

A physically-based stochastic boundary-layer perturbation scheme. Part I: formulation and evaluation in a convection-permitting model

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open Access

Clark, P. ORCID: <https://orcid.org/0000-0003-1001-9226>, Halliwell, C. and Flack, D. (2021) A physically-based stochastic boundary-layer perturbation scheme. Part I: formulation and evaluation in a convection-permitting model. *Journal of the Atmospheric Sciences*, 78 (3). pp. 727-746. ISSN 1520-0469 doi: <https://doi.org/10.1175/JAS-D-19-0291.1> Available at <https://centaur.reading.ac.uk/93960/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1175/JAS-D-19-0291.1>

Publisher: American Meteorological Society

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

A Physically Based Stochastic Boundary Layer Perturbation Scheme. Part I: Formulation and Evaluation in a Convection-Permitting Model

P. A. CLARK,^a C. E. HALLIWELL,^b AND D. L. A. FLACK^{a,c}

^a *Department of Meteorology, University of Reading, Reading, United Kingdom*

^b *MetOffice@Reading, Department of Meteorology, University of Reading, United Kingdom*

(Manuscript received 17 October 2019, in final form 26 October 2020)

ABSTRACT: We present a simple, physically consistent stochastic boundary layer scheme implemented in the Met Office's Unified Model. It is expressed as temporally correlated multiplicative Poisson noise with a distribution that depends on physical scales. The distribution can be highly skewed at convection-permitting scales (horizontal grid lengths around 1 km) when temporal correlation is far more important than spatial. The scheme is evaluated using small ensemble forecasts of two case studies of severe convective storms over the United Kingdom. Perturbations are temporally correlated over an eddy-turnover time scale, and may be similar in magnitude to or larger than the mean boundary layer forcing. However, their mean is zero and hence they, in practice, they have very little impact on the energetics of the forecast, so overall domain-averaged precipitation, for example, is essentially unchanged. Differences between ensemble members grow; after around 12 h they appear to be roughly saturated; this represents the time scale to achieve a balance between addition of new perturbations, perturbation growth, and dissipation, not just saturation of initial perturbations. The scheme takes into account the area chosen to average over, and results are insensitive to this area at least where this remains within an order of magnitude of the grid scale.

KEYWORDS: Boundary layer; Ensembles; Numerical weather prediction/forecasting; Parameterization; Stochastic models; Subgrid-scale processes


1. Introduction

In recent years, there has been a rapid growth in the development and use of so-called convective-scale numerical weather prediction (NWP) systems using “convection-permitting models” (CPMs) (Clark et al. 2016). In building quantitative NWP systems, it is widely recognized that a number of aspects of the system are only known with some degree of uncertainty. Ensemble-prediction systems (hereafter ensembles) are designed to translate our understanding of uncertainty into reliable probabilistic forecasts, in particular of hazardous weather events. The uncertainty can come from many areas including model uncertainty and initial and boundary condition uncertainty, all of which have been studied at the convective scale (e.g., Leoncini et al. 2010, 2013; Keil et al. 2014; Kühnlein et al. 2014). It is essential that these uncertainties are represented accurately.

All forms of uncertainty may have an important role in a forecast as errors, particularly at the convective scale, can, and do, generally grow quickly within forecasts (e.g., Lorenz 1969; Hohenegger and Schär 2007). The consensus in convective-scale studies is that errors grow primarily as a result of the processes occurring within convection (Zhang et al. 2003; Hohenegger et al. 2006). Initial studies considered the growth from initial conditions and indicated that not only was error growth faster

than in larger-scale models that use parameterized convection (Hohenegger and Schär 2007) but that forecasts could be improved in certain situations through better specification of the initial conditions (e.g., Melhauser and Zhang 2012). As expected, the predominant impact of the error growth from initial condition uncertainty occurred at the start of the forecast (Keil et al. 2014; Kühnlein et al. 2014), though of course this uncertainty will have some impact downstream. On the other hand, due to their regular refreshing, the boundary condition uncertainty led to consistent error growth throughout the forecasts (Keil et al. 2014; Kühnlein et al. 2014). Uncertainty also arises, for many reasons, from the representation of physical processes within the model. Investigations of error growth from simulated model physics variations in CPMs suggest that they have the greatest impact when initiated within the boundary layer (Lean 2006) and during the initiation and development of convective events (Leoncini et al. 2010; Keil et al. 2014). However, it is clear that all forms of uncertainty need to be represented in the model to allow for some form of variability on all scales from the start of the run, despite the different interactions and impacts of the error growth in the different convective situations (e.g., Raynaud and Bouttier 2016).

Uncertainty in the boundary layer can, itself, arise from a number of sources. The first is “model error” due to the fact that model parameterizations are not perfect. Put more precisely, it is important to recognize that, in the case of the boundary layer, boundary layer schemes (including shallow convection) are designed to predict the ensemble mean of realizations of a turbulent boundary layer in quasi-equilibrium.

 Denotes content that is immediately available upon publication as open access.

^c Current affiliation: Met Office, Exeter, United Kingdom.

Corresponding author: P. A. Clark, p.clark@reading.ac.uk



This article is licensed under a [Creative Commons Attribution 4.0 license](http://creativecommons.org/licenses/by/4.0/) (<http://creativecommons.org/licenses/by/4.0/>).

(It must be emphasized that the ensemble here is purely conceptual and nothing to do with ensemble-based NWP.) With given boundary conditions and forcing, we assume that there is a unique ensemble-mean solution, but no scheme always reproduces this perfectly. Schemes tend to have systematic errors; in an ideal world these would be correctable via a bias-correction scheme, leaving an unbiased error, but in many cases this has only been done through a form of assessment and manual tuning. Taking account of this has led to the use of “multiphysics” or “poor-man’s” ensembles, e.g., [Ebert \(2001\)](#). While undoubtedly important, this source of error is conceptually difficult as it is not based upon physics, but rather the inadequacies of our modeling of physics. As [Ebert \(2001\)](#) shows, it is essential that the performance of each model is properly taken into account before an optimal ensemble is designed. If not approached very carefully, it can lead to the artificial enhancement of ensemble spread to meet quantitatively desirable criteria by deliberately introducing less accurate parameterizations.

Physically based uncertainty arises from the heterogeneity of the boundary layer. In CPMs we want a space–time mean over a finite domain, not the ensemble mean, but we have no direct methods to establish how to parameterize this. Even a boundary layer with horizontally homogeneous forcing results in a highly inhomogeneous instantaneous turbulent flow. We assume that the stationary system is ergodic, i.e., that sufficiently long (wide) time (space) averages tend to the ensemble mean. However, as we look at increasingly small space and time periods, our ability to predict flux divergences is limited by the inherent variability of the turbulent flow.

Heterogeneity can also be introduced by the surface-exchange process. Many surface schemes, including that in the Met Office Unified Model (MetUM), account for surface inhomogeneity using techniques based upon the concept of blending height (e.g., [Mason 1988](#)) which accounts for inhomogeneity on scales of a few hundred meters; theory tells us that the boundary layer then “sees” an effective surface. The key point is that this heterogeneity is on a smaller scale than the convective–overturning scale considered in our scheme, so as a source of variability this should be less important. Heterogeneity at the model grid scale and larger is explicit in the model (though, of course, many questions surround whether the surface forcing should be smoothed for numerical reasons). Nevertheless, there is probably a range of scales in between where surface heterogeneity, if it exists, contributes to boundary layer variability.

In addition, there is uncertainty associated with surface characteristics. This may be essentially fixed or slowly varying (e.g., exactly how tall are the trees and hence what is their roughness length, or what is their leaf area index). This may be of very practical importance but is not a problem of physics but of information that can, in principle and, increasingly, in practice, be acquired. However, some variables, notably soil moisture, may have uncertainty that is impractical to reduce by measurement.

A number of studies exist concerning the impact of boundary layer uncertainty. One technique that has been used in atmospheric models when examining the predictability of convective events lumps small-scale stochastic variations in the boundary layer due to all unresolved turbulent processes into relatively arbitrary stochastic perturbation fields in the form of Gaussian “bumps” (e.g., [Done et al. 2012](#); [Leoncini et al. 2013](#); [Flack](#)

[et al. 2018](#)). While used as the only form of uncertainty represented in the model it has not produced large spread at the kilometer scale ([Flack et al. 2018](#)). However, the representation of the processes has been useful for determining different positions and magnitudes of scattered showers and convection in general. Here we go beyond this to develop a stochastic boundary layer perturbation scheme that specifically takes into account variation due to sampling the turbulent processes. The ideas below take this as a starting point, assuming that our ensemble-mean parameterization of the system is correct and that our task is to estimate the “sampling error” or noise that arises from averaging the system over a smaller time period and area than would be required for the average to be sufficiently converged to equal the ensemble mean. We do not take account of heterogeneity of the surface or model error, though the scheme is designed to be compatible with any boundary layer scheme.

It should be noted that the objective is to gain reasonable physically based estimates of the sampling uncertainty as a function of model resolution, as it must increase as resolution decreases. At current CPM scales, the studies cited above lead us to expect that mesoscale uncertainty in initial and boundary conditions is the primary contributor to ensemble spread.

The methodology discussed in this paper is inspired by [Plant and Craig \(2008\)](#) but can be traced back to the idea of shot-noise decomposition of turbulence, that is briefly reviewed in [section 2](#). The new scheme is described in [section 3](#). We first introduce a simple bulk parameterization of the convective boundary layer (CBL) in order to introduce scaling, but also as a means to illustrate the approach in a simple framework. [Section 4](#) outlines implementation of the scheme in the MetUM. Similar schemes have been implemented by others (e.g., A. Lock 2016, personal communication; [Kober and Craig 2016](#), hereafter [KC](#)) and we compare our scheme with these in [section 5](#).

We then test the scheme. This testing evaluates the impact of the scheme on its own; it is not expected that this source of uncertainty can account for the whole of forecast error, so evaluation is restricted in Part I of this paper to impact on ensemble spread and not on overall forecast reliability. The latter is addressed in [Flack et al. \(2021, hereafter Part II\)](#). The testing (in [Part II](#)) compares forecast uncertainty due to initial and boundary condition uncertainty with the uncertainty resulting from the scheme running continuously. It relies on the existence of realistic initial and boundary condition perturbations from the Met Office Global and Regional Ensemble Prediction System (MOGREPS) scheme; this only has one boundary layer scheme in operational use. At present it is therefore not feasible to test cleanly with another boundary layer scheme (though its design would make this straightforward in a model with multiple schemes). To test the scheme, diagnostics designed to test convective-scale ensemble rainfall forecasts have been used; these are described in [section 6](#) and applied to case studies in [section 7](#) and in [Part II](#). Finally, we summarize our conclusions regarding the scheme’s characteristic behavior in [section 8](#).

2. Shot-noise decomposition

The primary objective is to represent the variability that arises on space and time scales somewhat larger than the

characteristic scales of boundary layer eddies. We do not wish to describe the detailed behavior of these eddies, but, rather, to capture the consequences of their intermittent and chaotic occurrence. Our starting point is to assume that eddies occur randomly and independently. This is a restriction that shall be discussed below. The idea of modeling a turbulent process as a superposition of random events was first proposed by Lumley (1967), and a particular methodology was described by Moin and Moser (1989). By analogy with the electronic noise that occurs due to the carrying of current by independent, random electrons, this approach is known as “shot-noise” decomposition.

We assume that eddies occur as independent, random events, that one might term eddies. Suppose we consider just temperature, the warming provided by “unit eddy” is characterized by a function $\phi(z, \mathbf{x}, t)$ with \mathbf{x} and t relative to some reference point in the eddy. Thus, ϕ includes length and time scales over which this warming is spread. Note that we have included a z dependence but this, at least when averaged horizontally, must equal that of the mean field.

Eddies occur randomly in space and time, at a set of points $\{\mathbf{x}_i, t_i\}$ and amplitudes $\{a_i\}$, where a_i is an independent random variable with a given pdf. We can think of a function F made up of delta functions with these amplitudes and locations,

$$F(\mathbf{x}, t) = \sum_{i=-\infty}^{i=\infty} a_i \delta(\mathbf{x} - \mathbf{x}_i, t - t_i). \quad (1)$$

The probability of an event occurring in the infinitesimal interval $\delta x \delta y \delta t$ is $\rho \delta x \delta y \delta t$ follows Poisson statistics. Then the field of increments applied to the model state, denoted S , can be constructed convolving this with ϕ :

$$S(z, x, y, t) = F * \phi = \sum_{i=-\infty}^{i=\infty} a_i \phi(z, x - x_i, y - y_i, t - t_i). \quad (2)$$

It is then straightforward to show that the autocorrelation function of S (in time and/or space) is just that of ϕ . An important parameter of the autocorrelation function, the integral time scale for the process, τ_ϕ , is obtained by integrating the autocorrelation function over time. Equivalent results exist for spatial autocorrelation functions and integral length scales, $\ell_{x\phi}$ and $\ell_{y\phi}$.

We assume, without proof, that successive regions in space become uncorrelated if we average over a distance longer than the integral length scale. Similarly, if we average over areas greater than the product of integral length scales in x and y , successive regions in space become uncorrelated. We can then think of F as the source of a quantity such as temperature provided by the stochastic process. We assume that each “delivers” a_i over an average time scale τ_ϕ and area $A_i = \ell_{x\phi} \ell_{y\phi}$. Under these circumstances, the variability of the space/time averaged field is solely determined by the variability in a_i and the number of events. This may be illustrated by a simple example using a bulk parameterization of the CBL. This will help motivate the final scheme but also facilitate a discussion of relevant space and time scales compared with model resolution and time steps.

