

# *A physically-based stochastic boundary-layer perturbation scheme. Part II: perturbation growth within a super ensemble framework*

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open Access

Flack, D., Clark, P. ORCID: <https://orcid.org/0000-0003-1001-9226>, Halliwell, C., Roberts, N., Gray, S. ORCID: <https://orcid.org/0000-0001-8658-362X>, Plant, B. ORCID: <https://orcid.org/0000-0001-8808-0022> and Lean, H. (2021) A physically-based stochastic boundary-layer perturbation scheme. Part II: perturbation growth within a super ensemble framework. *Journal of the Atmospheric Sciences*, 78 (3). pp. 747-761. ISSN 1520-0469 doi: <https://doi.org/10.1175/JAS-D-19-0292.1> Available at <https://centaur.reading.ac.uk/93961/>

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-19-0292.1>

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](http://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 II: Perturbation Growth within a Superensemble Framework

D. L. A. FLACK,<sup>a,c</sup> P. A. CLARK,<sup>a</sup> C. E. HALLIWELL,<sup>b</sup> N. M. ROBERTS,<sup>b</sup> S. L. GRAY,<sup>a</sup>  
R. S. PLANT,<sup>a</sup> AND H. W. LEAN<sup>b</sup>

<sup>a</sup> *Department of Meteorology, University of Reading, Reading, United Kingdom*

<sup>b</sup> *MetOffice@Reading, Department of Meteorology, University of Reading, United Kingdom*

(Manuscript received 17 October 2019, in final form 23 October 2020)

**ABSTRACT:** Convection-permitting forecasts have improved the forecasts of flooding from intense rainfall. However, probabilistic forecasts, generally based upon ensemble methods, are essential to quantify forecast uncertainty. This leads to a need to understand how different aspects of the model system affect forecast behavior. We compare the uncertainty due to initial and boundary condition (IBC) perturbations and boundary layer turbulence using a superensemble (SE) created to determine the influence of 12 IBC perturbations versus 12 stochastic boundary layer (SBL) perturbations constructed using a physically based SBL scheme. We consider two mesoscale extreme precipitation events. For each, we run a 144-member SE. The SEs are analyzed to consider the growth of differences between the simulations, and the spatial structure and scales of those differences. The SBL perturbations rapidly spin up, typically within 12 h of precipitation commencing. The SBL perturbations eventually produce spread that is not statistically different from the spread produced by the IBC perturbations, though in one case there is initially increased spread from the IBC perturbations. Spatially, the growth from IBC occurs on larger scales than that produced by the SBL perturbations (typically by an order of magnitude). However, analysis across multiple scales shows that the SBL scheme produces a random relocation of precipitation up to the scale at which the ensemble members agree with each other. This implies that statistical postprocessing can be used instead of running larger ensembles. Use of these statistical postprocessing techniques could lead to more reliable probabilistic forecasts of convective events and their associated hazards.

**KEYWORDS:** Boundary layer; Ensembles; Numerical weather prediction/forecasting; Probabilistic Quantitative Precipitation Forecasting (POPF); Parameterization; Stochastic models


### 1. Introduction

Forecasting of convective events has had a “step change” in ability since the advent of convection-permitting models (e.g., Lean et al. 2008; Clark et al. 2016). In turn, this has led to improvements in the prediction of floods with a rapid rate of rise, i.e., both surface water and flash flooding (e.g., Roberts et al. 2009; Cuo et al. 2011). However, quantitative forecasting of convective precipitation still remains a key challenge due to uncertainty in spatial structure (e.g., Roberts and Lean 2008; Dey et al. 2014, 2016a; Flack et al. 2018), timing (e.g., Lean et al. 2008), storm structure (e.g., Stein et al. 2015) and intensity (e.g., Mittermaier 2014); these issues are covered in more detail by Clark et al. (2016).

Convection-permitting forecasts lead to improved forecasts of convective events (e.g., Clark et al. 2009) but the smaller scales represented have, in general, faster error growth than the larger scales represented in coarser-resolution systems (e.g., Hohenegger et al. 2006; Hohenegger and Schär 2007; Clark et al. 2010). While errors growing faster at smaller scales in the atmosphere is not a surprising result (e.g., Lorenz 1969),

the implication is that for most forecast lead times a probabilistic approach is required.

To help represent this uncertainty many operational centers use ensemble prediction systems (hereafter ensembles) at convection-permitting resolution (e.g., Seity et al. 2011; Baldauf et al. 2011; Hagelin et al. 2017) to indicate the range of plausible outcomes from subtle changes in initial conditions, boundary conditions and model physics (e.g., Buizza and Palmer 1995). However, there are still questions concerning error growth within ensembles, and as such convective-scale predictability (e.g., Zhang et al. 2003; Selz and Craig 2015; Johnson and Wang 2016). These questions need to be answered to allow for the effective design and implementation of convective-scale ensembles. While error growth is overall faster at these scales there are differences in the error growth that depend on the environmental flow, such as the presence or lack of a diurnal cycle (e.g., Nielsen and Schmachter 2016), and the scales at which the dominant growth occurs (e.g., Roberts 2008; Johnson et al. 2014; Flack et al. 2018). These factors need to be considered carefully in ensemble design to allow a reliable ensemble to be made, as they indicate that perturbations need to be made across a range of scales. Here we compare ensembles created by two different types of perturbations in the context of both magnitude and spatial aspects of perturbation growth.

