigate whether the scheme is providing a correct implementation of the theory for other heights. In this subsection we investigate whether or not m follows an exponential distribution at higher levels and how well equations 2 and 3 describe the distribution of M. Results

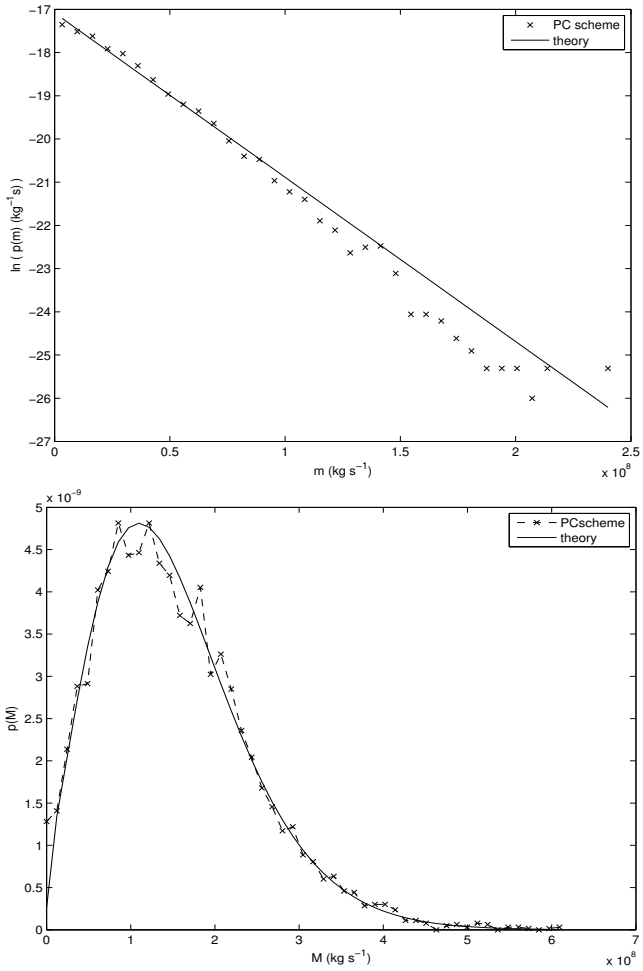

Figure 9. As figure 6, but for a cooling rate of $\dot{T}_0 = 12\,\mathrm{Kday}^{-1}$, and using the averaging strategy as specified in table II.

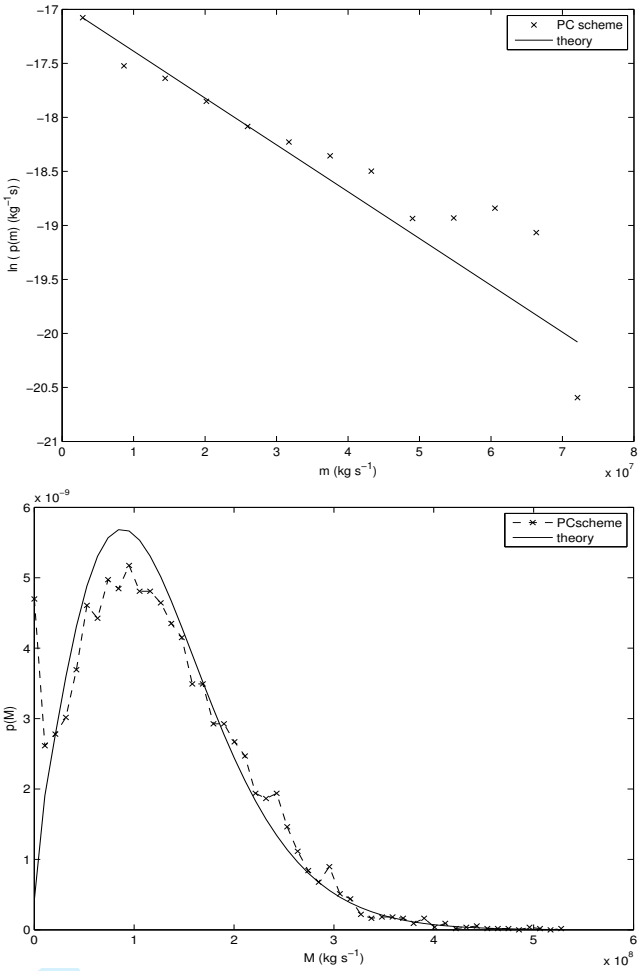

Figure 10. As figure 6, but for a model grid length of $16\,\mathrm{km}$, and using the averaging strategy as specified in table II.

are shown for the experiment with $32\,\mathrm{km}$ grid length, $\dot{T}_0 = 8\,\mathrm{Kday}^{-1}$ cooling rate and the standard-averaging strategy of Section 3.3.

Figure 12 shows PDFs of m at 3.51 and $8.75\,\mathrm{km}$. In both cases these agree with the exponential shape predicted by theory. A more detailed assessment of this agreement is found in figure 13. This shows the correlation coefficient between the probabilities taken from the model data and the probabilities taken from equation 1 for the same mean mass flux, $\langle m \rangle$. The agreement is slightly worse at the highest levels: this is likely to be due to the fact that there are fewer surviving clouds contributing to the experimental statistics, and that clouds with different strengths have different likelihoods of surviving to the top of the troposphere. It is also slightly worse at around $7\,\mathrm{km}$, where there are anomalously very slightly fewer strong clouds than in the corresponding exponential distribution. However, the correlation coefficient nonetheless remains above 0.9 at all heights.