3. A bulk parameterization

a. Ensemble-mean parameterization

We take as a starting point a bulk model of the CBL (e.g., Carson 1973). To illustrate our approach, we shall ignore moisture and assume that the top of the boundary layer is a rigid, impermeable lid. Including these is straightforward but adds nothing to the insights gained with our simplified model. The key parameters are thus the boundary layer depth h and the surface buoyancy flux (expressed as a heat flux), given by

$$H = \rho C_p \langle w' \theta' \rangle_0, \quad (3)$$

where ρ is the density of air, C_p is the specific heat capacity of air at constant pressure, and $\langle w' \theta' \rangle_0$ is the ensemble mean of the covariance of fluctuations of vertical velocity and potential temperature at the surface.

From these we can derive the convective velocity scale w_* , and thence the scale for temperature fluctuations, θ_* [see, e.g., Garratt (1994), though our notation differs slightly]:

$$w_* = \left(\frac{g H h}{\rho C_p \overline{\theta}} \right)^{1/3} \quad (4)$$

and, requiring $\langle w' \theta' \rangle_0 = w_* \theta_*$,

$$\theta_* = \frac{H}{\rho C_p w_*}. \quad (5)$$

Here, $\overline{\theta}$ is the vertical average of the ensemble-mean θ and g is the acceleration due to gravity. Note that the boundary layer literature often uses the opposite sign convention, but it is conceptually easier here to drop this sign convention as we are primarily considering CBLs. In a well-mixed box model with an impermeable lid the heat flux profile is

$$H(z) = H \left(1 - \frac{z}{h} \right). \quad (6)$$

The rate of change of potential temperature is given by

$$\rho C_p \frac{d\overline{\theta}}{dt} = - \frac{\partial H}{\partial z}, \quad (7)$$

which implies (ignoring the change of boundary layer depth)

$$\frac{d\overline{\theta}}{dt} = \frac{H}{\rho C_p h} = \frac{w_* \theta_*}{h} = \frac{\theta_*}{\tau_*}, \quad (8)$$

where we have introduced an “eddy-turnover time scale,” $\tau_* = h/w_*$.

As a numerical example, if $H = 400 \text{ W m}^{-2}$ and $h = 1000 \text{ m}$, $w_* \approx 2.2 \text{ m s}^{-1}$, and $\theta_* \approx 0.2 \text{ K}$. The eddy-turnover time scale τ_* is about 460 s or nearly 8 min.

b. A simple “shot-noise” version of the bulk CBL

We adopt a very similar approach to that of Plant and Craig (2008) in dealing with deep convection. We assume the following:

- 1) The mean heating is actually delivered to the boundary layer through a set of discrete thermals.
- 2) Our space–time averaging is on a scale at least the size of one thermal. (Spatial correlation will be discussed in a little more depth in section 5.)
- 3) The key source of variability is the *number* of thermals, in a given area, averaged over and we do not need to consider, in detail, the structure of a thermal.

A difference from Plant and Craig (2008) is that they also assume a distribution of thermal (i.e., cloud) properties based on statistical–mechanical arguments. We could, no doubt, extend our argument similarly, but for the present we shall concentrate on the number variability. It should be noted that this approach is the natural counterpart of the idea that the problem with the “gray zone” is that we are no longer averaging over a “large” number of clouds (as assumed by the pioneers of “mass flux”–based convection schemes; Arakawa and Schubert 1974).

Each thermal “event” delivers heating such that an area A_t is warmed (throughout the depth of the boundary layer) by θ_t by the thermal, in time scale τ_t ; the total heat delivered per thermal is thus $Q_t = \rho C_p \theta_t A_t h$. Note that A_t is the average area occupied by one thermal; it includes the upward motion (e.g., a convective thermal) and downward motion (i.e., the whole region influenced by the thermal). It makes sense to assume that $A_t \propto h^2$; the result is that we can express the relative variability in boundary layer heating entirely in terms of h and w_* or, equivalently, h and τ_* .

We assume that thermals occur entirely at random (independently) with probability p per unit area and time. The probability that a thermal occurs in an infinitesimal time dt in a given (small) area dA is $p dA dt$. The number of thermals that occur in finite time Δt and finite area ΔA is thus given by a Poisson distribution. The mean number is $\bar{n}(\Delta A, \Delta t) = p \Delta A \Delta t$, so the number of thermals (n) has the distribution

$$Pr(n) = \frac{\lambda^n}{n!} e^{-\lambda}, \quad (9)$$

with $\bar{n} = \lambda = p \Delta A \Delta t$.

The total heat delivered is $Q = p \Delta A \Delta t Q_t$, so the average rate of change of mean potential temperature is this divided by $\Delta A h \Delta t \rho C_p$; i.e.,

$$\frac{d\langle\theta\rangle}{dt} = p \theta_t A_t, \quad (10)$$

which implies

$$p = \frac{1}{\theta_t A_t} \frac{d\langle\theta\rangle}{dt}. \quad (11)$$

Note that, in a Poisson distribution, the variance in the number of events equals the mean, so the variance in the heat supplied to volume $\Delta A h$ in Δt is

$$\text{var}(Q) = p \Delta A \Delta t Q_t^2, \quad (12)$$

and the standard deviation in the heating rate is given by

$$\sigma_{\dot{\theta}} = \frac{\text{var}(Q)^{1/2}}{\Delta A h \Delta t \rho C_p} \equiv \bar{n}^{-1/2} \frac{d\langle\theta\rangle}{dt}. \quad (13)$$

If we apply this to the box model above, from Eqs. (8) and (11), we can write

$$p = \frac{1}{\theta_t A_t} \frac{w_* \theta_*}{h}. \quad (14)$$

If we assume $\theta_t = \theta_*$, (which makes sense to within a constant, which we shall assume is subsumed into the definitions of thermal area), then

$$p = \frac{1}{A_t \tau_*}, \quad (15)$$

and the Poisson distribution [Eq. (9)] has

$$\bar{n} = \lambda \equiv \frac{\Delta A \Delta t}{A_t \tau_*}. \quad (16)$$

The standard deviation of the heating can be written as

$$\sigma_{\dot{\theta}} = \left(\frac{\Delta A \Delta t}{A_t \tau_*} \right)^{-1/2} \frac{d\langle\theta\rangle}{dt}.$$

Thus, the *relative* standard deviation of the heating is proportional to the inverse square root of the averaging area and time.

This makes clear the intuitive simplicity of this model. If $\lambda = 1$ the *average* number of thermals in area ΔA and time Δt is 1. The *actual* heating rate over this area is

$$\frac{d\langle\theta\rangle}{dt} \text{ stochastic} = \frac{n}{\lambda} \frac{d\langle\theta\rangle}{dt} \text{ average} = \frac{n}{\lambda} \frac{\theta_*}{\tau_*}. \quad (17)$$

As stated above, the variability in heating is purely due to variability in n ; if we had included a distribution of thermal characteristics, that would appear as an additional factor here.

Thus, the stochastic forcing is best considered as multiplicative noise. It can, however, be written as additive, which helps comparison with other schemes. We can write

$$\frac{d\langle\theta\rangle}{dt} \text{ stochastic} = \frac{d\langle\theta\rangle}{dt} \text{ average} + \left(\frac{n}{\lambda} - 1 \right) \frac{d\langle\theta\rangle}{dt} \text{ average}, \quad (18)$$

or, in terms of the boundary layer scales,

$$\frac{d\langle\theta\rangle}{dt} \text{ stochastic} = \frac{d\langle\theta\rangle}{dt} \text{ average} + \left(\frac{n}{\lambda} - 1 \right) \frac{\theta_*}{\tau_*}. \quad (19)$$

In practice, numerical models work with discrete time steps, and model parameterization schemes produce an increment per time step $\Delta\theta$. Over time step δt the stochastically perturbed increment can be written as

$$\Delta\theta_{\text{stochastic}} = \Delta\theta_{\text{average}} + \varepsilon_p \Delta\theta_{\text{average}}, \quad (20)$$

or, in terms of boundary layer scales,

$$\Delta\theta_{\text{stochastic}} = \Delta\theta_{\text{average}} + \varepsilon_p \frac{\delta t}{\tau_*} \theta_*. \quad (21)$$

where $\varepsilon_p = n/\lambda - 1$ is a random number and the subscript is a reminder that this originates from a Poisson process, so a new value of ε_p arises only at intervals Δt (assumed, for simplicity, an integral multiple of δt). It has properties

$$E(\varepsilon_p) = 0, \tag{22}$$

$$E(\varepsilon_p^2) = \lambda^{-1}. \tag{23}$$

Furthermore, as λ increases, the Poisson distribution tends to a normal distribution, so ε_p tends to a normal variate with mean zero and variance λ^{-1} .

The factor $(\delta t/\tau_*)\theta_*$ is just the average increment per model time step $\Delta\theta_{\text{average}}$. The factor $\delta t/\tau_*$ is the average number of thermals per model time step at a point or, alternatively, since, in practice, $\delta t < \tau_*$, the average fraction of a thermal in a model time step at a point, as $\tau_*/\delta t$ is the average number of model time steps a thermal lasts at a point.

The variance of the stochastic increment, $\Delta\theta_{\text{inc}} = \varepsilon_p\Delta\theta_{\text{average}}$, is thus

$$E(\Delta\theta_{\text{inc}}^2) = \frac{A_t}{\Delta A} \frac{\delta t^2}{\Delta t \tau_*} \theta_*^2. \tag{24}$$

Recalling that n is the number of thermals occurring in time Δt (and area ΔA), Eqs. (21) and (24) can be applied in two different ways. If $\Delta t = \delta t$, so we apply all the heating from a thermal in a model time step, the standard deviation of the stochastic increment gets smaller with smaller model time step, but only as the square root of the model time step, as the relative sampling noise increases.

If, instead of the model time step, we choose an averaging time $\Delta t = \tau_*$, i.e., apply the same increment over a time period equal to the correlation time scale, then this becomes

$$E(\Delta\theta_{\text{inc}}^2) = \frac{A_t}{\Delta A} \frac{\delta t^2}{\tau_*^2} \theta_*^2, \tag{25}$$

so it is evident that we are merely spreading the overall increment over the number of time steps in τ_* , and the remaining factor depends just on the number of thermals in area ΔA at any one time. The convective boundary layer depth, h , is typically 1 km, so a model with a horizontal averaging length scale of around 1 km will have $\Delta A \approx A_t$ and standard deviation in temperature increment (added over τ_*) of θ_* .

In practice, the scales actually represented well by a model are several (at least five or six) grid lengths (Skamarock et al. 2014). Thus, $\Delta A \approx m^2 \Delta x^2 > A_t$, where m is about 5 or 6. Thus, in practice, the averaging area represents several thermals in models with horizontal grid length 1 km or larger, and the assumption of lack of horizontal correlation seems very reasonable. Hence, for models with grid length greater than around 100–200 m, there seems *no need to impose a horizontal correlation or structure* to the stochastic increments.

On the other hand, the time step of numerical models depends on factors such as the Courant number ($u\Delta t/\Delta x$, where u is the maximum horizontal wind speed). The MetUM has a semi-implicit, semi-Lagrangian dynamical core that is not restricted by the Courant number in terms of stability, but still

uses a time step with maximum Courant number a small integer (typically 5) to maintain accuracy. The operational 1.5 km UKV uses a 60 s time step. Many other models, such as WRF (Skamarock et al. 2008), are more restricted in time step. In the typical example in section 3a, $\tau_* = 460$ s. This tends to vary less than surface buoyancy flux and boundary layer depth as a large surface buoyancy flux (and hence large w_*) tends to produce a deep boundary layer depth. Thus, $\Delta t/\tau_*$ is often less than, or even much less than, 1 in many convection-permitting models with resolution of a few km. This suggests that we *do need to take into account the correlation of time perturbations*. Nevertheless, since we are averaging over the horizontal structure, there seems little justification in treating the temporal correlation in detail. We have chosen to include time correlation simply via a first-order autoregressive process with autocorrelation time scale determined by the characteristic eddy turnover time, τ_* .

c. Temporally correlated Poisson perturbations as a first-order autoregressive process

Let us introduce a more general stochastic process h_i , taken to be the rate of change of state variables in time step i . For simplicity, as above, we shall just consider θ , so

$$\Delta\theta_i = h_i\delta t. \tag{26}$$

We model h_i as a (stationary) first-order autoregressive stochastic process, in which we assume that a portion μh_{i-1} of the heating comes from thermals that “fired” in previous time steps, and a portion $n_i q$ comes from n_i new thermals in time step i and delivering q heating rate. Thus,

$$h_i = \mu h_{i-1} + n_i q. \tag{27}$$

Here $\mu \in [0, 1)$, and the number of new thermals (n_i) is assumed to follow the Poisson process.