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

<sup>c</sup> 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/>).

Recent work examining convective-scale error growth has considered the spatial aspects of the growth for a range of cases (e.g., Johnson et al. 2014; Surcel et al. 2016). Generally these studies indicate that more widespread precipitation results in a greater areal extent of error growth than more localized precipitation. However, more localized precipitation is less predictable compared to larger areas of precipitation (e.g., Roberts 2008). There are also other factors that determine the spatial aspects of error growth. For example, Flack et al. (2018) indicated the scales at which the error growth was dominating were partly linked to the large-scale synoptic forcing. Indeed, for their experiments it was shown that cases with weaker synoptic forcing had perturbation growth that dominated on scales  $O(1)$  km whereas for cases that were strongly forced there was an order of magnitude difference, so growth dominated on scales  $O(10)$  km.

Many more studies have considered the magnitude of error growth across multiple cases (e.g., Done et al. 2006; Keil and Craig 2011; Done et al. 2012). These studies showed that the total (area-averaged) precipitation had reduced spread between ensemble members in strong synoptically forced compared to weakly forced cases. These results were then developed by Keil et al. (2014) and Kühnlein et al. (2014) to consider the response of convection to different perturbation strategies. It was indicated that model physics perturbations had a greater influence on the total precipitation spread in weakly forced cases compared to strongly forced conditions, particularly around the initiation time of events, in agreement with Surcel et al. (2017). This agrees with previous studies considering convective cases that found that model physics perturbations have their greatest impact at convective initiation (e.g., Zhang et al. 2003; Hohenegger et al. 2006; Leoncini et al. 2010).

Intrinsic predictability experiments yield the theoretical lowest amount of uncertainty possible for an event whereas practical predictability experiments yield the uncertainty in models for actual cases (based on current capabilities). In a forecasting context, both intrinsic predictability experiments and practical predictability experiments have their uses for forecast interpretation. Generally studies [including most previously discussed, with the exception of Keil et al. (2014) and Kühnlein et al. (2014)] have focused on intrinsic rather than practical predictability. However, there are now more studies considering practical predictability (e.g., Melhauser and Zhang 2012; Sun and Zhang 2016). Both of these studies considered the up/downscale growth of perturbations and show that if the errors on large scales (of roughly 1000 km) are large then the forecasts can be improved via more accurate initial conditions, whereas if the errors on the large scale are small then, regardless of improvements in initial conditions, there will be limited improvement in the forecasts on the mesoscale. This result was also found by Durran and Gingrich (2014) and Weyn and Durran (2017), though the latter study notes that there is no upscale/downscale growth within their idealized simulations and the errors grow up-amplitude on all scales simultaneously. These discrepancies show that further work needs to go into these practical predictability experiments as this

will help indicate where forecasts can be improved further, for example through better specification of initial conditions or better representation of unresolved processes such as turbulent eddies.

In Clark et al. (2021, hereafter Part I) we discussed the formulation of our physically based stochastic boundary layer (SBL) perturbation scheme and tested it for two distinct cases (18 July and 5 August 2017) over the United Kingdom. Our physically based stochastic scheme is designed to represent the sampling error from unresolved turbulent eddies within the boundary layer. It depends upon the average number of thermals triggered over an area in a set time and is such that situations with, on average, more thermals result in relatively smaller stochastic increments. Testing showed that the scheme does not result in any significant systematic change in overall precipitation, but generates significant differences from a control simulation at the convective cell scale over a forecast of several hours and so can form the basis of an ensemble designed to represent the impact of this form of uncertainty. The stochastic scheme is designed to be relatively insensitive to the spatial scale the perturbations are applied on, and testing confirmed this; some sensitivity to the magnitude of the perturbations was observed, though a factor-of-10 increase was required to produce significantly more displacement in the convective precipitation from the control simulations.

The magnitude of stochastic increments appears very small (around 0.01 K), but this is because the boundary layer heating is similarly small on the same time scale. In fact, at the scales applied, the variability of increments can easily match the size of the mean. As discussed in Part I in more depth, this irreducible variability must exist in even the most idealized smoothly forced circumstances, and one of our objectives is to determine how significant this source of variability is. Other sources of uncertainty exist, including uncertainty in surface parameters, and so-called structural uncertainties due to the inaccuracy of the parameterization scheme. The former depends on knowledge of surface characteristics (or lack thereof) and is difficult to model universally. For example, the “uncertainty” in evapotranspiration would be larger in a model using climatological values of, for example, leaf area index compared with that using a measured value from satellite-based remote sensing. Clearly, the objective with such uncertainty is to reduce it using more or better measurements (though again there is likely to be an irreducible limit to be determined). “Structural” uncertainty is not a well enough defined concept, but we take it to mean that the ensemble mean response to forcing is likely to be in error. Such errors tend to be systematic, often leading to different quasi-equilibrium profiles for given forcing, and it is very hard to argue that the representation of such errors should be stochastic on small scales without introducing the physical reasoning behind our scheme. Our scheme represents the variability about the ensemble mean, which increases as the space and time averaging scale decreases. Of course, the mean is zero so the question is how much of the variability is retained and grows. We therefore would argue that the variability represented by our scheme must be considered

at high resolution, and in this paper we do so cleanly, comparing its effect with that of a well-defined and separate source of uncertainty.