The ability of the underlying theory to describe the distribution of M at different heights is shown in figure 14. This displays the correlation coefficient between probabilities predicted by equation 2 and normalized number frequencies taken from the data: the coefficients are computed in the same way as those discussed in Section 3.3. Results were obtained using four different-sized groups of grid boxes: for example, the results for an area of $(128\,\mathrm{km})^2$ were obtained by dividing the domain into 4×4 groups, each of 4×4 grid boxes.

The graph in figure 14 displays some interesting non-linearities in the coefficient with both height and averaging area. On inspection of individual PDF comparisons (not shown), it is apparent that, away from the LCL, the Cohen-Craig theory is less well adhered to when there is an

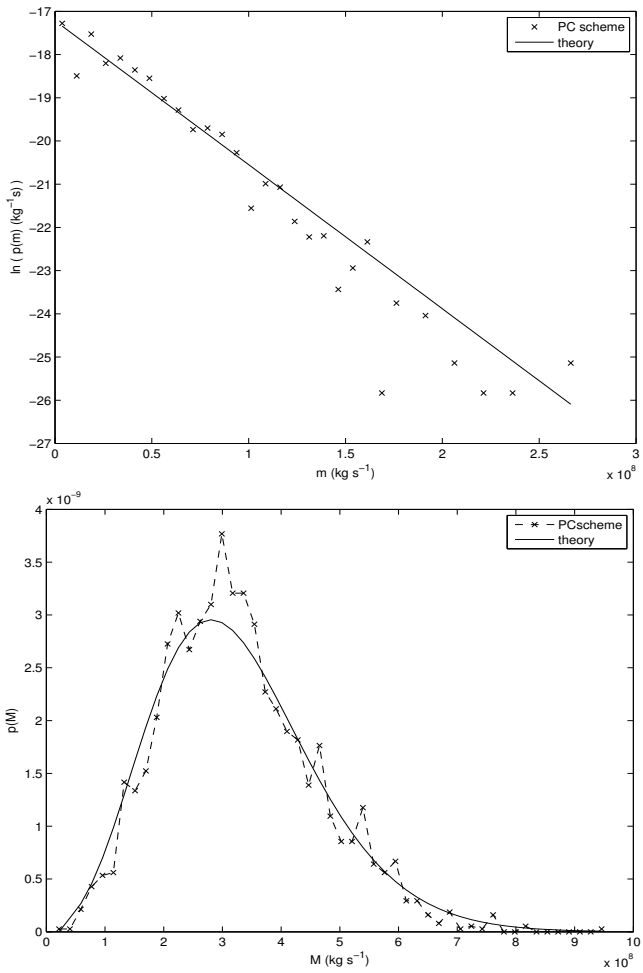

Figure 11. As figure 6, but for a model grid length of $51.2\,\mathrm{km}$, and using the averaging strategy as specified in table II. The PDF of M in the bottom panel is over a horizontal area of $(102.4\,\mathrm{km})^2$.

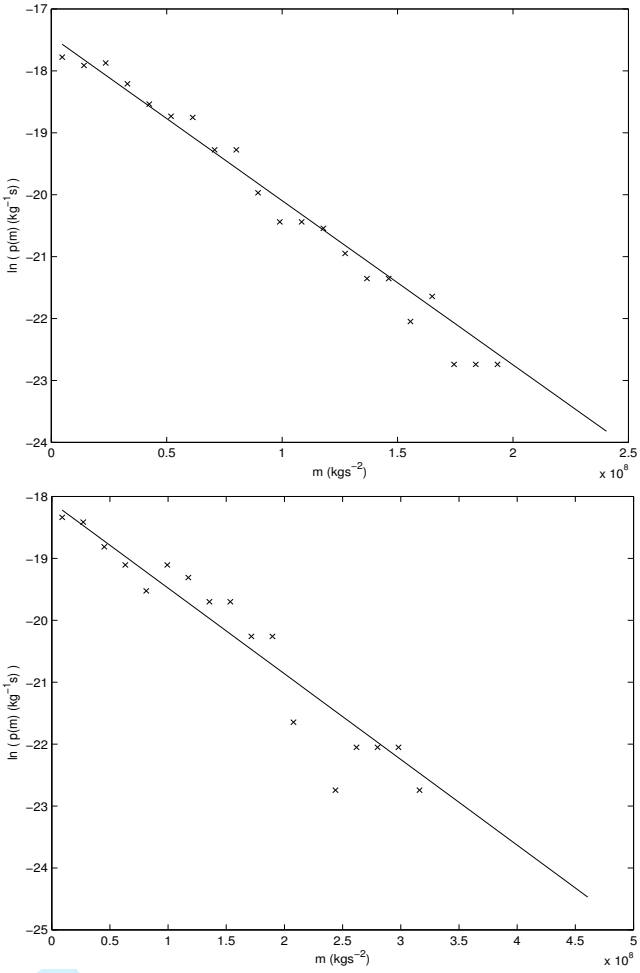

Figure 12. PDFs of m at heights of $3.51\,\mathrm{km}$ (top) and $8.75\,\mathrm{km}$ (bottom). Clouds that were produced by the parameterization, but that did not reach the height in question, are not included.