Taking expected values (i.e., ensemble mean) of this first-order autoregressive stochastic process leads to

$$E(h_i) = \bar{h} = \frac{\lambda q}{1 - \mu}, \tag{28}$$

with $\lambda = \bar{n}$; hence,

$$q = \frac{(1 - \mu)\bar{h}}{\lambda}. \tag{29}$$

As above, this may be rearranged as an additive stochastic process; thus,

$$\Delta\theta_i = \bar{h}\delta t + \left(\frac{h_i}{\bar{h}} - 1\right)\bar{h}\delta t. \tag{30}$$

If we write

$$f_i \equiv \frac{h_i}{\bar{h}} - 1, \tag{31}$$

then

$$\Delta\theta_{\text{stochastic}} = \Delta\theta_{\text{average}} + f_i\Delta\theta_{\text{average}}. \tag{32}$$

This is the same stochastic model as Eq. (20), except instead of simple multiplicative noise we have a multiplicative stochastic process with

$$f_i = \mu f_{i-1} + (1 - \mu)\varepsilon_i, \quad (33)$$

where

$$\varepsilon_i \equiv \frac{n_i}{\lambda} - 1. \quad (34)$$

This has the properties

$$E(\varepsilon_i) = E(f_i) = 0, \quad (35)$$

$$E(\varepsilon_i^2) = \lambda^{-1}, \quad (36)$$

$$E(f_i^2) = \frac{1 - \mu}{1 + \mu} E(\varepsilon_i^2) = \frac{1 - \mu}{1 + \mu} \lambda^{-1}, \quad (37)$$

and

$$E(f_i f_{i+m}) = E(f_i^2) \mu^{|m|}. \quad (38)$$

Note that the variance of the correlated factors is less than that of the uncorrelated noise because the same heating is delivered over a longer time determined by μ . We need to parameterize μ . The “integral time scale” for this process is given by

$$\sum_{m=0,\infty} \mu^m \delta t = \frac{1}{1 - \mu} \delta t. \quad (39)$$

We assume that this is proportional to τ_* ; i.e.,

$$\frac{1}{1 - \mu} \delta t = \alpha \tau_*, \quad (40)$$

where α is a tuning parameter of order one. Clearly, negative μ makes no physical sense, so we can write

$$\mu = \max\left(0, 1 - \frac{\delta t}{\alpha \tau_*}\right). \quad (41)$$

If $\mu = 0$ perturbations are uncorrelated in time and are assumed to occur entirely in one time step δt . The option to assume uncorrelated perturbations everywhere has been implemented but, in the following, time-correlated perturbations are always used.

From Eq. (29),

$$q = \frac{\bar{h}}{\lambda \alpha \tau_*} \delta t = \frac{A_t}{\alpha \Delta A} \bar{h}. \quad (42)$$

Note that a single thermal delivers total heating (summed over all time steps) given by

$$\Delta \theta_{\text{thermal}} = q \sum_{m=0,\infty} \mu^m \delta t = \frac{A_t}{\Delta A} \theta_*, \quad (43)$$

as expected.

We can write the autocorrelation function in exponential form:

$$\mu^{|m|} = e^{-|m| \delta t / \tau_e}, \quad (44)$$

so the exponential correlation time scale is $\tau_e = -\delta t / \ln \mu$. For small $\delta t / \alpha \tau_*$, $\ln \mu \approx -\delta t / \alpha \tau_*$ so $\tau_e \approx \alpha \tau_*$. Thus, for small $\delta t / \alpha \tau_*$ the scheme approximates to an exponential decay of the heating from each thermal. However, the parameterization of μ in terms of τ_* , Eq. (41), gives us a natural cutoff. If $\delta t \geq \alpha \tau_*$ then we revert to an “uncorrelated” formulation without any “memory,” with $E(f_i^2)$ naturally tending to λ^{-1} as in the uncorrelated case.

4. Implementation in the MetUM

a. Procedure

As noted above, in practice, the scales actually represented well by a model are several (at least five or six) grid lengths. It is thus unlikely that variability on the grid scale will couple efficiently with the dynamics of the model. We take the area ΔA to be the area covered by $n_g \times n_g$ grid boxes and so assume that each area ΔA is spatially uncorrelated. The integer n_g is an adjustable parameter of the scheme. Over each area we obtain an average λ [using Eq. (16)] and μ [using Eq. (41)] with $\Delta A = n_g^2 \Delta x^2$, Δt the model time step, $A_t = h^2$, $\tau_* = h / w_m$, and

$$w_m = \max[0.4, (u_*^3 + 0.25 w_*^3)^{1/3}]. \quad (45)$$

This is taken directly from the MetUM boundary layer scheme (Lock et al. 2000) and represents a generalized boundary layer time scale that includes the effect of wind shear through the friction velocity u_* as well as the convective velocity scale. The latter is given by

$$w_*^3 = h \overline{w' b'}, \quad (46)$$

with $\overline{w' b'}$ the surface buoyancy flux, essentially Eq. (4) including the effect of moisture. The form of this essentially comes from surface layer similarity and was first proposed by Panofsky et al. (1977) and is discussed in Holtslag and Boville (1993), though they use a value 0.6 rather than 0.25; the choice of constant 0.25 is as in the Lock et al. (2000) scheme, though its origin is not entirely clear. This extends the use of the scheme to both neutral and (noting that a negative surface buoyancy flux reduces w_m) stable conditions, though the minimum value of 0.4 m s^{-1} , used for consistency with the boundary layer scheme, may override this in practice. This minimum has little impact in practice, as the scheme produces very small perturbations in stable or even very weakly convective conditions, but it does serve to ensure that extremely short correlation time scales do not occur.

Note that it would probably be more consistent to average the surface buoyancy flux, friction velocity and boundary layer depth over the $n_g \times n_g$ grid boxes to derive λ and μ but our approach was adopted to be more efficient on a parallel computer and has little practical impact in relatively homogeneous terrain. Furthermore, because of differences in land use, the surface buoyancy flux can vary substantially from grid box to grid box, while λ tends to vary less. As discussed above, this heterogeneity due to differences in land use is an explicit and

deterministic source of small-scale variability. The multiplicative nature of the scheme means that this heterogeneity carries through to the stochastic forcing.

For each of the grid boxes on the reduced two-dimensional grid (i.e., each box is $n_g \times n_g$ model grid boxes) a random integer n_i , drawn from a Poisson distribution, is generated. Note that with λ substantially less than 1, as is often the case, this distribution is highly skewed, with 0 occurring frequently. The random factor f_i is then generated using Eqs. (34) and (33). The latter reduces to $f_i = \varepsilon_i$ if $\mu = 0$, i.e., uncorrelated perturbations. Finally, a perturbation f_i times the boundary layer increment is added to the model physics forcing. The option is implemented to apply this selectively to the potential temperature, moisture, or wind fields; in practice we have applied it to all. The scheme has been implemented outside the boundary layer scheme as an additional, stochastic forcing scheme. Ideally, it should be moved inside the scheme and the stochastic increments applied to the surface exchange as well, in order to conserve energy. Currently, energy is only conserved to the extent that the ensemble average of the stochastic increments is zero.

The scheme can be summarized thus:

$$\Delta\phi_{\text{stochastic}} = \Delta\phi_{\text{average}} + f_i \Delta\phi_{\text{average}}, \quad (47)$$

where ϕ represents the model u , v , θ and q_v . If $\Delta t < \alpha\tau_*$, f_i is a first-order autoregressive stochastic process, randomly forced by the Poisson-related variable ε_i [Eq. (34)] and with autocorrelation time scale $\alpha\tau_*$. Since ε_i has zero mean and variance λ^{-1} , f_i has zero mean and variance, given by Eq. (37). Thus, inclusion of time correlation greatly reduces the dependency of the noise on time step, especially when $\Delta t \ll \alpha\tau_*$, as it often is at resolutions $O(1)$ km.

A time step dependence is, of course, present as we have multiplied the increment in a time step by this noise factor. In a bulk model of the CBL our scheme reduces to

$$\Delta\theta_{\text{stochastic}} = \Delta\theta_{\text{average}} + f_i \frac{\delta t}{\tau_*} \theta_*. \quad (48)$$

The variance of the stochastic increment $\Delta\theta_{\text{inc}} = \Delta\theta_{\text{stochastic}} - \Delta\theta_{\text{average}}$ is thus

$$E(\Delta\theta_{\text{inc}}^2) = \frac{\alpha^{-1} A_i}{2 - \frac{\delta t}{\Delta A}} \left(\frac{\delta t}{\tau_*} \theta_* \right)^2. \quad (49)$$

Note that the area factor $A_i/\Delta A$ in Eq. (49), can be written $(h/n_g \Delta x)^2$ so the standard deviation of the noise will depend on $h/(n_g \Delta x)$.

b. Key features

The scheme outlined has the following physically sensible features:

- 1) The key parameter is \bar{n} [Eq. (16)], the average number of thermals triggered in the selected area ΔA in time Δt .
- 2) The stochastic increments have zero mean on average (in a steady system).

- 3) The stochastic increments have an amplitude proportional to the mean increments times $\lambda^{-1/2}$. Thus, more thermals means smaller stochastic increments.
- 4) The stochastic increments have time correlation (via an AR1 process) related to the turnover-time. If this is less than the time step, stochastic increments are uncorrelated from time step to time step.

5. Comparison with other schemes

The implemented scheme uses the parameterized increments [using the boundary layer parameterization described by Lock et al. (2000)]. However, in a CBL and cumulus-capped CBL, the results broadly resemble the bulk scheme described above, so we shall use this as a basis for comparison.

A relatively simple stochastic scheme has already been implemented by A. Lock (2016, personal communication) with more recent details in Beare et al. (2019). This assumes

$$\Delta\theta_{\text{stochastic}} = \Delta\theta_{\text{average}} + g_i \frac{\delta t}{\delta t_0} \theta_*. \quad (50)$$

Here g_i can be either a uniformly distributed variable ε_U in the range $[-1, 1]$ or a temporally correlated variable with

$$g_{i+1} = \mu_L g_i + \sqrt{1 - \mu_L^2} \varepsilon_U. \quad (51)$$

The factor μ_L here is given by

$$\mu_L = \exp(-\delta t/\tau), \quad (52)$$

with τ a fixed autocorrelation time scale, usually taken to be 10 min. Thus, the temporal correlation is very similar to our scheme, apart from the fixed choice of time scale instead of the time scale varying with conditions. In practice, the fixed time scale of 10 min is quite representative of the values of τ_* found in the CBL.

The factor $\delta t/\delta t_0$ scales the increment with time step, with $\delta t_0 = 60$ s the time step of the UKV model the scheme was first implemented in. Both g_i and ε_U have variance 1/3 so the variance of the θ increment is

$$E(\Delta\theta_{\text{inc}}^2) = \frac{1}{3} \left(\frac{\delta t}{\delta t_0} \theta_* \right)^2. \quad (53)$$

This may be compared with the variance of the stochastic increment in our scheme applied to a bulk CBL [Eq. (49)]. The dependence on θ_* is essentially the same. The scheme is implemented with a single value g_i for each box of a ‘‘supergrid’’ of $n_g \times n_g$ points similar to our scheme but the dependence on ratio of eddy area to averaging area is absent in the Lock scheme, meaning that the same amplitude of noise is applied irrespective of the averaging area. As discussed in section 4a, our scheme reduces the standard deviation by a factor roughly $h/(n_g \Delta x)$. The time step dependence is the same, but normalized by δt_0 rather than τ_* . These differ by an order of magnitude. The remaining factors, about 1/2 [Eq. (49)] compared with 1/3 [Eq. (53)], are a consequence of different assumed probability density functions. Overall, in conditions with

$h \approx \Delta x$ and $\tau_* \approx 10\text{min}$, the Lock scheme will produce stochastic increments about a factor $6n_g$ larger than our scheme.

KC describe a scheme designed to address similar issues. A stochastic field is added to the model physics tendencies [KC Eq. (1)]; thus,

$$\left(\frac{\partial\Phi}{\partial t}\right)^{\text{stoch}} = \frac{\partial\Phi}{\partial t} + \alpha_{\text{sh}} \eta \langle \Phi^2 \rangle^{1/2}, \quad (54)$$

where α_{sh} is a constant, η is a Gaussian random number field and $\langle \Phi^2 \rangle^{1/2}$ is the variance of variables given by the turbulence scheme in their model. The scheme is applied to variables temperature T , vertical velocity w , and specific humidity q . The Gaussian random number field contains spatial and temporal correlations. The temporal correlation arises from holding the field fixed for 10 min, then generating a new field. The 10 min represents the eddy turnover time. Spatial correlation in η is provided by convolving a set of random numbers drawn from a Gaussian distribution with a horizontal two-dimensional Gaussian with standard deviation 2.5 grid points, representing a correlation length scale of $5\Delta x$ (in our notation). The resulting field is normalized to have mean 0 and variance $E(\eta^2) = 1$.

This has similarities with our scheme; in our scheme, we choose a fixed area in space, generally larger than the expected horizontal correlation length scale, on the grounds that the horizontal scales well represented by the model are much larger than the spatial correlation scale. However, we explicitly represent the temporal correlation, on the grounds that the time step is shorter than the temporal correlation scale. Thus, the spatial pattern is stepped (or “checkerboarded”) but the temporal pattern varies relatively smoothly from time step to time step. The KC scheme does the reverse, representing smooth correlations explicitly in space on scales roughly the smallest that are well-represented by the model, but with a stepped time profile, each fixed period given by an equal 10 min with no correlation between 10-min periods. Thus, the physical idea behind temporal correlation is consistent with our scheme but the implementation does not adapt to different conditions in the same way.

Note that a revised version of the KC scheme has been published more recently (after first submission of this paper) in Hirt et al. (2019); this, among other changes, implements an autoregressive temporal correlation scheme similar to that of the Lock scheme discussed above. The scheme is implemented as in Eq. (33), with ε a spatially correlated “Gaussian bump” field and the same fixed correlation time as used in KC. Though the end result is very similar, we regard a key difference in our scheme as the link to the Poisson process expressed in the starting-point equation, Eq. (27), which provides a self-consistent scale adaptivity, as well as the physically based correlation.

The “checkerboard” in space in our scheme is necessary as the amplitude and correlation time scale vary in space (and time). No doubt some smoothing could be applied to the spatial pattern of parameters, if necessary, but in practice the increments in a time step are so small that this structure does not seem to have any significant impact. Both schemes assume vertically coherent perturbations.