Thus, here in Part II of this study, we wish to determine the impact of the SBL compared to initial and boundary condition (IBC) perturbations on forecast uncertainty and so determine the spatial scales at which these perturbations act. We consider the perturbation growth in a superensemble (SE) framework using the same two cases in practical predictability experiments. An SE is a large ensemble that consists of several subensembles in which different types of perturbations are used. This is a useful but expensive tool. This expense arises from the need to consider a large number of ensemble members, either  $n^m$  or (if each factor has a different ensemble size)<sup>1</sup>  $n_0 \times n_1 \times \dots \times n_m$ , where  $n$  represents the ensemble size and  $m$  represents the number of factors being considered (i.e., for our situation  $m = 2$  to compare the influence of IBC perturbations and SBL perturbations) to be able to determine the impact of each factor.

The SE framework is a simple and effective method for determining the (relative) impact of different sources of uncertainty upon the forecast (e.g., Kühnlein et al. 2014; Keil et al. 2014). Since the SBL perturbations are small scale, we wish to address a second question. Practical ensembles do not contain enough members to enable probabilities of, for example, precipitation to be derived simply and directly. Some postprocessing is needed to smooth the predicted probabilities, often based on “neighborhood” methods (as discussed in section 4d). The SE provides us with a tool to compare both the scales of variability due to the SBL with that assumed in the neighborhood processing, and the predicted rainfall probabilities. If the postprocessed ensemble is similar to the full SE it implies that the postprocessing acts to artificially increase the ensemble size thus saving the computational expense of running an SE operationally, particularly at the convective scale. While this paper acts to test our scheme in an operational context, the questions considered in the manuscript apply more widely to all forms of SBL perturbations.

Thus, through our SE we consider two questions:

- (i) How does the perturbation growth induced by our SBL compare to the growth from IBC perturbations, and
- (ii) how does the impact of our SBL scheme compare to that from postprocessing an ensemble without our SBL scheme using neighborhood-based diagnostics?

The remainder of this paper is set out as follows. The construction of the SE is discussed in section 2, a brief overview of the cases is given in section 3 and diagnostics considered here are explained in section 4, with a particular emphasis on those not used in Part I. The magnitude of the perturbation growth is considered in section 5 and the

spatial aspects are considered in section 6; finally, conclusions are drawn in section 7.

## 2. The superensemble

Here we have taken an operational convection-permitting ensemble and expanded it into a much larger ensemble using the perturbations from our SBL scheme. We have termed this larger ensemble an SE as it is one large ensemble made of many subensembles. The SE (Fig. 1) is constructed using the Met Office Unified Model (MetUM) at version 10.6. The MetUM is a nonhydrostatic, semi-implicit, semi-Lagrangian model that uses the Even Newer Dynamics for General Atmospheric Modeling of the Environment (ENDGAME) formulation for its dynamical core (Wood et al. 2014). We use the standard MetUM parameterizations for the boundary layer (Lock et al. 2000), microphysics (Wilson and Ballard 1999), radiation (Edwards and Slingo 1996) and surface-layer scheme (Porson et al. 2010). A convection scheme is not used as convection is treated explicitly.

The SE is constructed from 12 members of the operational Met Office Global and Regional Ensemble Prediction System for the United Kingdom (MOGREPS-U.K.; Hagelin et al. 2017). The MOGREPS-U.K. configuration of the MetUM is a 2.2 km grid-length ensemble. It is closely connected to the U.K. variable resolution (UKV) configuration of the MetUM operational at the time of the case studies (except the UKV has a 1.5 km grid length over the United Kingdom). This configuration uses 4DVAR data assimilation to produce an analysis every 3 h; in practice analysis increments are “nudged” into a forecast started from the 1 h forecast from the previous analysis. MOGREPS-U.K. follows a similar process, starting with the same UKV 1 h forecast and analysis increments reconfigured to the 2.2 km grid, but each ensemble member also has added the downscaled perturbations for IBCs from the 33 km grid-length global ensemble (MOGREPS-G; Bowler et al. 2008; Tennant et al. 2011). The intention is thus to retain both the high-resolution information from the UKV analysis and the mesoscale perturbations from MOGREPS-G. This setup is identical to the setup described by Hagelin et al. (2017) except that the UKV analysis increments have since been updated to 4DVAR analysis increments instead of 3DVAR analysis increments. Each of the 12 MOGREPS-U.K. members forms the basis of a 12-member subensemble by generating a further 11 members with our SBL scheme using 11 different random seeds. This process results in a set of 12 subensembles each with 12 members, and thus an SE with 144 members. The IBC perturbed components of the SE are those generated by the operational MOGREPS-U.K. system.

In our experiments, unlike in the operational version of MOGREPS-U.K., we do not use the operational stochastic potential temperature ( $\theta$ ) perturbations or the random parameter scheme to produce model physics perturbations (discussed in McCabe et al. 2016; Hagelin et al. 2017). We run the SBL scheme discussed in Part I instead.

Our SBL scheme is designed to represent the variation due to unresolved turbulent processes that is not accounted for in traditional boundary layer schemes. In the SE the SBL scheme is set up to perturb  $\theta$ ,  $q$ ,  $u$ , and  $v$  (where  $q$ ,  $u$ , and  $v$  represent

<sup>1</sup> This generalization of the SE size leads to greater ambiguity than the former when defining the size of each ensemble member and could, perhaps, imply that a differing size ensemble for one factor leads to more weight for that metric than another.