intermediate number of areas with zero mass flux – in this case instances of zero M are overestimated and instances of low M are underestimated. For the smallest area size, $32\,\mathrm{km}$, the PDF is dominated by instances of zero M at cloud base, and this does not change higher up as clouds disappear. However, looking at the $64\,\mathrm{km}$ areas, the theory is well adhered to near to cloud base, but the clouds seem to disappear in "clusters", so that there are too many instances where all the clouds in an area have disappeared (and too many where no clouds have disappeared) relative to what would be expected if they were to disappear uniformly. Higher up still, the uniformity is restored, and the PDF is now dominated by instances of zero M, similar to with the $32\,\mathrm{km}$ area at all heights. For the $128\,\mathrm{km}$ area, a similar phenomenon occurs: lower down, there are no instances at all of zero M (as the area is sufficiently large always

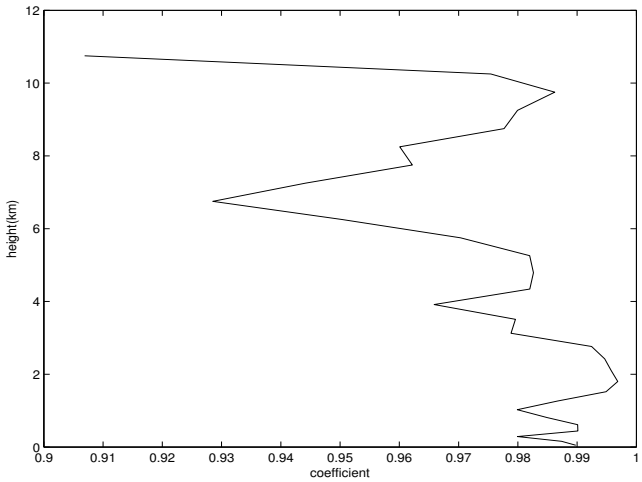

Figure 13. Correlation coefficient between theoretical and experimental PDFs of m, as a function of height.

to capture at least one cloud), but as the clouds disappear with height, the distribution crosses into the intermediate range which is less well captured by the theory. Finally, the relatively low values when dividing the domain into four

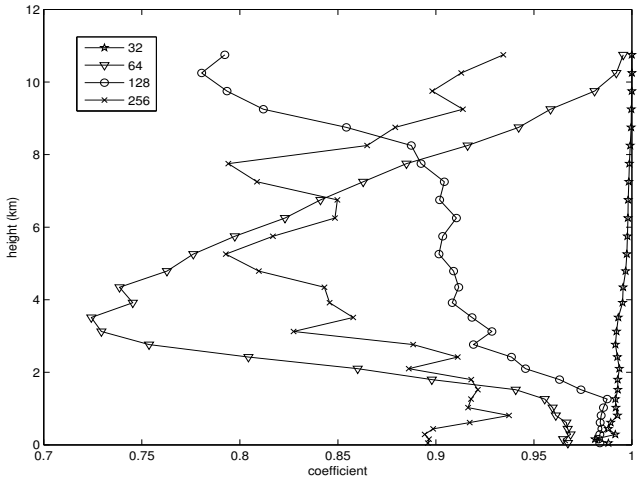

Figure 14. Correlation coefficients between theoretical and experimental PDFs of M, as a function of height. The PDFs were constructed for horizontal areas of $(32\,\mathrm{km})^2$ (stars), $(64\,\mathrm{km})^2$ (triangles), $(128\,\mathrm{km})^2$ (circles) and $(256\,\mathrm{km})^2$ (crosses).

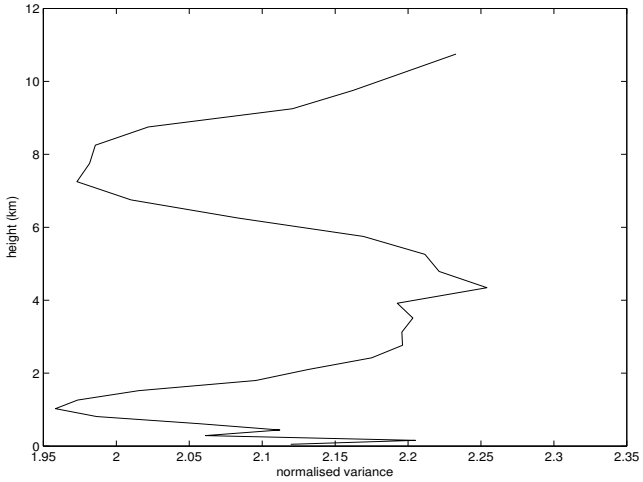

Figure 15. Variation of normalized variance $\langle (M - \langle M \rangle)^2 \rangle / (\langle M \rangle \langle m \rangle)$ with height, for M determined from individual grid-box values.