According to KC, the parameter α_{sh} is given by

$$\alpha_{\text{sh}} = \alpha_{\text{sh},\Phi} \frac{l_{\infty}}{5\Delta x} \frac{1}{dt}, \quad (55)$$

where l_{∞} is the asymptotic mixing length describing the average size of an eddy assumed in the turbulence parameterization, dt is equal to the “temporal resolution of the model,” 25 s, and $\alpha_{\text{sh},\Phi}$ is a tunable parameter set equal to 2 for all variables. This does not seem to make physical sense. The stochastic increment in each time step δt over the 10 min that a choice of η applies, is thus

$$\Delta\Phi = \alpha_{\text{sh}} \eta \langle \Phi^2 \rangle^{1/2} \delta t = \alpha_{\text{sh},\Phi} \frac{l_{\infty}}{5\Delta x} \eta \langle \Phi^2 \rangle^{1/2} \frac{\delta t}{dt}, \quad (56)$$

i.e., independent of δt if $\delta t \equiv dt$. This means that if one were to halve the model time step, twice the total increment would be applied in the 10-min period η applies for. Given the time step of 25 s quoted, this would also lead to a very large increment. Equation (56) only makes physical sense if dt is a fixed time scale independent of the time step. Indeed, the revised scheme described in the appendix of Rasp et al. (2018) replaces this with τ_{eddy} . The tuning parameter was also changed (to 7.2 from 2). This partly offsets the change in time scale from 25 to 600 s, but clearly a change from 2 to 48 would be required to completely offset the change in time scale. The authors state that the “tuning factor was chosen so that the effects of the PSP scheme were noticeable but reasonable”; we understand this to mean the same methodology as KC was applied to optimize the amount and onset time of domain integrated precipitation for a strongly nonequilibrium case followed by a check that the same choice of tuning constant gave a similar improvement in other nonequilibrium cases, and while leaving the total precipitation largely unchanged in equilibrium cases. Whatever it means, clearly the size of perturbations applied in the revised scheme should be a factor of around 7 smaller than in the KC scheme. The scheme reported by Hirt et al. (2019) also uses τ_{eddy} .

If we denote the revised scheme of Rasp et al. (2018) KCR, considering temperature T , the increment ΔT_{KCR} in a time step δt is given by

$$\Delta T_{\text{KCR}} = \alpha_{\text{sh}} \eta \langle T^2 \rangle^{1/2} \delta t \quad (57)$$

and so has variance

$$E(\Delta T_{\text{KCR}}^2) = \alpha_{\text{sh},T}^2 \left(\frac{l_{\infty}}{5\Delta x}\right)^2 E(\eta^2) \langle T^2 \rangle \frac{\delta t^2}{\tau_{\text{eddy}}^2}, \quad (58)$$

with $E(\eta^2) = 1$

This may be compared with Eq. (25). Overall, the ratio of the KC scheme standard deviation in temperature increment per time step to that in our scheme is

$$\frac{n_g \alpha_{\text{sh},T} l_{\infty} \langle T^2 \rangle^{1/2} \tau_*}{5 h \theta_* \tau_{\text{eddy}}}. \quad (59)$$

Clearly $\tau_* \sim \tau_{\text{eddy}}$, the only difference being that in our scheme the local value is used rather than a global constant. The first three terms highlight the key physics in the scheme:

the horizontal-averaging length scale (n_g), the length scale of turbulent eddies, and the variability in temperature due to turbulent eddies. The averaging scale is based upon a very similar premise and differs only in being more (hopefully realistically) physically flexible in our scheme (i.e., if we choose a larger averaging scale, we get smaller variability). A key difference in formulation is that we take the length scale as the length scale occupied, on average, by one eddy, rather than the mixing length in a diffusion approach, which is assumed to be the scale of the eddy itself. Thus l_{∞} , typically 150 m, is much smaller than h (and, indeed, in some schemes is taken to be $0.15h$). However, including the factor of $\alpha_{sh,T} = 7.2$ in this term means $\alpha_{sh,T} l_{\infty} / h \approx 1$.

Garratt (1994) suggests T^2/θ_*^2 ranging from about 1.85 in the middle of the boundary layer to 10 or more near the top and bottom. Thus, in practice, the KCR scheme applies perhaps 3 or 4 times larger perturbations near the top and bottom of the boundary layer than our scheme. While perhaps of practical value, we would argue that this is physically inconsistent as the turbulent length scale clearly decreases in these regions. For example, Kaimal et al. (1976) (Fig. 5) show smaller length scales lower in the boundary layer where $\langle \theta^2 \rangle$ is larger. Thus, if the local variance from the turbulence scheme is used at a given height, it would seem more appropriate to use a height-varying length scale. The effect would be to at least partially cancel the variability in each.

We have focused primarily on amplitude of perturbations, and particularly on temperature. The recent enhancements of the KC scheme by Hirt et al. (2019) focus on a nondivergent wind perturbation which is found to force the vertical velocity more effectively. This seems to be a useful innovation; Hirt et al. (2019) emphasize that this enhances coupling with the resolved scales as less of the perturbation projects onto sound waves, so comparing magnitudes may not give a true reflection of impact. Whether this is physically important at the scales we are applying perturbations at (several km) may be questionable, but the approach may become more important as models move more into the subkilometer “turbulence gray zone.” We did not pursue this in this initial implementation, in part, because the MetUM already has a kinetic energy backscatter scheme and, in part, to maintain simplicity. It is certainly likely that small-scale vertical motions at the top of the boundary layer play a crucial role in coupling to shallow cumulus, but this is likely to be at a scale much smaller than those we are considering, and future work will focus on extending the scheme to include parameterized coupling with unresolved vertical motion in the cumulus scheme along with the backscatter scheme.

In summary, at least when considering the magnitude of temperature fluctuations, the schemes are broadly similar in both magnitude and dependencies. The two schemes will produce different vertical profiles of perturbation; ours is simple and follows the parameterization profile. The KC scheme has a profile dependent on the profile of variability, but seems slightly flawed in not including the compensating profile of turbulence length scale. Our scheme seems better justified when considering the relative sizes of grid box and time step compared with spatial and, in particular, temporal correlation scales, respectively. Furthermore, the Poisson-based stochastic

process is properly scale adaptive (at least as far as the assumptions go), so naturally gives higher amplitudes when run at smaller scales that average to the consistent amplitudes at larger scales. However, in practice, the manner in which correlations are applied probably makes little difference (as discussed further below), and the main difference is that temporal correlation in our scheme depends explicitly on the local boundary layer properties, and our random function is likely to be substantially more skewed.

Overall, despite their apparent differences in formulation, the schemes are notable more for their similarity than their differences, the main key difference being the nondivergent increments in the later Hirt et al. (2019) variant.

6. Diagnostics

a. Introduction to diagnostics

Diagnostics for testing the scheme have been chosen to allow examination of both the changes in magnitude induced by the scheme and the spatial differences. As the scheme is written for convective-scale models, careful interpretation of more “traditional” techniques is required. To combat the differences between interpretation of diagnostics on the convection-permitting compared to convection-parameterizing scales either the metrics have been adapted (as discussed below) or cell statistics have been used. All of the diagnostics have been calculated over the analysis boxes shown in Fig. 1.

b. Mean square difference

The mean square difference (MSD) is a simple measure of the difference of magnitude between two forecasts, and is more frequently used in its square root form (e.g., Hohenegger and Schär 2007; Clark et al. 2009; Johnson et al. 2014). Here, however, we use it in the squared form as in Leoncini et al. (2010, 2013) and, like Flack et al. (2018), we define the MSD as

$$\text{MSD} = \frac{1}{\Sigma P_c^2} \Sigma (P_p - P_c)^2, \quad (60)$$

where P_p is the hourly accumulation of precipitation in the perturbed forecast, and P_c is the hourly accumulation of precipitation in the control forecast. The normalization factor implies that the MSD will not be altered by the total amount of precipitation in the forecast and so multiple cases with different precipitation amounts can be compared fairly.

The MSD is kept in its square form so that, when considering precipitation forecasts, it can be broken down into the MSD arising from precipitating points only in the control forecast, those points only in the perturbed forecast and those points that have precipitation in the same location in both the perturbed and the control (common points). Considering only the common points (MSD_{common}) between both forecasts removes the “double penalty” problem faced by all gridpoint diagnostics at the convective scale. By removing the “double penalty” problem the interpretation of this diagnostic at the convective scale reverts to its traditional interpretation of magnitude differences as opposed to a mixture of magnitude and spatial differences.

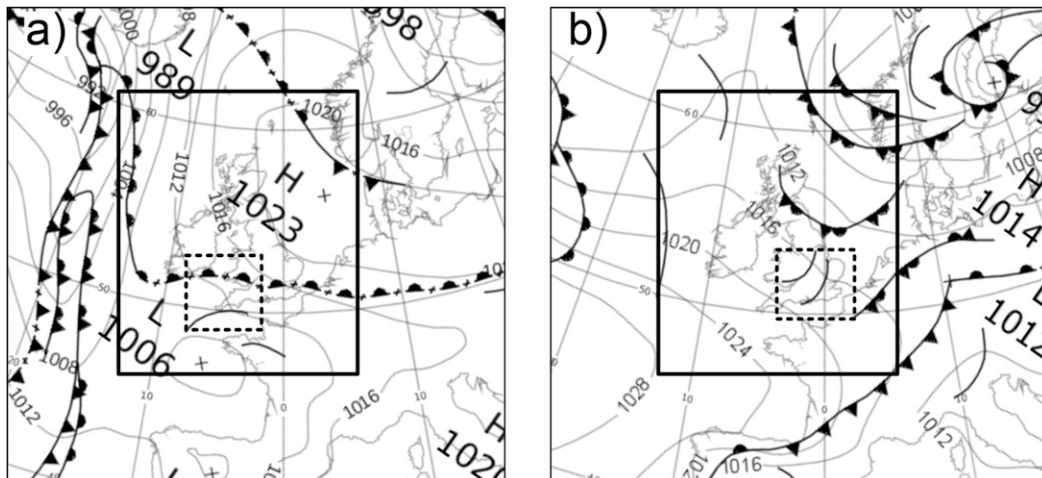


FIG. 1. Met Office analysis charts valid at (a) 0000 UTC 18 Jul (Coverack case) and (b) 0000 UTC 5 Aug 2017 (Kent case). Both show the model domain (black rectangle) and analysis domain (black dashed rectangle).

The threshold for the precipitation used within this study is an hourly accumulation of 1 mm.

c. Fraction of common points

The fraction of common points (F_{common}) is a useful concept alongside $\text{MSD}_{\text{common}}$ as it helps determine the fraction of the precipitating points used within the calculation of $\text{MSD}_{\text{common}}$. The fraction is defined by the total number of precipitating points in common between two forecasts (above an arbitrary threshold: in this case hourly precipitation accumulations exceeding 1 mm) divided by the worst case scenario of precipitating points (i.e., when both forecasts have all the rainfall in different locations); i.e.,

$$F_{\text{common}} = \frac{N_{p,c}}{N_p + N_c - N_{p,c}}. \quad (61)$$

Here, N is the number of precipitating points above the threshold and the subscripts p and c refer to the forecasts being compared (perturbed and control forecast, respectively). As with the MSD, this diagnostic has been used in Leoncini et al. (2010), and here we take the normalization of Flack et al. (2018) so that F_{common} ranges between 0 and unity, where unity represents a forecast where all of the precipitation is in the same location, and 0 represents a forecast where none of the precipitation is in the same location. Had the subtraction of $N_{p,c}$ not occurred in the denominator then the fraction would range between 0.0 and 0.5.

As in Flack et al. (2018), the F_{common} values are compared to the fraction of common points that would be yielded from a random relocation of precipitating cells, based upon the number of cells in the control forecast (F_{chance}). A value of F_{common} smaller than or close to F_{chance} implies that the spatial forecasts of the events are near to a random forecast.

d. Cell statistics

The precipitation field from each forecast comprises discrete areas, or cells, of rain. To determine the distribution of size and

strength of these cells, the cells are first identified using a tracking algorithm described in Hanley et al. (2015). Each identified cell is then described as having an area-equivalent diameter and a mean rain rate; area-equivalent diameter is defined as the diameter a circle with the equivalent area would have. Area and rain rate thresholds are applied to the precipitation fields so that only cells with a specified minimum size and rain rate are considered in the evaluation. In this study, a minimum area of $2\Delta x \times 2\Delta x$ is used and rain rate thresholds of 1 and 4 mm h⁻¹ are imposed, representing light and heavy rain, respectively. After thresholds have been imposed, distributions of cell size and magnitude can be produced, in the form of histograms.

7. Testing the scheme

a. Test cases

While the scheme has been applied to a number of convective cases during its development, two test cases from summer 2017 are presented here. These cases exhibit very different convective behavior and have been chosen to show how the scheme behaves in a situation that is more stochastic in nature (scattered showers) compared to one that is not [a mesoscale convective system (MCS)]. The convective adjustment time scale (e.g., Done et al. 2006) is used to show how different these cases are by considering their placement in the spectrum of convective regimes (e.g., Flack et al. 2018). The cases have also been named after the areas in which either there was flooding or was the most dynamically active.