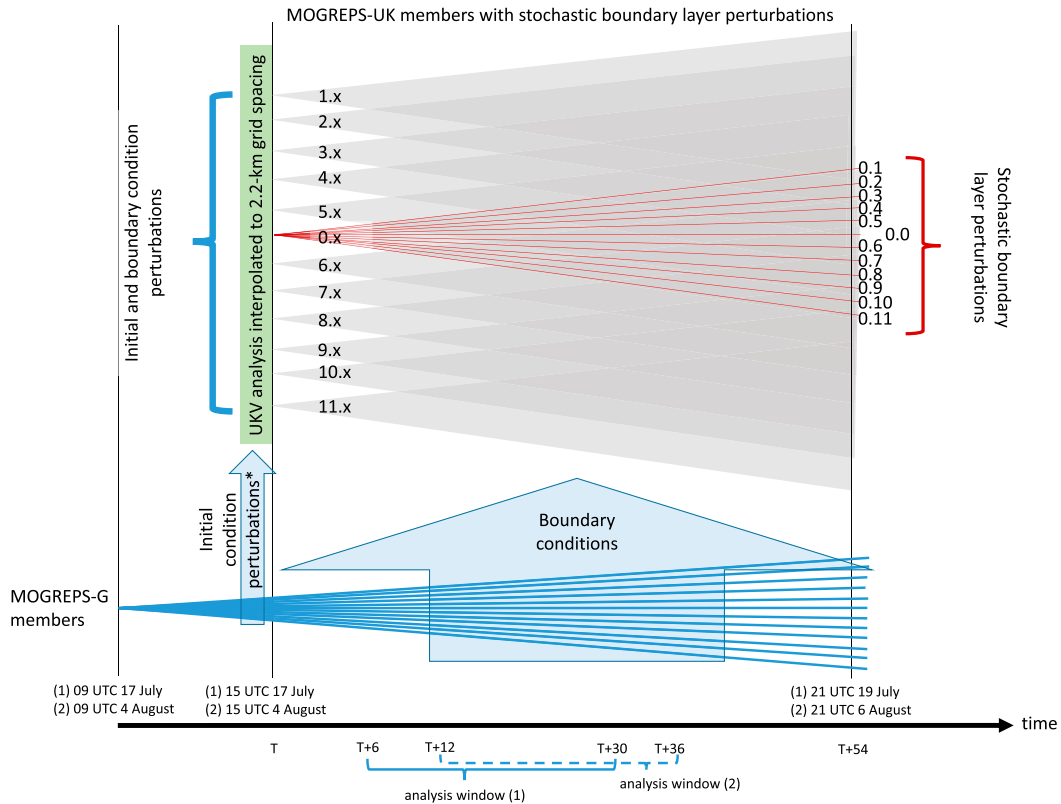


FIG. 1. A schematic of the superensemble, showing where all<sup>2</sup> the perturbations come from and its relationship to the operational MOGREPS-G and MOGREPS-U.K. It also includes the labels for the superensemble members which have the form  $a.x$ , where  $a$  refers to the IBC member and  $x$  refers to the stochastic member.

specific humidity, and wind components in the zonal and meridional directions, respectively) over a region of  $8 \times 8$  grid boxes that is repeated in a “checkerboard” effect across the domain. The magnitudes of the perturbations are set to a value that is physically appropriate based on boundary layer scalings and is not multiplied by an extra factor. On this eddy turnover time scale the scheme adds perturbations with standard deviation roughly  $(A_i/\Delta A)^{1/2}\theta_*$ , where  $A_i$  is the area occupied by one eddy,  $\Delta A$  is the averaging area, and  $\theta_*$  is the free convective temperature scale, typically of order 0.1 K; with an  $8 \times 8$  checkerboard and 2.2 km grid box, so  $A_i/\Delta A \approx 1/16$ , this is roughly 0.1/16 K. Thus, perturbations are very small. They could have been applied over a smaller area, and thus been larger, but the results of Part I suggest the results would not be significantly different. The scheme is applied on all model levels that the boundary layer scheme runs on and at every time step throughout the run. In practice this means all model levels, but the perturbations outside the actual boundary layer are generally much smaller. Full justification for these choices and sensitivity of the scheme is discussed in Part I.

In the SE experiments the two cases considered are initiated at 1500 UTC the day prior to the event of interest (17 July and

4 August 2017, respectively). This allows the event of interest to occur at a time in the forecast (approximately  $T + 24$  h) when all the perturbations have had time to grow to produce a similar influence on the forecast, as we will demonstrate in section 5.

Throughout the rest of the paper the following notation (used within Fig. 1) is used to describe the different ensemble members within the SE:  $a.x$  where  $a$  refers to the IBC member and  $x$  refers to the stochastic member. Thus member 0.0 of the SE is the control of MOGREPS-U.K. and there are no stochastic perturbations added (i.e., it is the unperturbed control member of the entire SE). Furthermore, we define two types of subensembles (IBC and SBL subensembles) e.g., 1.x refers to the subensemble with IBC member 1 with all 12 stochastic members (i.e., an IBC subensemble) whereas  $a.1$  is the subensemble with stochastic member 1 with all 12 different IBCs (i.e., a SBL subensemble). We also refer to subensemble  $a.0$  as the control subensemble (the ensemble with no stochastic perturbations), which is our equivalent to MOGREPS-U.K. without any stochastic perturbations.