$(256\,\mathrm{km})^2$ areas are at least partly attritutable to the fact that the quantity of experimental data is relatively low.

The normalized variance, $\langle (M - \langle M \rangle)^2 \rangle / (\langle M \rangle \langle m \rangle)$, is plotted as a function of height in figure 15, for values of M determined at the level of the individual grid box. This further demonstrates the agreement with the underlying theory, to a similar extent as the agreement obtained from CRM data (Cohen and Craig 2006). Thus, although the PC scheme imposes the theoretical distribution at the LCL, the underlying theory remains appropriate to describe the distributions that are produced at all heights.

4. Rainfall statistics

Bulk mass flux schemes such as Gregory and Rowntree (1990); Kain (2004) consider a single, effective cloud rather than a spectrum of clouds carrying different mass fluxes (Plant 2010). Such schemes therefore do not yield explicit results for m. In order to compare the convective fluctuations produced by different convection schemes, it is therefore more convenient to consider fluctuations in rainfall rather than mass flux. Following Shutts and Palmer (2007), the Craig and Cohen (2006) theory can be generalized for convective rainfall, since the same important equilibrium constraints apply to this variable as to the mass flux. Specifically, in the absence of any large-scale contribution to the rainfall, then at equilibrium the ensemble-mean total convective rainfall must balance the (evaporative) moisture source. Moreover, the total rainfall is comprised of contributions from a variable, but finite number of approximately independent rain sources. Thus, a similar distribution to equation 2 is expected to describe the statistics of the convective rainfall:

$$p(C) = \delta(C)\mathrm{e}^{-\frac{\langle C \rangle}{\langle c \rangle}} + \frac{1}{\langle c \rangle}\sqrt{\frac{\langle C \rangle}{C}}\mathrm{e}^{-\frac{C+\langle C \rangle}{\langle c \rangle}} I_1\left(\frac{2}{\langle c \rangle}\sqrt{C\langle C \rangle}\right), \quad (9)$$

where C is the total rainfall within a given area, and c is the rainfall produced from each contributing source.

Rainfall data produced by each of the Plant-Craig (PC), Kain-Fritsch (KF) and Gregory-Rowntree (GR) schemes were used to construct frequency distributions for the total rainfall over various horizontal areas. Rainfall values were recorded every 8 hours for a total of 20 days, for each scheme. Results from the PC scheme are presented in figure 16, in which they are also compared with Eq. 9. The focus here is on the scaling of the distribution with area. Thus a suitable value to choose for $\langle c \rangle$ in plotting the theoretical curves is that obtained by fitting the experimental data obtained at $(256\,\mathrm{km})^2$ to equation 9 (separately for each scheme, with $\langle C \rangle$ being prescribed

simply from the mean of the experimental data). The resulting value for $\langle c \rangle$ was then used to produce the theoretical curves for the other three areas.

The agreement with theory for the PC scheme is generally good, a highly non-trivial result for this change of fluctuating variable. The exception is at the grid-scale itself, where there are too many instances of light rain. Given that the scheme imposes the theoretical distribution of cloud-base mass flux at the grid scale and given that the rainfall scales appropriately with area over larger areas, it becomes tempting to hypothesize that this may point to a problem with the implicit mass-flux-rainfall relationship that is produced by the plume model being used. However, any modifcations to the plume model which might improve the situation are beyond the scope of the present study.

Table III compares the results from the PC scheme with those of the KF and GR schemes. The correlation coefficient, used in previous sections, was found not to discriminate sufficiently the level of agreement between equation 9 and the experimental data, particularly with the smaller areas, over which instances of zero rainfall are an important aspect of the PDF. Therefore, another measure has been adopted for assessing the agreement between theory and experiment. It is obtained simply by calculating the total area between the two PDFs,

$$A = \sum_i |p(C_i)\mathrm{d}C - \nu_i / N_{\mathrm{data}}| \qquad (10)$$

where i labels the bin, ν_i is the number of data points in bin i and N_{data} is the total number of data points. This measure should not, in principle, require any normalization because the total area under the PDFs is always unity. However, care must be taken that the sampling of the data (in particular the number of bins and the upper and lower limits) is done in the same way for each scheme for a fair comparison to be made. Here, the number of bins was set to be equal to the number of data points divided by 10, subject to a minimum of 100 and a maximum of 500. The lower limit was always