1) COVERACK CASE: 18 JULY 2017

This case is highlighted by a flash flood that swept through the village of Coverack in southern Cornwall (Cornwall Council 2017). The convective event formed off the coast of Brittany at approximately 1200 UTC and progressed northward, reaching Coverack at around 1400 UTC. Part of the event anchored over Coverack and produced intense rainfall

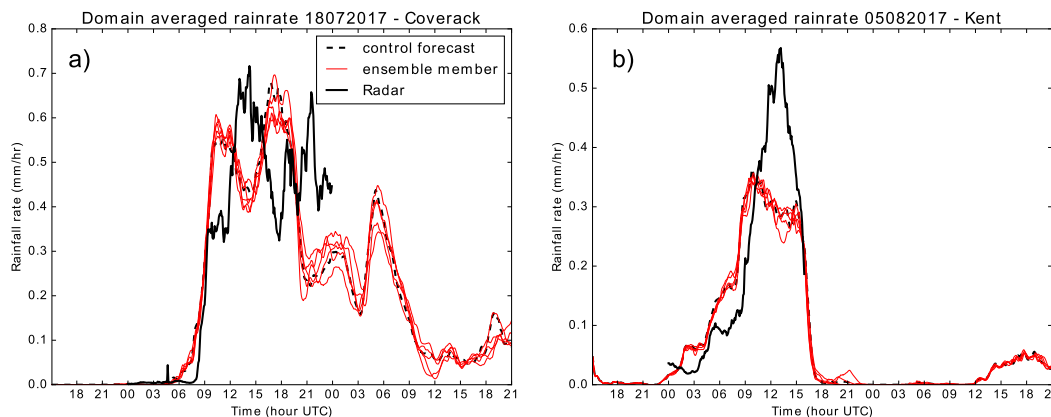


FIG. 2. Domain-averaged rain rate for (a) Coverack and (b) Kent cases.

in the area for approximately 3 h, the main part of the system continued past Coverack and into the surrounding moorland. The event formed as a result of a surface trough (Fig. 1a), and this forcing continued with the event, with it eventually forming into an MCS. This rapidly decayed and later in the day another MCS formed over France, which then moved into southeast England causing surface-water flooding in Reading (Davies 2017). The convective adjustment time scale for this case begins at 4.2 h and reduces to 0.4 h by the end of the convection. Therefore, it sits somewhere in the middle of the regimes for the United Kingdom (cf. with Flack et al. 2016). Using the time scale's value and combining it with the knowledge that local forcing existed to keeping the storm anchored places this case more toward the nonequilibrium end of the spectrum (despite the marginal time scale).

2) KENT CASE: 5 AUGUST 2017

A high pressure center was located over continental Europe and the United Kingdom was dominated by westerly flow (Fig. 1b). This led to showers forming in the lee of the Welsh mountains which were then advected by the flow across England. As the showers formed, they intensified as a result of localized forcing from a surface trough. This surface trough then tracked with the showers and they eventually organized into squall lines over central England. These lines then progressed eastward into East Anglia, the Thames estuary, and North Kent. During the event, one of the authors (D. L. A. Flack) was in North Kent. Along with the intense precipitation associated with this convective line, at the southern end of the line (over the North Kent coast) he observed a mesocyclone and three funnel clouds from 1502 to 1534 UTC. This case has a low convective adjustment time scale throughout the convective life cycle varying between 1.1 and 0.1 h. These low values imply that the case is closer toward the convective quasi-equilibrium end of the spectrum of convective regimes.

b. Model setup

The scheme has been tested in the 1.5 km resolution, UKV configuration of the MetUM at version 10.6 with the Even Newer Dynamics for General Atmospheric Modeling of the Environment (ENDGAME) dynamical core (Wood et al.

2014). The model domain is shown in Fig. 1. This configuration uses the Lock et al. (2000) boundary layer scheme, the Wilson and Ballard (1999) microphysics scheme, the Edwards and Slingo (1996) radiation scheme and the Porson et al. (2010) surface-exchange scheme. The two cases considered here are run such that the most intense convection of the case studies occurs at approximately 24 h into the simulation. The Coverack case is initiated at 1500 UTC 17 July 2017 and the Kent case is initiated at 1500 UTC 4 August 2017. These times are chosen to allow time for the perturbations to spin up so that they can influence the forecast of the convective events.

c. Effect of the scheme

This section illustrates how the introduction of the stochastic scheme described in sections 3 and 4 influences the forecast in the MetUM using the diagnostics introduced in section 6. As described in section 4, the area ΔA over which we are considering the thermals is covered by $n_g \times n_g$ grid boxes. The perturbations to the field are calculated on a reduced two-dimensional grid (with each horizontal dimension reduced by n_g), with each original grid box within an $n_g \times n_g$ gridbox area using identical random perturbations. For each grid box in the original grid, the associated perturbation from the reduced grid is then multiplied by the perturbation to the field due to the boundary layer scheme in that grid box to obtain the resulting perturbation to be used in the scheme. This method is applied to potential temperature, moisture and horizontal wind fields as described in Eq. (48). A value of $n_g = 8$ has been chosen so as to be close to the 5 or 6 Δx required to resolve features on a grid and yet not too large so that the stochastic perturbations are negligible. Sensitivity to this choice is tested in section 7d.

For each case, the MetUM was initially run without the addition of any stochastic perturbations, to produce a control run. The model run was then repeated five times with the scheme included, each time with a different random seed, to produce five ensemble members. This gives an indication of the spread of the forecasts when using the scheme. Figure 2 shows that in both cases the overall domain-averaged rainfall was quite well forecast but certainly not perfectly. Initiation was (perhaps unusually) more gradual in the model than in radar rainfall in the Coverack case. The initiation time of the convection

Rainrate at 14:00 (UTC) 18-07-2017

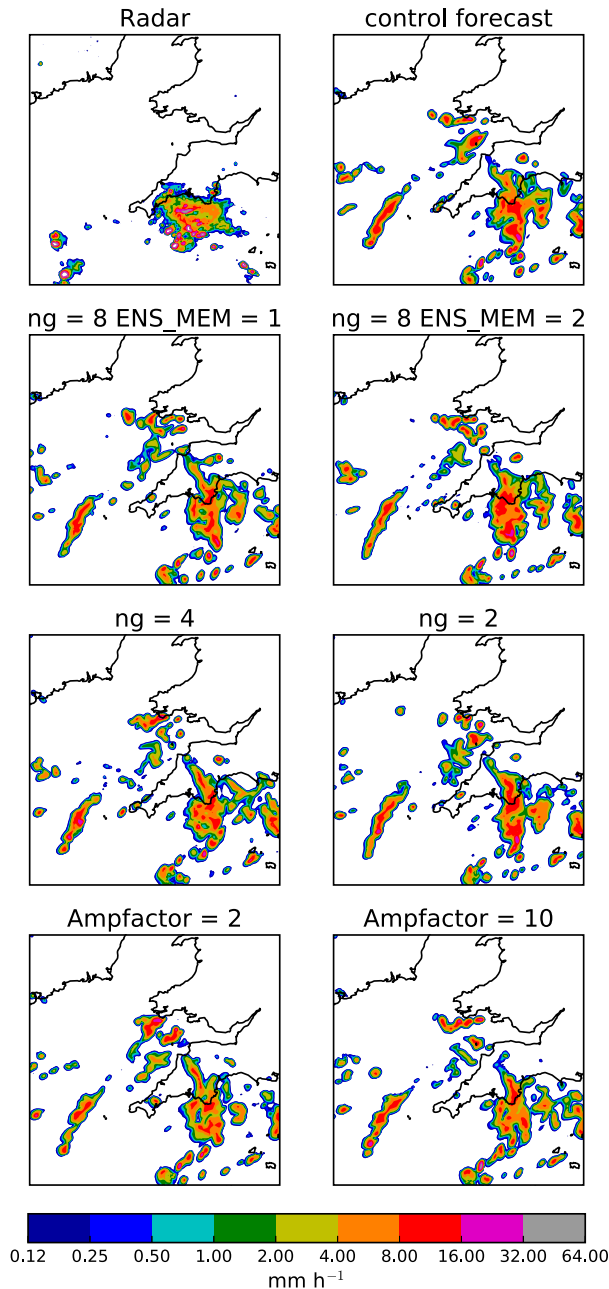


FIG. 3. Coverack: (top left) Radar and precipitation fields at 1400 UTC from (top right) control model run, (second row) ensemble members with different seeds, (third row) ensemble member 1 with different n_g as described in section 4a and (bottom) different values of Ampfactor, as described in section 7e.

is not significantly changed comparing the control run and any of the ensemble members. Furthermore, during the first 6 h after the convection is initiated, the domain-averaged rain rate is almost identical. The largest differences appear later into the run;

Rainrate at 12:00 (UTC) 05-08-2017

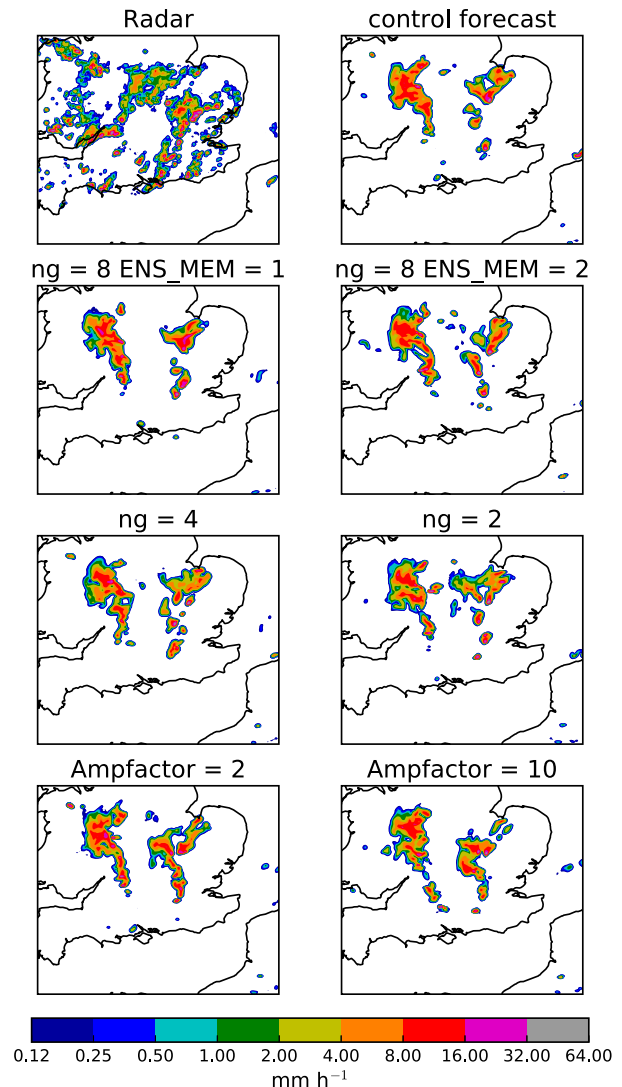


FIG. 4. Kent: (top left) Radar and precipitation fields at 1200 UTC from (top right) control run, (second row) ensemble members with different seeds, (third row) ensemble member 1 with different n_g as described in section 4a and (bottom) different values of Ampfactor, as described in section 7e.

although the peaks and troughs in the precipitation are similar, the magnitude varies by up to 10% or 20%. There is a qualitative impression that the growth of differences between members has saturated well before the end of the forecast—this is discussed more quantitatively below.

This variation is also evident in instantaneous precipitation fields in Fig. 3 for Coverack and Fig. 4 for Kent. In both cases the forecasts, while having larger-scale structure in common with the radar rainfall, have substantial errors at the cloud-cluster scale. The ensemble members show similar larger regions of precipitation (cloud clusters) but differ in the finer detail (cloud scale).

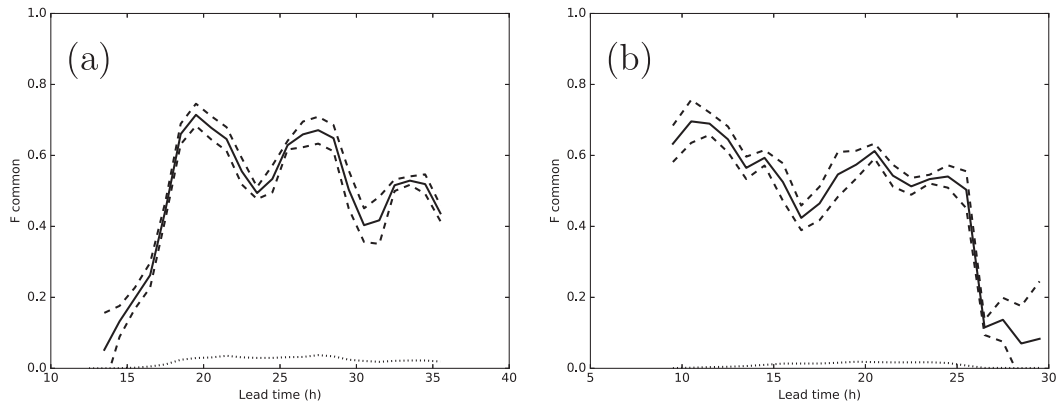


FIG. 5. F_{common} for (a) Coverack and (b) Kent cases showing mean (solid line), ± 1 standard deviation (dashed lines), and F_{chance} (dotted line) for all ensemble members.

The fraction of common points between the control forecast and the ensemble member forecasts (using an hourly accumulation threshold of 1 mm) can be seen in Fig. 5, with the mean of all ensemble members shown as the solid line and the dashed lines indicating ± 1 standard deviation from the mean. When the convection is fully developed in the domain, at a lead time of around 18 h, the fraction of common points generally has a value between 0.4 and 0.8, with a decreasing trend toward the lower end of the range. This supports observations that, while the location of the main regions of precipitation are superficially the same, the details in location and/or timing of the heaviest rain differ between forecasts.

The increases in F_{common} are associated with (i) the formation of new cells collocated in the ensemble members, (ii) enlargement/intensification of cells to encompass other smaller cells in the other members, or (iii) cells coming into the analysis domain from the boundaries, as the boundary conditions remain unchanged throughout each ensemble member. Subsequent falls in the diagnostic are expected as the forecasts begin to diverge from one another.