### 3. Case studies

As discussed in the introduction we use the same cases here as we did in Part I; however, we provide a brief overview of the cases here to set the scene and the terminology around the cases. Both cases are given names based primarily after the

<sup>2</sup> Perturbations from the UKV analysis are added to the initial conditions from the MOGREPS-G members into MOGREPS-U.K. but are not included in the plot for simplicity.

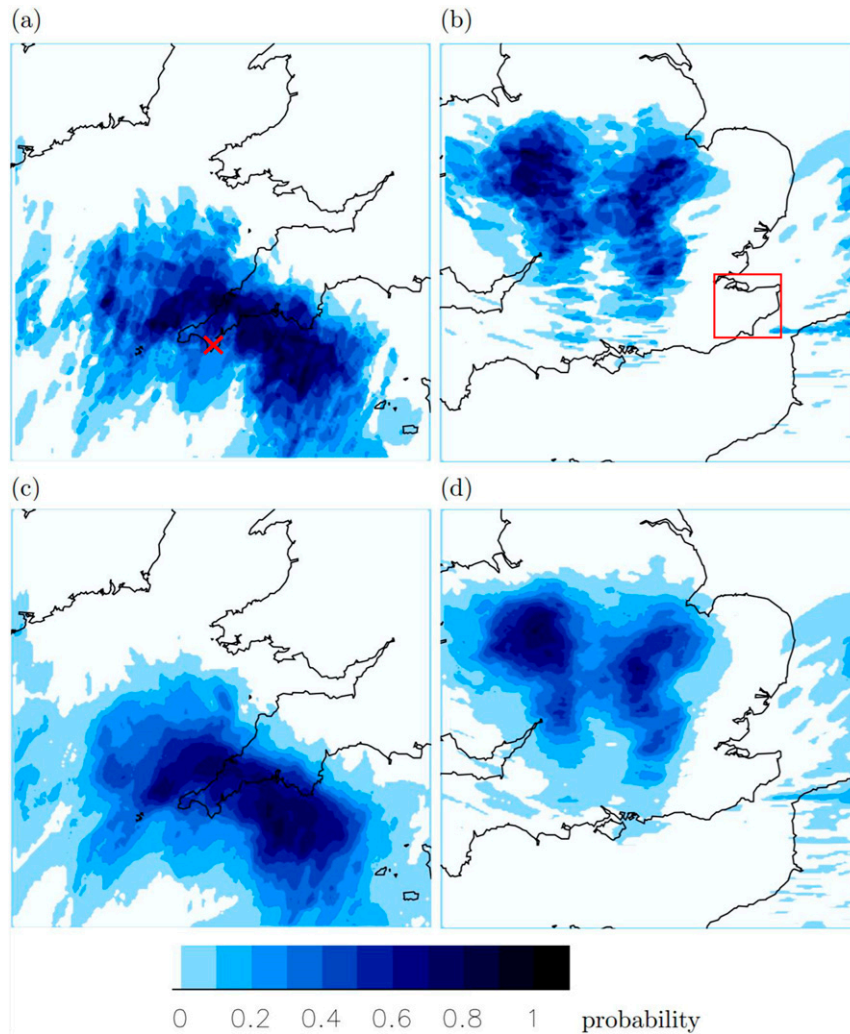


FIG. 2. An example of the precipitation probabilities for exceeding an hourly precipitation accumulation of 1 mm in (a),(b) the control subensemble (*a.0*) and (c),(d) the full SE from 1400 to 1500 UTC for (a),(c) the Coverack case and 1300 to 1400 UTC for (b),(d) the Kent case. The domains plotted correspond to the analysis domains used and are substantially smaller than the MORGREPS-U.K. domain used for the model runs. The red cross in (a) indicates Coverack, and the red box in (b) indicates the county of Kent.

locations where the convection was observed to be most intense or dynamically active, rather than from the analysis domains. Figure 2 shows the probability of reaching an hourly-precipitation accumulation of at least 1 mm for these events generated from the control subensemble (*a.0*) and the entire SE, for both the Coverack case (Figs. 2a,c) and the Kent case (Figs. 2b,d). The cases were chosen to show different types of convection, and via the convective adjustment time scale (e.g., Done et al. 2006) can be shown to occur within different places along the spectrum of convective regimes (e.g., Flack et al. 2018).

#### a. Coverack case: 18 July 2017

In this case a mesoscale convective system (MCS) progressed toward the United Kingdom after forming off the coast of Brittany at 1200 UTC. The MCS moved over Cornwall at

1400 UTC bringing intense precipitation that resulted in a devastating flood for the village of Coverack (Essex 2018) as part of the MCS became anchored over Coverack for approximately 3 h from 1400 UTC. The convective adjustment time scale for this case is initially 4.2 h and over time reduces to 0.4 h. Combining this with the local forcing keeping the storm anchored places this case toward the nonequilibrium end of the spectrum (despite the marginal time scale).

#### b. Kent case: 5 August 2017

The second case began as scattered showers forming in the lee of the Welsh mountains before aggregating as they traveled across England. Upon reaching east England (East Anglia) at 1400 UTC, they had formed into two S–N oriented squall lines (see Figs. 2b,d). The eastern squall line then moved along the

north Kent coast (not shown). By chance, the lead author was there at the time and from 1502 to 1534 UTC he witnessed multiple mesocyclones and three funnel clouds as the squall line passed directly overhead. The rainfall associated with the southern squall line was intense and could have led to flooding had it been further south over land. However, most of the precipitation occurred along the coast either onto marshland or into the sea. This case has a low convective adjustment time scale of initially 1.1 h dropping to 0.1 h and so is placed on the convective quasi-equilibrium end of the spectrum of convective regimes.