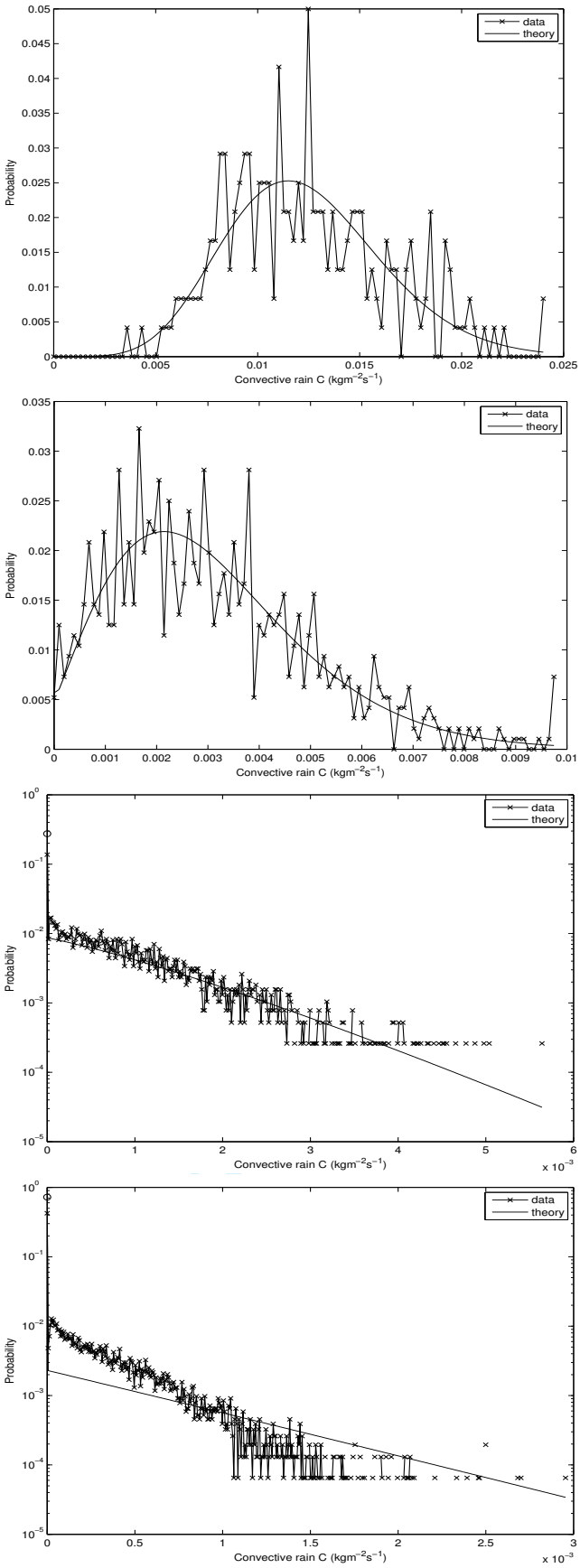

Figure 16. PDF for the total convective rainfall C over horizontal areas of $(256\,\mathrm{km})^2$ (top), $(128\,\mathrm{km})^2$ (second), $(64\,\mathrm{km})^2$ (third) and $(32\,\mathrm{km})^2$ (bottom). The crosses show results from a simulation with the PC scheme for a grid length of $32\,\mathrm{km}$, a cooling rate of $\dot{T}_0 = 8\,\mathrm{Kday}^{-1}$ and standard-averaging for the input to the scheme. Also shown are the theoretical predictions (solid line) given by equation 9. For the two lower scales, the y axis is logarithmic, and the circle denotes the value of $p(0)\mathrm{d}C$.

Table III. Measure A of difference between theoretical and experimental rainfall distributions using three different convection schemes, and computed for four different areas. The top row lists the length (in km) of the side of the accumulation area. In each case the model grid length was 32 km and the cooling rate was $\dot{T}_0 = 8\,\mathrm{K\,day^{-1}}$.

Scheme	$\langle c \rangle$ ($\mathrm{kgm^{-2}s^{-1}}$)	256	128	64	32
PC	6.0×10^{-4}	0.369	0.258	0.397	0.684
KF	5.0×10^{-4}	0.402	0.283	0.658	1.250
GR	2.8×10^{-3}	0.443	0.585	1.259	1.197

set at zero and the upper limit such that 99.5 % of the area under the theoretical PDF was included within the limits.

The PC scheme does the best job of rescaling the theoretical rainfall distribution across the different areas (a lower value of A denotes a better agreement between theory and experiment). Perhaps unsurprisingly, all three schemes perform well at the largest scale, at which $\langle c \rangle$ has been fit and for which a deterministic, 'single effective cloud' method should be appropriate. At the grid scale, all three of the schemes produce too much light rain, and not enough instances of no rain, such behaviour being in fact much less pronounced for the PC scheme than for the others.

Figure 17 shows the PDFs from the KF scheme over $(32\,\mathrm{km})^2$ and $(64\,\mathrm{km})^2$. Interestingly, the issue of too much light rain at the grid scale has become an issue of insufficient cases of very light rain, and too much light to moderate rain at $(64\,\mathrm{km})^2$. We suggest that this provides an indication of a problem with the upscaling of rainfall variability: the rainfall field becomes too smooth too readily.

It must be noted here that alternative methods are possible for setting the values of $\langle c \rangle$ used for each scheme. We have investigated some other approaches, and do find some effects on the values of A obtained. However, any uncertainties in the choice of $\langle c \rangle$ have no effect on our general conclusions about the upscaling of rainfall variability, which is much improved with the PC scheme compared to the conventional deterministic schemes. It is also worth noting that the spatial and temporal averaging strategy does not strongly affect values of A obtained for the PC scheme.