In the Coverack case, the entry of the MCS into the domain is preceded by the formation of small and very scattered shower cells, which develop in random locations in each

ensemble member roughly 13 h into the forecast. The large area of organized precipitation then moves into the domain accompanied by smaller cells (again, in different locations depending on the ensemble member) and an increase in F_{common} . It is fully in the domain by 1500 UTC ($T + 24$ h) when F_{common} starts to increase again. By 1700 UTC, there are fewer small cells and then the organized system moves out of the domain. It would seem that the increases in F_{common} coincide with the decrease in the number of smaller cells.

The sharp drop at $T + 26$ h in Fig. 5 in the Kent case, is associated with the squall lines leaving the analysis domain and so reflects different timings alongside a drop in the number of precipitating cells, as indicated by the drop in F_{chance} .

The difference in magnitude for the common points between the control and the ensemble forecasts is shown in Fig. 6, where the mean square difference of common points between the control and each ensemble member, is less than 0.5 for all runs. For the Coverack case, the ensemble-averaged $\text{MSD}_{\text{common}}$ remains fairly constant over time (though with a peak at $T + 18$ h). This is probably because the forecast was dominated by the MCS. The Kent case shows more temporal variations that can be associated with the development of the showers and merger into squall lines.

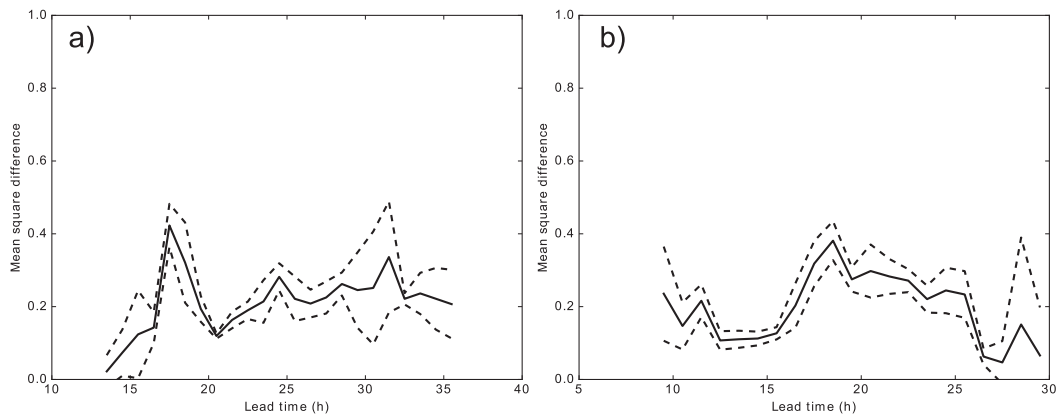


FIG. 6. The MSD over common points for (a) Coverack and (b) Kent cases showing the average MSD (solid line), and ± 1 standard deviation (dashed lines).

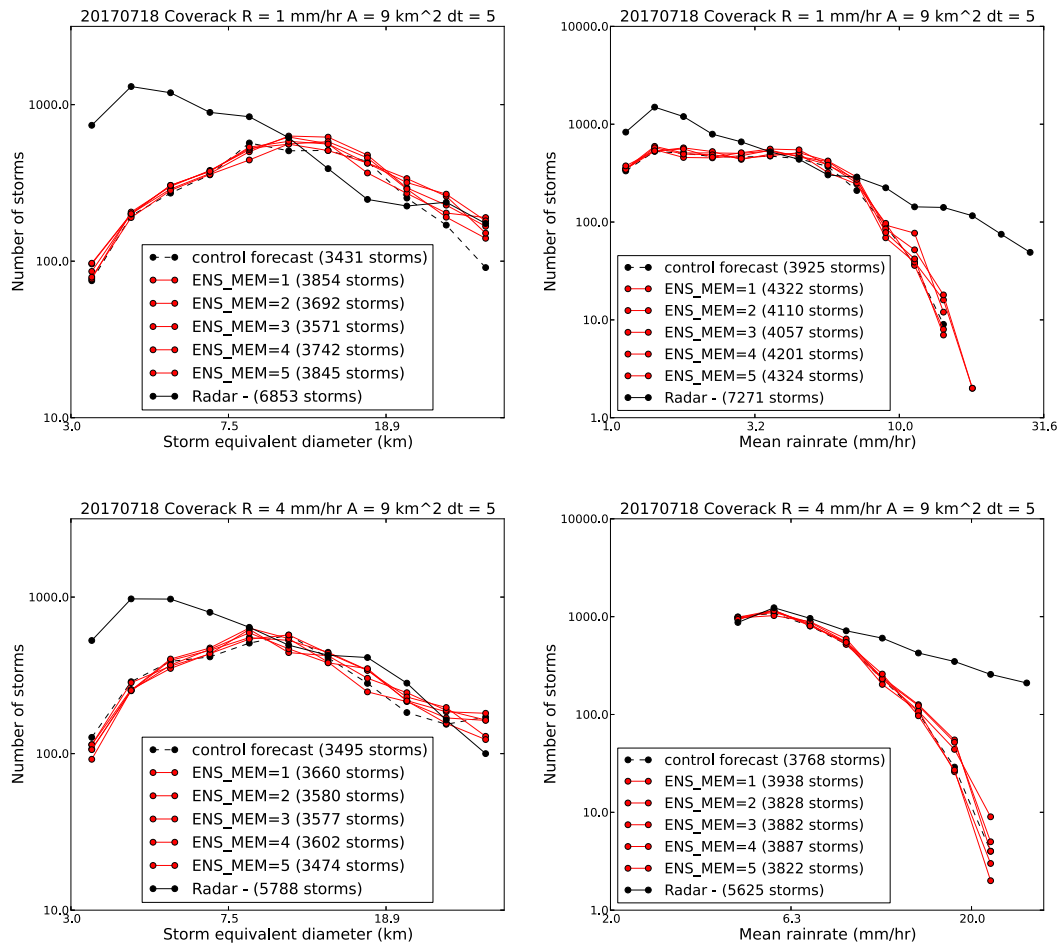


FIG. 7. Coverack: Cell statistics for radar, the control run, and all ensemble members, using the stochastic BL scheme with $n_g = 8$ (as described in section 4a) and different seeds. Rain-rate and area thresholds are indicated in the title of each subplot.

Cell statistics in Fig. 7 (for Coverack) and Fig. 8 (for Kent) show that the distribution of cell size and magnitude is largely similar for each ensemble member. Overall, the total number of precipitation cells ($>1 \text{ mm h}^{-1}$ threshold) appears to increase when the scheme is used and there is a hint that there are more larger cells in the Coverack case, although in general, the distribution of size and magnitude of precipitation cells are similar for the control and each ensemble member. Together with the MSD and F_{common} plots, we can conclude that the cells have similar characteristics for the control and each ensemble member but appear in different locations at any particular time (with evidence that there may be more of the larger cells with 1 mm h^{-1} threshold). On the other hand, the Kent case shows a general reduction in the total number of precipitation cells and is likely to be linked to slightly stronger upper-level forcing in this case compared to the Coverack case.

d. Sensitivity to number of grid boxes, n_g

As described in section 4, the area ΔA in the scheme is defined using n_g grid boxes. This adjustable parameter determines how the stochastic perturbations vary spatially. A value

of 8 is chosen by default for reasons described previously. However, it is still possible to use a smaller (or larger) value of n_g and this section explores the sensitivity of the scheme to the value of n_g in model runs for the cases described above. Specifically, values for n_g of 2 and 4 are compared with the $n_g = 8$ runs already described.

The key features in section 4 highlight that a smaller area ΔA would hold fewer thermals and thus the stochastic increments would be larger in a physically consistent way. This is confirmed in the distributions of the instantaneous perturbations shown in Fig. 9 for both cases. Note that the instantaneous perturbations are those added in a time step; as discussed in section 7c, the total perturbation amplitude delivered by temporally correlated perturbations is of order $\tau_*/\Delta t$, typically an order of magnitude, larger.

There are some differences in the distributions between the cases though. Specifically, in the Kent case with increased n_g , the distribution shows greater frequency of points with a perturbation value of zero, and there is also a narrowing of the distribution (cf. $n_g = 8$ and $n_g = 2$; Fig. 9b), whereas for the Coverack case (cf. $n_g = 8$ and $n_g = 4$; Fig. 9a), the distribution

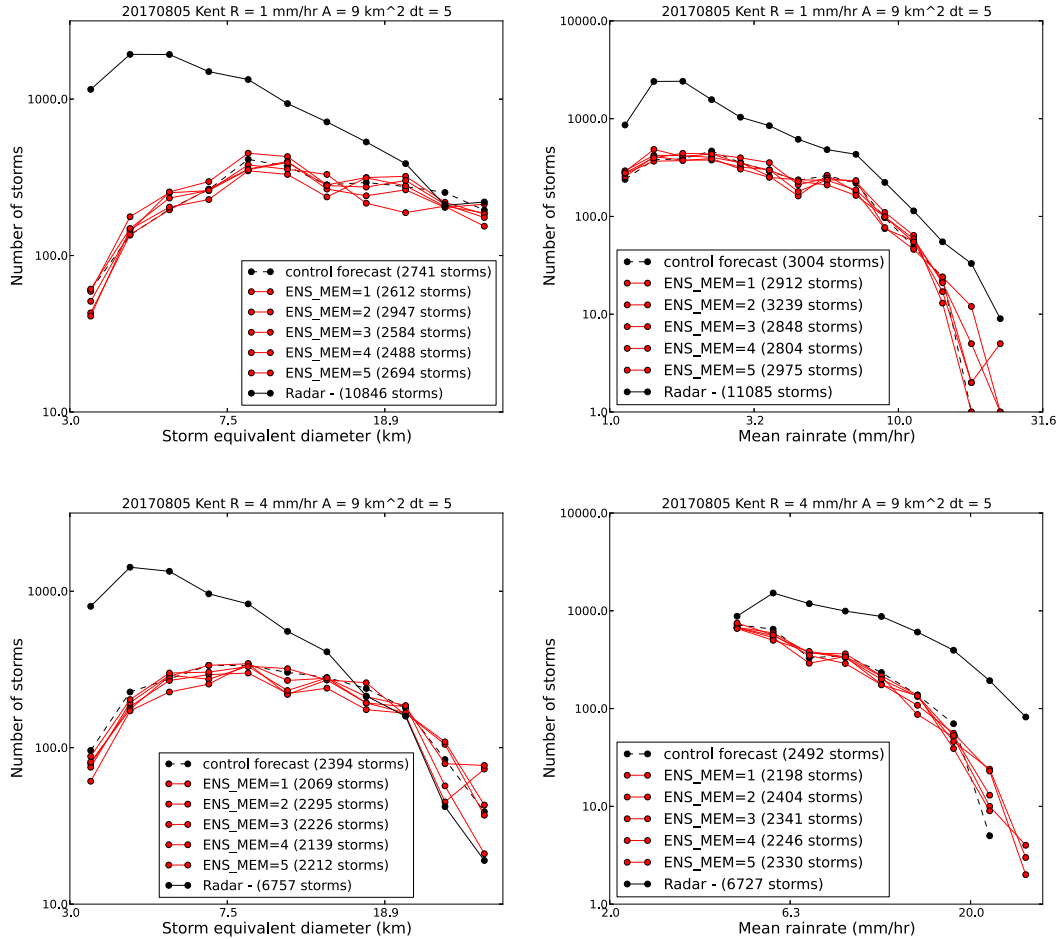


FIG. 8. Kent: Cell statistics for radar, the control run, and all ensemble members, using the stochastic BL scheme with $n_g = 8$ (as described in section 4a) and different seeds. Rain-rate and area thresholds are indicated in the title of each subplot.

only becomes narrower and there is no change in the frequency of points with a perturbation size of zero. These changes in the distribution are linked to the more numerous convection, and hence perturbation occurrence, in the Kent case, compared to

the more locationally restricted perturbations (mainly around the MCS) in the Coverack case.

The sensitivity of the forecast to a change in value of n_g is shown by the time series of average rain rate in Figs. 10a and 10b

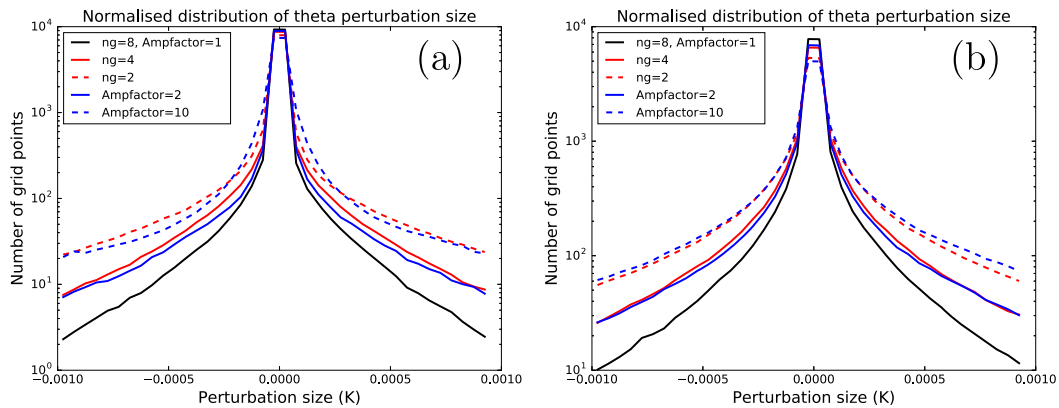


FIG. 9. Distribution of amplitude of perturbations for model with different values of n_g (as described in section 4a) and amplitude factor (as described in section 7e). The black line is the baseline run ($n_g = 8$, Ampfactor = 1). (a) Coverack and (b) Kent case.