#### 4. Diagnostics

Three diagnostics were utilized in this study and are now described. Alongside the mean square difference (MSD) previously discussed in Part I and defined in Flack et al. (2018), a variance diagnostic and a diagnostic that considers the spatial aspects of the forecasts is also used: the temperature variance and the ensemble agreement scale (EAS; Dey et al. 2016a,b). All analysis using the MSD is performed over a region of  $205 \times 205$  grid boxes ( $451 \text{ km} \times 451 \text{ km}$ ) which includes the formation locations for each event. The temperature variance is calculated for the full forecasts and the interior domain (2.2 km) of MOGREPS-U.K., while the EAS is calculated across the entire domain, but shown over the same analysis domain as the MSD. Figure 2 indicates the analysis domains for each case, which are identical to those used in Part I. The diagnostics are considered for both forecasts at times specific to the life cycle of the event across the full SE including formation and decay or leaving the United Kingdom. These times are  $T + 12 \text{ h}$  to  $T + 36 \text{ h}$  for the Coverack case and  $T + 6 \text{ h}$  to  $T + 30 \text{ h}$  for the Kent case (Fig. 1). They are further chosen to allow at least 1% of points within the domain to have precipitation as otherwise it becomes difficult to separate numerical artifacts due to the small number of points from physical differences.

##### a. Mean square difference

The MSD was used and discussed in detail in Part I in the testing of our SBL scheme. We repeat the formula here for convenience:

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

for  $P_c$  the hourly precipitation accumulations in the control forecast and  $P_p$  the hourly-precipitation accumulations in the perturbed forecast, evaluated at each grid square within the analysis domain. As in Part I the MSD is considered only over the common points, so is referred to as  $\text{MSD}_{\text{common}}$ . The “common” points are those at which precipitation occurs in the same location in both the control and perturbed forecasts. The  $\text{MSD}_{\text{common}}$  is used to help diminish the “double penalty” problem as it neglects points where precipitation only occurs in one forecast. An arbitrary hourly precipitation accumulation threshold for the identification of convective precipitation is used here for  $\text{MSD}_{\text{common}}$ . This threshold is set at 1.0 mm to keep consistency with Part I

although the conclusions are insensitive to reasonable changes in this value (not shown).

For the calculation of the MSD the ensembles have been bootstrapped with replacement for 10 000 samples to produce confidence intervals on the mean, and reliable 95th and 5th percentiles. Furthermore, times during this analysis period with a low number of precipitating points are separated by vertical dot-dashed lines on the figures. Times before that, indicated by a line near the start of the analysis period, or after that, indicated by a line near the end of the analysis period, are less statistically reliable, and hence conclusions are not drawn from these periods.

##### b. Temperature variance

The temperature variance has been chosen as a diagnostic because it is one of the components of the difference total energy (DTE), which consists of the difference kinetic energy and a thermal component (e.g., Zhang et al. 2003), and is frequently used to consider error growth (e.g., Zhang et al. 2003; Selz and Craig 2015; Durran and Gingrich 2014). Here, we consider the evolution of the temperature variance ( $\text{DTE}_T$ ) at 850 hPa averaged across the interior domain. The temperature variance is given by

$$\text{DTE}_T = \frac{c_p}{T_{\text{ref}}} T' T',$$

where  $c_p$  is the specific heat capacity,  $T_{\text{ref}}$  is a reference temperature, here taken to be 273 K, and  $T'$  denotes the difference between the control (0.0) and the perturbed forecasts ( $ax$ ). This diagnostic allows a direct inference of the size of the temperature perturbations to indicate the impact of the spatial scale of different size perturbations, and the behavior can be used to help infer the different growth mechanisms.

##### c. Ensemble agreement scale

At small scales ensemble members are more likely to be in disagreement with each other and observations because of differences in positioning and intensity which means low predictability and low skill for any given member. A lack of predictability at small scales will also lead to noisy (spatially fragmented) probability forecasts from an ensemble unless there are either sufficient ensemble members to account for the uncertainty or neighborhood processing is used to effectively add members. At larger scales there is typically more agreement so a “skillful scale” can be defined as the smallest scale at which the members are in agreement. Here we use the EAS defined by Dey et al. (2016a) to determine a scale for each individual grid point that can be used to establish appropriate neighborhood sizes to be used when generating probabilities from the ensembles.

The calculation of the EAS starts by comparing pairs of fields and is applied to each grid point. First, for each grid point, a comparison is made with its equivalent and then successively larger square neighborhoods are tested until a neighborhood size is found in which the precipitation forecasts are found to have sufficient agreement with one another (they “suitably” agree) as defined by Eqs. (1) and (2) below. Usually,



the overall EAS for each grid point  $(i, j)$  is defined as the average agreement scale between all member-member pairs at that grid point (Dey et al. 2016a,b). However, given the size of the SE we restrict this to the average agreement scale between the control and perturbed member pairs.