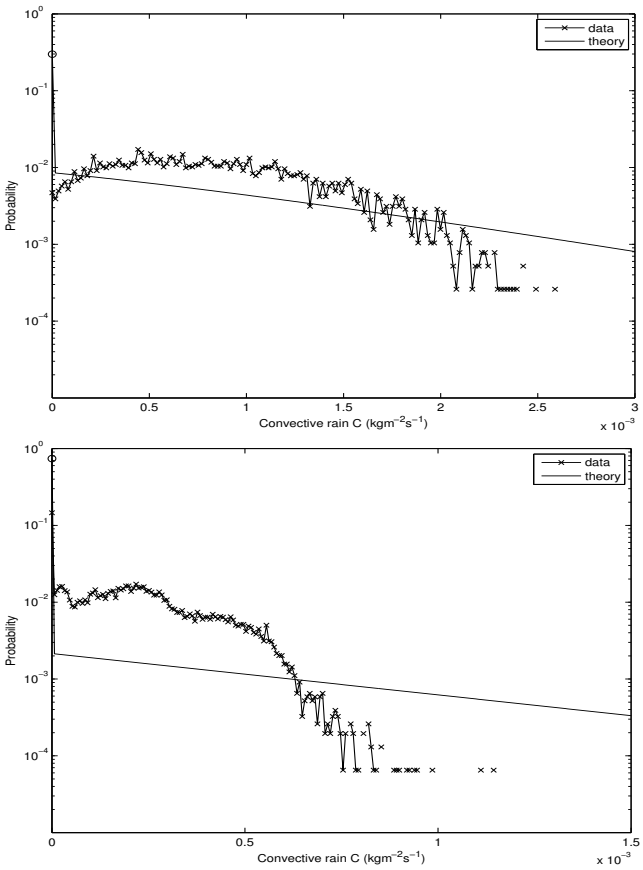

Figure 17. PDF for the total convective rainfall C over horizontal areas of $(64\,\mathrm{km})^2$ (top) and $(32\,\mathrm{km})^2$ (bottom). The crosses show results from a simulation with the KF scheme for a grid length of 32 km and a forcing of $\dot{T}_0 = 8\,\mathrm{K\,day^{-1}}$. Also shown are the theoretical predictions (solid line) given by equation 9. The circle in each plot denotes the value of $p(0)\mathrm{d}C$.

5. Discussion

The mass flux formalism for deep convective parameterization is based on representing the collective behaviour of an ensemble of convective clouds, subject to a known large-scale forcing. A conventional deterministic parameterization assumes that the grid-scale of the parent model can be identified with the large scale, whereas the Plant-Craig (PC) scheme makes an explicit distinction between those scales. Natural consequences of that distinction are that convective parameterization should be stochastic, and that the input large-scale state used for closure calculations should be a space-time average over a suitable region. The implications of that first consequence were addressed by Plant and Craig (2008), while implications of the second have been considered here by developing the PC scheme for use in three dimensions. The resulting scheme has been shown to produce mean vertical profiles which agree

well with established conventional schemes, and with CRM simulations. The statistics of the PC scheme also agree well with the underlying Cohen-Craig theory (Craig and Cohen 2006), to a similar extent as their corresponding CRM simulations (Cohen and Craig 2006). In this way, the PC scheme has been shown to produce the same convective statistics as CRM simulations, a key test for any stochastic scheme. The agreement between PC and theory (and, therefore, CRM) is good for a range of model heights, suggesting that the theory remains applicable away from the cloud base. The agreement is also good for a range of model grid lengths, indicating that the PC scheme can adapt correctly within variable resolution models.

We have investigated what averaging length and time scales are required to produce an input state with which the scheme can yield the correct statistics. An averaging length of $160\,\mathrm{km}$ and time of $50\,\mathrm{minutes}$ was found to be sufficient for the $\dot{T}_0 = 8\,\mathrm{K day}^{-1}$ experiments. Given that these experiments typically included 307 clouds, on average, over the $512\,\mathrm{km}$ square domain, and with a cloud lifetime of $45\,\mathrm{minutes}$, this corresponds to a requirement of roughly 33 clouds (full lifetime equivalent) in the being averaged.

This requirement can be used to produce a dynamical averaging strategy for operational use of the PC scheme. Of course, the averaging length and time scales vary depending on weather regime, and so the fundamental requirement, at each timestep and in each grid box, would be to ensure that a large enough area is taken to include 33 clouds (full lifetime equivalent), based on an estimate of the number of clouds in the grid box in question (obtained, for example, by taking the averaged value from the previous timestep).

However, although the spatial averaging is clearly beneficial, there is certainly scope for using less averaging without greatly degrading the statistics. Indeed, a full range of averaging strategies is available from, at one end, the ideal target of 33 clouds (full lifetime equivalent) to, at the other end, the possibility of foregoing the averaging completely if this is necessary due to operational constraints. Although the temporal averaging yields no significant benefit in this paper, it is less computationally intensive than the spatial averaging, and does not require communication between processors in a parallel environment. Given that it does not degrade the performance of the scheme in the current simulations, and that it may yield benefit in other situations, it is suggested here that temporal averaging may still be used as an alternative, or a complement, to spatial averaging if the aforementioned constraints render the spatial averaging unwieldy.