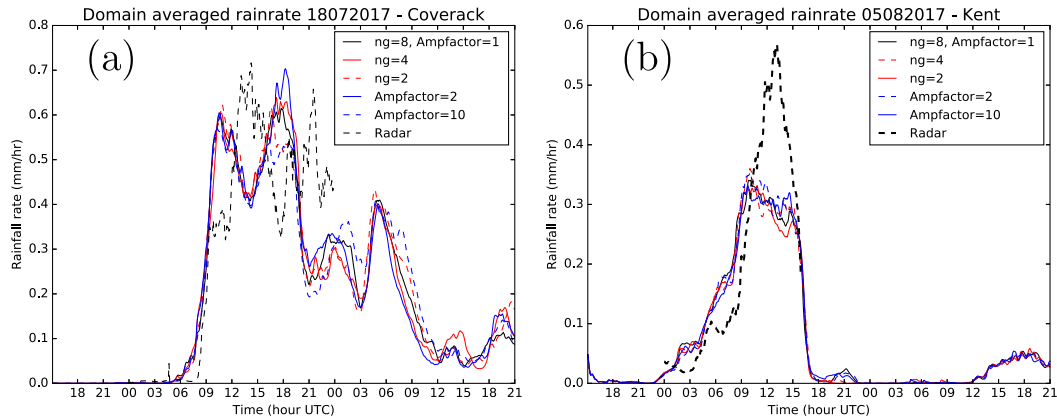


FIG. 10. Average rain rate for radar and model with different values of n_g (as described in section 4a) and amplitude factor (as described in section 7e). The black line is the baseline run ($n_g = 8$, Ampfactor = 1). (a) Coverack and (b) Kent.

in the Coverack and Kent cases, respectively. This shows that using a lower value of n_g does not significantly change the precipitation amount; the rainfall rate using these lower values of n_g actually fall within the range of values from the ensemble member forecasts (not shown here). Instantaneous rain rate plots over the Coverack area at the time when the precipitation was anchored over the village (third row of Fig. 3—and equivalent Fig. 4 for Kent), also show that the main regions of heavy precipitation are very similar in all ensemble members and runs with lower value of n_g although, again, the finer detail is different. Cell statistics (see Figs. 13a and 13b for Coverack and Kent, respectively) show that the distribution of the cell size and intensity are not sensitive to the value of n_g .

The number of common points between the control (no scheme) forecast and those with the scheme but varying values of n_g are largely similar for both cases. Small changes in F_{common} are observed with smaller perturbation areas (see Figs. 11a and 11b for Coverack and Kent, respectively). This is echoed in the mean square difference of the common points (see Figs. 12a and 12b for Coverack and Kent, respectively). This may reflect the idea that it is the variability at the scales

resolvable that matters and the scheme is designed to produce similar variability on this scale even if a small scale is used in the scheme.

e. Sensitivity to amplitude

The magnitude of the perturbations at each time step produced by this scheme are of order 20 times smaller than those from the (A. Lock 2016, personal communication) scheme described in section 5. It was demonstrated in the previous section that changing the seed in the calculation of the perturbations or the area over which the perturbations change did not significantly change the overall forecast compared with a forecast without the scheme (though substantial differences develop at small scales). This section considers the effect of the magnitude of the perturbation on the results. This is achieved by multiplying the second term on the right-hand side of Eq. (48) by a factor (hereafter referred to as Ampfactor) thus increasing the perturbations by that factor. We use Ampfactor values of 2 and 10. These values of Ampfactor are used purely for these sensitivity tests, they do not have any physical basis. This lack of physical basis implies that the results would be less

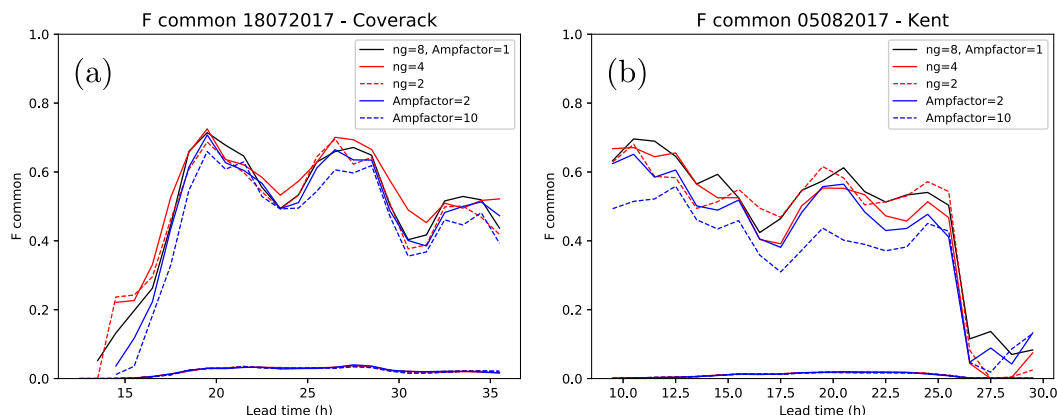


FIG. 11. F_{common} for model with different values of n_g (as described in section 4a) and Ampfactor (as described in section 7e) with associated F chance in the lower lines. The black line is the baseline run ($n_g = 8$, Ampfactor = 1). (a) Coverack and (b) Kent.

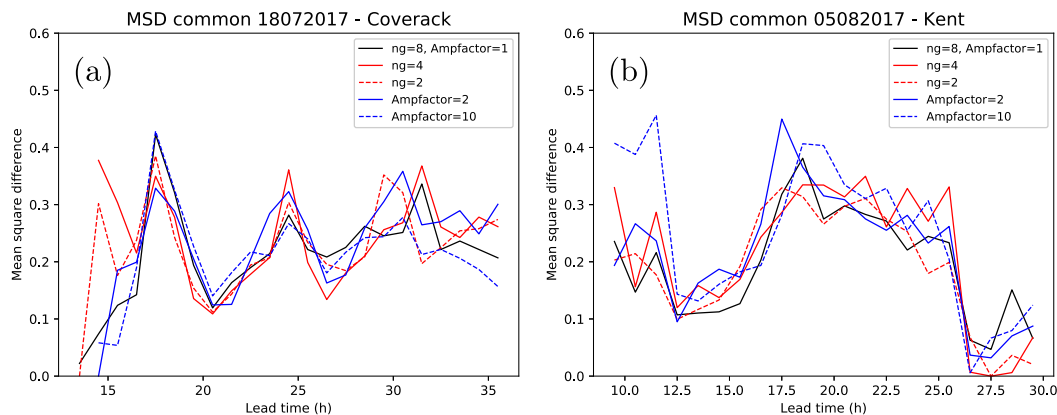


FIG. 12. Ensemble-averaged MSD common for model with different values of n_g (as described in section 4a) and Ampfactor (as described in section 7e). The black line is the baseline run ($n_g = 8$, Ampfactor = 1). (a) Coverack and (b) Kent.

scientifically justified for use within a forecast. Distributions of instantaneous perturbations for the experiments are shown in Fig. 9.

Once again, there are no obvious differences in the larger-scale structure of precipitation fields or distribution of cell size with the larger perturbation. These larger perturbations do not appear to significantly affect the average rain rate (see Fig. 10) and the area of precipitation is largely similar with differences only in the detail (see bottom row in Figs. 3 and 4).

The detail can be studied through the other diagnostics. Differences between Ampfactor 1 and Ampfactor 2 simulations, in comparison with the control and ensemble spread, are minimal. Larger differences occur for multiplying the perturbations by an Ampfactor of 10. There is a consistent decrease in F_{common} , particularly for the Kent case, during the mature stages of the convective events (Fig. 11b). The ensemble-averaged $\text{MSD}_{\text{common}}$ (Fig. 12) shows similar behavior for Ampfactor 1 and 2 simulations, and even the Ampfactor 10 results show a similar spread in the ensemble, though again with some suggestion of more impact in the Kent case. Similarly, the size and magnitude of the cells, shown in Figs 13a and 13b (for Coverack and Kent, respectively), do not change significantly with perturbation size.

8. Conclusions

A physically consistent but simple stochastic boundary layer scheme has been developed that takes a form most naturally expressed as multiplicative noise. The noise is Poisson in character and, consideration of magnitudes shows that the distribution can be highly skewed at convection-permitting scales. The scheme has been derived via a very different route but has been shown to have much in common with the KC scheme, though with the advantage that some parameters are more physically based.

Quantitative investigation of the scheme shows that with horizontal grid lengths around 1 km temporal correlation is far more important than spatial. Above all, however, it is notable that the size of perturbations is very small (overall heating, for

example, generally less than 0.01 K in temperature delivered over a correlation time of around 10 min); at least an order of magnitude smaller than in some more ad hoc schemes. As a result, they have very little impact on the energetics of the forecast, so overall domain-averaged precipitation, for example, is essentially unchanged.

Differences between ensemble members initialized from the control do grow to produce significant differences between ensemble members at the scale of individual convective cells. Thus, they can plausibly account for lack of predictability at this scale. After around 12 h they appear to be roughly saturated. However, it must be emphasized that perturbations are continuously added to the system, so this represents the time scale to achieve a balance between addition of new perturbations, perturbation growth and dissipation, not just saturation of initial perturbations. On these time scales there is no strong evidence for substantial growth of ensemble spread at larger scales. However, it is likely that, in addition to this being limited by the domain size, the cases chosen have a finite lifetime of very active convection.

The results of the scheme are not sensitive to the choice of averaging area the perturbations are chosen to represent, at least where this remains within an order of magnitude of the grid scale. This was the intention behind the design of the scheme and is consistent with the idea that the greater amplitude of perturbations when applied closer to the grid scale do not couple with the dynamics, and only those at well-resolved scales are effective. However, we have not proven this by looking at the initial perturbation growth, as it seems clear that the scheme has impact through the longer-term saturation of the continuously applied perturbations.

Some sensitivity to magnitude in difference diagnostics compared to no scheme is evident, which suggests that the amplitude of perturbations at resolved scales does have some effect, but the sensitivity is not strong. A factor-of-2 increase in magnitude produces results that are within the ensemble spread produced by the scheme but an order of magnitude increase in magnitude does lead to ensemble spread that is somewhat more rapid, and generally marginally more spatial spread when considering the Kent case.

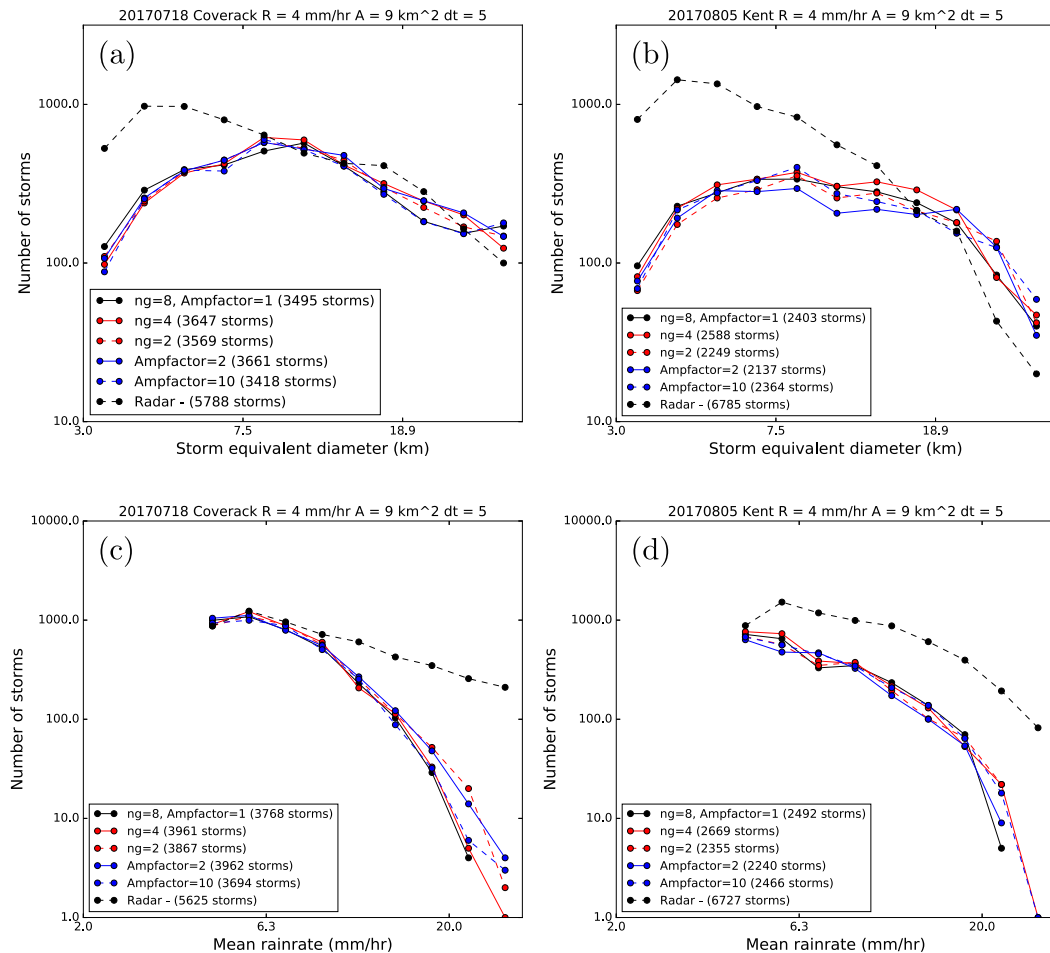


FIG. 13. Cell statistics using a rain-rate threshold of 4 mm h^{-1} for radar and model with different values of n_g (as described in section 4a) and Ampfactor (as described in section 7e). The black line is the baseline run ($n_g = 8$, Ampfactor = 1). (top) Storm equivalent diameters are for (a) Coverack and (b) Kent. (bottom) Mean rain rates are for (c) Coverack and (d) Kent.