The agreement scale,  $S_{ij}^A$ , for each control-perturbed member pair, at each grid point, is the neighborhood width for which the two forecasts are in agreement. If two forecast agree at the scale  $n \times n$  grid points, the agreement scale is  $(n - 1)/2$  grid lengths. For example, if the forecasts agree at the grid scale, they have an agreement scale of 0 whereas if a neighborhood of size  $3 \times 3$  grid boxes is required, they have an agreement scale of 1 and for  $5 \times 5$  grid boxes the agreement scale is 2, etc. The EAS is therefore defined as the minimum scale  $S$  in terms of number of complete grid boxes, that satisfies

$$D_{ij}^S \leq D_{crit,ij}^S, \tag{1}$$

where  $D_{ij}^S$  is the normalized ratio of the squared differences between the fields and  $D_{crit,ij}^S$  is the critical value which determines if the values suitably agree with one another. For hourly precipitation accumulations for the control-perturbed member pairs,  $D_{ij}^S$  is calculated as

$$D_{ij}^S = \begin{cases} \frac{(P_{p,ij}^S - P_{c,ij}^S)^2}{(P_{p,ij}^S)^2 + (P_{c,ij}^S)^2}, & \text{if } P_{p,ij}^S > 0 \text{ or } P_{c,ij}^S > 0, \\ 1, & \text{if } P_{p,ij}^S = 0 \text{ and } P_{c,ij}^S = 0, \end{cases} \tag{2}$$

where  $P_{p,ij}^S$  and  $P_{c,ij}^S$  represent the precipitation within the neighborhood of width  $S$  centered on grid box  $(i, j)$  for the control ( $c$ ) and perturbed ( $p$ ) forecasts. The critical value is determined by

$$D_{crit,ij}^S = \alpha + (1 - \alpha) \frac{S}{S_{lim}},$$

where  $\alpha$  is a bias tolerance between the forecasts varying between zero and unity (here it is set to 0.5) and  $S_{lim}$  is a pre-determined fixed maximum scale where (1) will always be satisfied [here it is set to 80 as in Dey et al. (2016a)]. Recall that for forecasts that agree at the grid scale  $S$  is zero and so  $D_{ij}^S = \alpha$ . The assessment of agreement is performed iteratively, incrementing  $S$  from the initial value of zero until an agreement scale, or  $S_{lim}$ , is reached.

Given that (1) is for the minimum when the criterion is met the EAS will range from zero (an acceptably spatially identical forecast) to  $S_{lim}$  which either implies that there is only precipitation in one forecast over the area corresponding to  $S_{lim}$ , or that there is no precipitation in either forecast over the area, or that there is no spatial agreement between forecasts.

*d. Postprocessing with the EAS*

The EAS is used to define a neighborhood size for generating probabilities that can vary with each grid point in the domain rather than have a fixed size for every grid point. The use of the EAS, as developed by Dey et al. (2016a), has

included applications for the United Kingdom (Dey et al. 2016b), China (e.g., Chen et al. 2017) and the United States (Blake et al. 2018). The postprocessing here follows three simple steps:

- 1) The ensemble probabilities are calculated at each individual grid point, as standard.
- 2) The EAS is calculated using the method outlined above.
- 3) At each grid point the neighborhood length is defined by the EAS for that grid point. The postprocessed probability of rainfall at each grid point is then calculated as the average of the probabilities within the neighborhood [the neighborhood ensemble probability (NEP) as defined by Schwartz et al. (2010) and Schwartz and Sobash (2017)]; e.g., for a grid point with an EAS of 5, an average over the probabilities in the  $11 \times 11$  grid points centered on that grid point is calculated.

**5. Magnitude analysis**

Here we analyze the precipitation intensity within the SE. Figure 3 shows the cumulative precipitation for both of the cases, alongside the maximum hourly accumulations within the analysis domain. It indicates that the control member ( $a.0$ ) of each of the ensembles lies toward the center of the precipitation distribution and that the spread is increasing with increasing lead time. The dashed lines representing the stochastic members remain close to their corresponding IBC member, implying that there is more spread from the IBC perturbations than from the SBL perturbations, and that the SBL scheme serves the purpose of ‘‘filling the gaps’’ associated with having a small ensemble. This impression is confirmed by the standard deviation (not shown) and the subensemble-averaged range of the IBC subensembles being an order of magnitude larger than that of the stochastic subensembles. For the Coverack case the range of the IBC and SBL subensembles at  $T + 48$  h are 6.8 and 0.5 mm, respectively; for the Kent case the same ranges are 1.6 and 0.2 mm, respectively. The order of magnitude differences between the range of the subensembles also, qualitatively, holds throughout the forecast after the initial perturbation growth. The process of ‘‘filling in the gaps’’ in itself is a useful property as it may enable further confidence in the probabilities generated by the ensemble forecasts, and thus a better interpretation of the forecast. It is also worth noting that any bias introduced by the scheme for these cases is minimal and has no meteorological significance.