The scheme has been shown to yield the correct scaling for the variability of rainfall rate, across a range of scales. Even for such a simple model setup, the established conventional schemes do not do this correctly. The improvement of the scaling produced by the PC scheme, over that produced by conventional schemes, indicates that the stochastic character of the PC scheme captures physically-realistic convective variability that will provide a worthwhile improvement to NWP and GCMs.

Work is underway on investigating the impact of the PC scheme on mesoscale NWP forecasts. Groenemeijer and Craig (2011) have shown that the scheme yields significant amounts of variability, as compared to those amounts yielded by perturbations in the initial and boundary conditions, and have investigated how this partition between "internal" and "external" variability changes for different weather regimes, within an ensemble forecast. Additionally, we are currently conducting a study into the impact on ensemble verification scores, when the conventional convection scheme in the UM is replaced by the PC scheme.

Acknowledgements

This research was supported by the NERC grant NE/D011493/1. Computing support was provided by the NCAS Computer Modelling Service. The authors would also like to thank the following for useful discussions during the course of the project: W. J. Tennant, N. E. Bowler, R. N. Swinbank, S. E. Beare and other staff at the Met Office;

and P. H. Groenemeijer and G. C. Craig from DLR/LMU, Munich.

References

Ball MA, Plant RS. 2008. Comparison of stochastic parameterization approaches in a single-column model. *Phil. Trans. Roy. Soc. A* **366**: 2605–2623.

Bechtold P, Köhler M, Jung T, Doblas-Reyes F, Leutbecher M, Rodwell MJ, Vitart F, Balsamo G. 2008. Advances in simulating atmospheric variabilty with the ECMWF model: From synoptic to decadal time-scales. *Q. J. R. Meteorol. Soc.* **134**: 1337–1351.

Bowler NE, Arribas A, Mylne KR, Robertson KB, Beare SE. 2008. The MOGREPS short-range ensemble prediction system. *Q. J. R. Meteorol. Soc.* **134**: 703–722.

Bright DR, Mullen SL. 2002. Short-range ensemble forecasts of precipitation during the southwest monsoon. *Weather and Forecasting* **17**: 1080–1100.

Buizza R. 1997. Potential forecast skill of ensemble prediction and spread and skill distributions of the ECMWF ensemble prediction system. *Monthly Weather Rev.* **125**: 99–119.

Buizza R, Houtekamer PL, Toth Z, Pellerin G, Wei M, Zhu Y. 2005. A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems. *Monthly Weather Rev.* **133**: 1076–1097.

Buizza R, Miller M, Palmer TN. 1999. Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System. *Q. J. R. Meteorol. Soc.* **125**: 2887–2908.

Cohen BG. 2001. Fluctuations in an ensemble of cumulus clouds. PhD thesis, University of Reading.

Cohen BG, Craig GC. 2006. Fluctuations in an equilibrium convective ensemble. Part II: Numerical experiments. *J. Atmos. Sci.* **63**(8): 2005–2015.

Craig GC, Cohen BG. 2006. Fluctuations in an equilibrium convective ensemble. Part I: Theoretical formulation. *J. Atmos. Sci.* **63**(8): 1996–2004.

Davies L. 2008. Self-organization of convection as a mechanism for memory. PhD thesis, University of Reading.

Davies L, Jakob C. 2011. On convective cloud properties at radiative-convective equilibrium. *To be submitted to: J. Geophys. Res.* .

Davies T, Cullen MJP, Malcolm AJ, Mawson MH, Staniforth A, White AA, Wood N. 2005. A new dynamical core for the Met Office's global and regional modelling of the atmosphere. *Q. J. R. Meteorol. Soc.* **131**: 1759–1782.

Davoudi J, McFarlane NA, Birner T. 2010. Fluctuation of mass flux in a cloud-resolving simulation with interactive radiation. *J. Atmos. Sci.* **67**: 400–418.

Emanuel KA. 2000. Quasi-equilibrium thinking. In: *General circulation model development*, Randall DA (ed), ch. 8, Elsevier, p. 225255.

Gregory D, Rowntree PR. 1990. A mass flux convection scheme with representation of cloud ensemble characteristics and stability-dependent closure. *Monthly Weather Rev.* **118**: 1483–1506.

Groenemeijer PH, Craig GC. 2011. Ensemble forecasting with a stochastic convective parametrization based on equilibrium statistics. *Submitted to Atmos. Chem. Phys.* .

Held IM, Zhao M. 2008. Horizontally homogeneous rotating radiative-convective equilibria at GCM resolution. *J. Atmos. Sci.* **65**: 2003–2013.

Held IM, Zhao M, Wyman B. 2007. Dynamic radiative-convective equilibria using GCM column physics. *J. Atmos. Sci.* **64**: 228–238.

Holloway CE, Neelin JD. 2010. Temporal relations of column water vapor and tropical precipitation. *J. Atmos. Sci.* **67**: 1091–1105.

Jones TR, Randall DA. 2011. Quantifying the limits of convective parameterizations. *J. Geophys. Res.* **116**: D08 210.

Kain JS. 2004. The Kain-Fritsch convective parameterization: An update. *J. Appl. Meteor.* **43**: 170–181.