The scheme represents only the variability due to boundary layer turbulence in a horizontally homogeneous system. Other sources of boundary layer variability undoubtedly exist, such as small-scale heterogeneity in land surface and soil moisture. More ad hoc schemes may well represent actual boundary layer variability more quantitatively, but at the cost of having less basis in physical processes. Nevertheless, the scheme has been shown to be sufficient to produce substantial differences in forecasts at the scale of convective cells; in Part II of this paper, the differences between ensemble members is compared with that resulting from mesoscale differences in the initial and boundary conditions.

Acknowledgments. The authors acknowledge the use of the MONSooN system, a collaborative facility supplied under the Joint Weather and Climate Research Programme, which is a strategic partnership between the Met Office and the Natural Environment Research Council (NERC). We would also like to acknowledge the three reviewers, two anonymous and George Craig, for their input that has undoubtedly improved the clarity of this paper.

This work has been funded under the work program Towards end-to-end flood forecasting and a tool for real-time catchment susceptibility (TENDERLY) as part of the Flooding From Intense Rainfall (FFIR) project by NERC under Grant NE/K00896X/1. The data used are available by contacting D. Flack at david.flack1@metoffice.gov.uk and are subject to licensing.

REFERENCES

- Arakawa, A., and W. Schubert, 1974: Interaction of a cumulus cloud ensemble with large-scale environment, part I. *J. Atmos. Sci.*, **31**, 674–701, [https://doi.org/10.1175/1520-0469\(1974\)031<0674:IOACCE>2.0.CO;2](https://doi.org/10.1175/1520-0469(1974)031<0674:IOACCE>2.0.CO;2).
- Beare, S., W. Tennant, and C. Sanchez, 2019: Stochastic physics code in the UM. Met Office Documentation Paper 081, 22 pp., https://code.metoffice.gov.uk/doc/um/latest/papers/umdp_081.pdf.
- Carson, D. J., 1973: The development of a dry inversion-capped convectively unstable boundary layer. *Quart. J. Roy. Meteor. Soc.*, **99**, 450–467, <https://doi.org/10.1002/qj.49709942105>.
- Clark, A. J., W. A. Gallus Jr., M. Xue, and F. Kong, 2009: A comparison of precipitation forecast skill between small

- convection-allowing and large convection-parameterizing ensembles. *Wea. Forecasting*, **24**, 1121–1140, <https://doi.org/10.1175/2009WAF2222222.1>.
- Clark, P., N. Roberts, H. Lean, S. P. Ballard, and C. Charlton-Perez, 2016: Convection-permitting models: A step-change in rainfall forecasting. *Meteor. Appl.*, **23**, 165–181, <https://doi.org/10.1002/met.1538>.
- Cornwall Council, 2017: Coverack flood. Accessed March 2018, <http://www.cornwall.gov.uk/coverack>.
- Davies, R., 2017: UK—Emergency rescues after flash floods in southern England. FloodList, accessed March 2018, <http://floodlist.com/europe/united-kingdom/flash-floods-south-england-july-2017>.
- Done, J. M., G. C. Craig, S. L. Gray, P. A. Clark, and M. E. B. Gray, 2006: Mesoscale simulations of organized convection: Importance of convective equilibrium. *Quart. J. Roy. Meteor. Soc.*, **132**, 737–756, <https://doi.org/10.1256/qj.04.84>.
- , —, —, and —, 2012: Case-to-case variability of predictability of deep convection in a mesoscale model. *Quart. J. Roy. Meteor. Soc.*, **138**, 638–648, <https://doi.org/10.1002/qj.943>.
- Ebert, E. E., 2001: Ability of a poor man's ensemble to predict the probability and distribution of precipitation. *Mon. Wea. Rev.*, **129**, 2461–2480, [https://doi.org/10.1175/1520-0493\(2001\)129<2461:AOAPMS>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<2461:AOAPMS>2.0.CO;2).
- Edwards, J., and A. Slingo, 1996: Studies with a flexible new radiation code. I: Choosing a configuration for a large-scale model. *Quart. J. Roy. Meteor. Soc.*, **122**, 689–719, <https://doi.org/10.1002/qj.49712253107>.
- Flack, D. L. A., R. S. Plant, S. L. Gray, H. W. Lean, C. Keil, and G. C. Craig, 2016: Characterisation of convective regimes over the British Isles. *Quart. J. Roy. Meteor. Soc.*, **142**, 1541–1553, <https://doi.org/10.1002/qj.2758>.
- , S. L. Gray, R. S. Plant, H. W. Lean, and G. C. Craig, 2018: Convective-scale perturbation growth across the spectrum of convective regimes. *Mon. Wea. Rev.*, **146**, 387–405, <https://doi.org/10.1175/MWR-D-17-0024.1>.
- , P. A. Clark, C. E. Halliwell, N. M. Roberts, S. L. Gray, R. S. Plant, and H. W. Lean, 2021: A physically based stochastic boundary layer perturbation scheme. Part II: Perturbation growth within a superensemble framework. *J. Atmos. Sci.*, **78**, 747–761, <https://doi.org/10.1175/JAS-D-19-0292.1>.
- Garratt, J. R., 1994: *The Atmospheric Boundary Layer*. Cambridge University Press, 316 pp.
- Hanley, K. E., R. S. Plant, T. H. M. Stein, R. J. Hogan, J. C. Nicol, H. W. Lean, C. Halliwell, and P. A. Clark, 2015: Mixing-length controls on high-resolution simulations of convective storms. *Quart. J. Roy. Meteor. Soc.*, **141**, 272–284, <https://doi.org/10.1002/qj.2356>.
- Hirt, M., S. Rasp, U. Blahak, and G. C. Craig, 2019: Stochastic parameterization of processes leading to convective initiation in kilometer-scale models. *Mon. Wea. Rev.*, **147**, 3917–3934, <https://doi.org/10.1175/MWR-D-19-0060.1>.
- Hohenegger, C., and C. Schär, 2007: Predictability and error growth dynamics in cloud-resolving models. *J. Atmos. Sci.*, **64**, 4467–4478, <https://doi.org/10.1175/2007JAS2143.1>.
- , D. Lüthi, and C. Schär, 2006: Predictability mysteries in cloud-resolving models. *Mon. Wea. Rev.*, **134**, 2095–2107, <https://doi.org/10.1175/MWR3176.1>.
- Holtlag, A. A. M., and B. A. Boville, 1993: Local versus nonlocal boundary-layer diffusion in a global climate model. *J. Climate*, **6**, 1825–1842, [https://doi.org/10.1175/1520-0442\(1993\)006<1825:LNVBLD>2.0.CO;2](https://doi.org/10.1175/1520-0442(1993)006<1825:LNVBLD>2.0.CO;2).
- Johnson, A., and Coauthors, 2014: Multiscale characteristics and evolution of perturbations for warm season convection-allowing precipitation forecasts: Dependence on background flow and method of perturbation. *Mon. Wea. Rev.*, **142**, 1053–1073, <https://doi.org/10.1175/MWR-D-13-00204.1>.
- Kaimal, J. C., J. C. Wyngaard, D. A. Haugen, O. R. Coté, Y. Izumi, S. J. Caughey, and C. J. Readings, 1976: Turbulence structure in the convective boundary layer. *J. Atmos. Sci.*, **33**, 2152–2169, [https://doi.org/10.1175/1520-0469\(1976\)033<2152:TSITCB>2.0.CO;2](https://doi.org/10.1175/1520-0469(1976)033<2152:TSITCB>2.0.CO;2).
- Keil, C., F. Heinlein, and G. Craig, 2014: The convective adjustment times-scale as indicator of predictability of convective precipitation. *Quart. J. Roy. Meteor. Soc.*, **140**, 480–490, <https://doi.org/10.1002/qj.2143>.
- Kober, K., and G. C. Craig, 2016: Physically based stochastic perturbations (PSP) in the boundary layer to represent uncertainty in convective initiation. *J. Atmos. Sci.*, **73**, 2893–2911, <https://doi.org/10.1175/JAS-D-15-0144.1>.
- Kühnlein, C., C. Keil, G. Craig, and C. Gebhardt, 2014: The impact of downscaled initial condition perturbations on convective-scale ensemble forecasts of precipitation. *Quart. J. Roy. Meteor. Soc.*, **140**, 1552–1562, <https://doi.org/10.1002/qj.2238>.
- Lean, P. W., 2006: The predictability of convective storms over the ocean. Ph.D. thesis, University of Reading, 179 pp.
- Leoncini, G., R. Plant, S. Gray, and P. Clark, 2010: Perturbation growth at the convective scale for CSIP IOP 8. *Quart. J. Roy. Meteor. Soc.*, **136**, 653–670, <https://doi.org/10.1002/qj.587>.
- , —, —, and —, 2013: Ensemble forecasts of a flood-producing storm: Comparison of the influence of model-state perturbations and parameter modifications. *Quart. J. Roy. Meteor. Soc.*, **139**, 198–211, <https://doi.org/10.1002/qj.1951>.
- Lock, A. P., A. R. Brown, M. R. Bush, G. M. Martin, and R. N. B. Smith, 2000: A new boundary layer mixing scheme. Part I: Scheme description and single-column model tests. *Mon. Wea. Rev.*, **128**, 3187–3199, [https://doi.org/10.1175/1520-0493\(2000\)128<3187:ANBLMS>2.0.CO;2](https://doi.org/10.1175/1520-0493(2000)128<3187:ANBLMS>2.0.CO;2).
- Lorenz, E. N., 1969: The predictability of a flow which possesses many scales of motion. *Tellus*, **21**, 289–307, <https://doi.org/10.3402/tellusa.v21i3.10086>.
- Lumley, L., 1967: The structure of inhomogeneous turbulent flows. *Atmospheric Turbulence and Radio Wave Propagation*, A. M. Yaglom and V. I. Tatarsky, Eds., Nauka, 166–176.
- Mason, P. J., 1988: The formation of areally-averaged roughness lengths. *Quart. J. Roy. Meteor. Soc.*, **114**, 399–420, <https://doi.org/10.1002/qj.49711448007>.
- Melhauser, C., and F. Zhang, 2012: Practical and intrinsic predictability of severe and convective weather at the meso-scales. *J. Atmos. Sci.*, **69**, 3350–3371, <https://doi.org/10.1175/JAS-D-11-0315.1>.
- Moin, P., and R. D. Moser, 1989: Characteristic-eddy decomposition of turbulence in a channel. *J. Fluid Mech.*, **200**, 471–509, <https://doi.org/10.1017/S0022112089000741>.
- Panofsky, H. A., H. Tennekes, D. H. Lenschow, and J. C. Wyngaard, 1977: The characteristics of turbulent velocity components in the surface layer under convective conditions. *Bound.-Layer Meteor.*, **11**, 355–361, <https://doi.org/10.1007/BF02186086>.
- Plant, R., and G. Craig, 2008: A stochastic parameterization for deep convection based on equilibrium statistics. *J. Atmos. Sci.*, **65**, 87–105, <https://doi.org/10.1175/2007JAS2263.1>.
- Porson, A., P. A. Clark, I. N. Harman, M. J. Best, and S. E. Belcher, 2010: Implementation of a new urban energy budget scheme in the MetUM. Part I: Description and idealized simulations. *Quart. J. Roy. Meteor. Soc.*, **136**, 1514–1529, <https://doi.org/10.1002/qj.668>.

- Rasp, S., T. Selz, and G. C. Craig, 2018: Variability and clustering of midlatitude summertime convection: Testing the Craig and Cohen theory in a convection-permitting ensemble with stochastic boundary layer perturbations. *J. Atmos. Sci.*, **75**, 691–706, <https://doi.org/10.1175/JAS-D-17-0258.1>.
- Raynaud, L., and F. Bouttier, 2016: Comparison of initial perturbation methods for ensemble prediction at convective scale. *Quart. J. Roy. Meteor. Soc.*, **142**, 854–866, <https://doi.org/10.1002/qj.2686>.
- Skamarock, W. C., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp., <https://doi.org/10.5065/D68S4MVH>.
- , S.-H. Park, J. B. Klemp, and C. Snyder, 2014: Atmospheric kinetic energy spectra from global high-resolution nonhydrostatic simulations. *J. Atmos. Sci.*, **71**, 4369–4381, <https://doi.org/10.1175/JAS-D-14-0114.1>.
- Wilson, D. R., and S. P. Ballard, 1999: A microphysically based precipitation scheme for the UK Meteorological Office Unified Model. *Quart. J. Roy. Meteor. Soc.*, **125**, 1607–1636, <https://doi.org/10.1002/qj.49712555707>.
- Wood, N., and Coauthors, 2014: An inherently mass-conserving semi-implicit semi-Lagrangian discretisation of the deep-atmosphere global nonhydrostatic equations. *Quart. J. Roy. Meteor. Soc.*, **140**, 1505–1520, <https://doi.org/10.1002/qj.2235>.
- Zhang, F., C. Snyder, and R. Rotunno, 2003: Effects of moist convection on mesoscale predictability. *J. Atmos. Sci.*, **60**, 1173–1185, [https://doi.org/10.1175/1520-0469\(2003\)060<1173:EOMCOM>2.0.CO;2](https://doi.org/10.1175/1520-0469(2003)060<1173:EOMCOM>2.0.CO;2).