Kain JS, Fritsch JM. 1990. A one-dimensional entraining / detraining plume model and its application in convective parameterization. *J. Atmos. Sci.* **47**(23): 2784–2802.

Kuo YH, Bresch JF, Cheng MD, Kain J, Parsons DB, Tao WK, Zhang DL. 1997. Summary of a mini workshop on cumulus parameterization for mesoscale models. *Bull. Am. Meteorol. Soc.* **78**: 475–491.

Lander J, Hoskins BJ. 1997. Believable scales and parameterizations in a spectral transform model. *Monthly Weather Rev.* **125**: 292–303.

Larson K, Hartmann DL. 2003. Interactions among cloud, water vapor, radiation and large-scale circulation in the tropical climate. Part I: Sensitivity to uniform sea surface temperature changes. *J. Clim.* **16**: 1425–1440.

LeMone MA, Zipser EJ. 1980. Cumulonimbus vertical velocity events in GATE. Part II: Synthesis and model core structure. *J. Atmos. Sci.* **37**: 2458–2469.

Lin JWB, Neelin JD. 2003. Toward stochastic deep convective parameterization in general circulation models. *Geophys. Res. Lett.* **30**(4): 1162.

Lock AP, Brown AR, Bush MR, Martin GM, Smith RNB. 2000. A new boundary layer mixing scheme. Part I: Scheme description and single-column model tests. *Monthly Weather Rev.* **128**: 3187–3199.

Lucas C, Zipser EJ, LeMone MA. 1994. Vertical velocity in oceanic convection off tropical Australia. *J. Atmos. Sci.* **51**: 3183–3193.

Martin GM, Ringer MA, Pope VD, Jones A, Dearden C, Hinton TJ. 2006. The physical properties of the atmosphere in the new Hadley centre global environmental model (HadGEM1). Part I: Model description and global climatology. *J. Clim.* **19**: 1274–1301.

May PT, Rajopadhyaya DK. 1999. Vertical velocity characteristics of deep convection over darwin, Australia. *Monthly Weather Rev.* **127**: 1056–1071.

Moron V, Robertson AW, Ward MN, Camberlin P. 2007. Spatial coherence of tropical rainfall at the regional scale. *J. Clim.* **20**: 5244–5263.

Parodi A, Emanuel K. 2009. A theory for buoyancy and velocity scales in deep moist convection. *J. Atmos. Sci.* **66**: 3449–3463.

Paulius O, Garner S. 2006. Sensitivity of radiative-convective equilibrium simulations to horizontal resolution. *J. Atmos. Sci.* **63**: 1910–1923.

Plant RS. 2010. A review of the theoretical basis for bulk mass flux convective parameterization. *Atmos. Chem. Phys.* **10**: 3529–3544.

Plant RS, Craig GC. 2008. A stochastic parameterization for deep convection based on equilibrium statistics. *J. Atmos. Sci.* **65**(1): 87–105.

Ricciardulli L, Sardeshmukh PD. 2002. Local time- and space scales of organized tropical deep convection. *J. Clim.* **15**: 2775–2790.

Robe FR, Emanuel KA. 1996. Moist convective scaling: Some inferences from three-dimensional cloud ensemble simulations. *J. Atmos. Sci.* **53**(22): 3265–3275.

Shutts G. 2005. A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Q. J. R. Meteorol. Soc.* **131**: 3079–3102.

Shutts GJ, Gray MEB. 1999. Numerical simulations of convective equilibrium under prescribed forcing. *Q. J. R. Meteorol. Soc.* **125**: 2767–2787.

Shutts GJ, Palmer TN. 2007. Convective forcing fluctuations in a cloud-resolving model: Relevance to the stochastic parameterization problem. *J. Clim.* **20**: 187–202.

Smith DF, Gasiewski AJ, Jackson DL, Wick GA. 2005. Spatial scales of tropical precipitation inferred from TRMM microwave imager data. *IEEE Trans. Geophys. Remote Sens.* **43**: 1542–1551.

Stiller O. 2009. Efficient moist physics schemes for data assimilation. II: Deep convection. *Q. J. R. Meteorol. Soc.* **135**: 721–738.

Willett MR, Milton SF. 2006. The tropical behaviour of the convective parameterization in aquaplanet simulations and the sensitivity to timestep. Forecasting Research Technical Report 482, Met Office, UK.

Williams PD, Palmer TN (eds). 2009. *Stochastic physics and climate modelling*. Cambridge University Press. 496pp.

Wilson DR, Ballard SP. 1999. A microphysically based precipitation scheme for the UK Meteorological Office Unified Model. *Q. J. R. Meteorol. Soc.* **125**: 1607–1636.

Xu KM, Arakawa A, Krueger SK. 1992. The macroscopic behavior of cumulus ensembles simulated by a cumulus ensemble model. *J. Atmos. Sci.* **49**(24): 2402–2420.

Zhang F, Snyder C, Rotunno R. 2003. Effects of moist convection on mesoscale predictability. *J. Atmos. Sci.* **60**: 1173–1185.

Zhang GJ, McFarlane NA. 1995. Sensitivity of climate simulations to the parameterization of cumulus convection in the Canadian climate centre general circulation model. *Atmosphere-Ocean* **33**: 407–446.