When the largest precipitation totals are considered, which become particularly meaningful when considering a flooding or potential flooding situation, our stochastic scheme introduces an increase in the number of events. This increase is shown particularly in hourly accumulations over 50 mm (Figs. 3c,d) where the probability of exceedance in the SE is  $32/144 = 22.2\%$  compared to 0% in the control subensemble in the Coverack case; for the Kent case the equivalent probabilities are  $16/144 = 11.1\%$  and  $1/12 = 8.3\%$  between  $T + 12$  h and  $T + 36$  h. The larger SE is able to sample more extreme tails of the

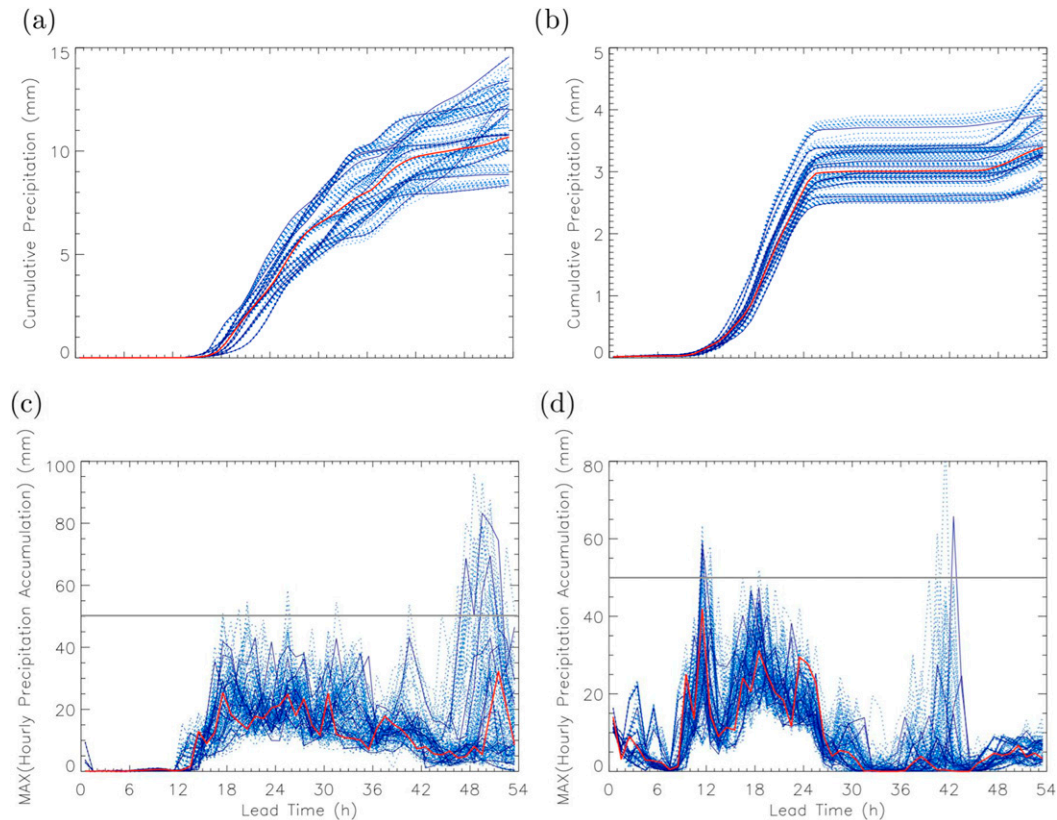


FIG. 3. (a),(b) Cumulative rainfall averaged over all points in the analysis domain and (c),(d) maximum hourly accumulation of any grid point within the analysis domain across the forecast for (a),(c) the Coverack case and (b),(d) the Kent case. The solid lines represent the control subensemble members ( $a.0$ ), the red line is the control forecast of the SE (0.0), and the dashed lines are the stochastic ensemble members ( $a.x$ ). The gray horizontal line across (c) and (d) indicates 50 mm.

precipitation rate distribution compared to the control subensemble, which is equivalent to MOGREPS-U.K. without any stochastic perturbations. This production of larger precipitation rates from the SBL scheme would have been beneficial for operational meteorologists in the Coverack case as it showed increased potential for large precipitation rates, and hence risk of flooding.

The magnitude of the perturbation growth from the scheme is considered further through the use of the MSD, addressing the first question we posed in section 1. When considering the MSD for every point within the domain it is clear that there is more spread produced from the IBC than from the SBL perturbations (not shown). However, from the full MSD it is not clear whether the “double penalty” problem is influencing the results. Hence in the remainder of this section we shall discuss the perturbation growth magnitude by considering only the common points in both forecasts ( $MSD_{\text{common}}$ ).

Figure 4 shows  $MSD_{\text{common}}$  for both cases and for all of the IBC subensembles and the SBL subensembles.

As expected for both of the cases there is a larger confidence interval for the  $MSD_{\text{common}}$  in the IBC subensembles compared to the SBL perturbation subensembles. The initial

period of growth in the analysis period is hard to interpret because of the limited number of precipitating points meeting the required threshold ( $<1\%$  of points in the analysis domain) for both periods (and also at the end of the analysis period for the Kent case (Fig. 4). Throughout both forecasts the impact of the SBL perturbations retains a similar magnitude whereas the impact of the IBCs varies in magnitude. For the Coverack event (Fig. 4a) the  $MSD_{\text{common}}$  values in the perturbation subensembles remain statistically distinguishable from each other (at the 5% statistically significance level) until  $T + 26$  h, 14 h after the start of the precipitation in the forecast. Until this time,  $MSD_{\text{common}}$  for the forecasts in the SBL subensembles remains smaller than that for the forecasts in the IBC subensembles. On the other hand,  $MSD_{\text{common}}$  values from the perturbation subensembles in the Kent case are statistically indistinguishable, at the 5% significance level, throughout the forecast after 10 h from the start of the run (which is 4 h into the precipitation). There is a short period of time at  $T + 25$  h where the subensembles do split and this is associated with departure of the squall line from the analysis domain at different times.

Further insight into why there are differences between the IBC and SBL perturbation growth can be gained from the

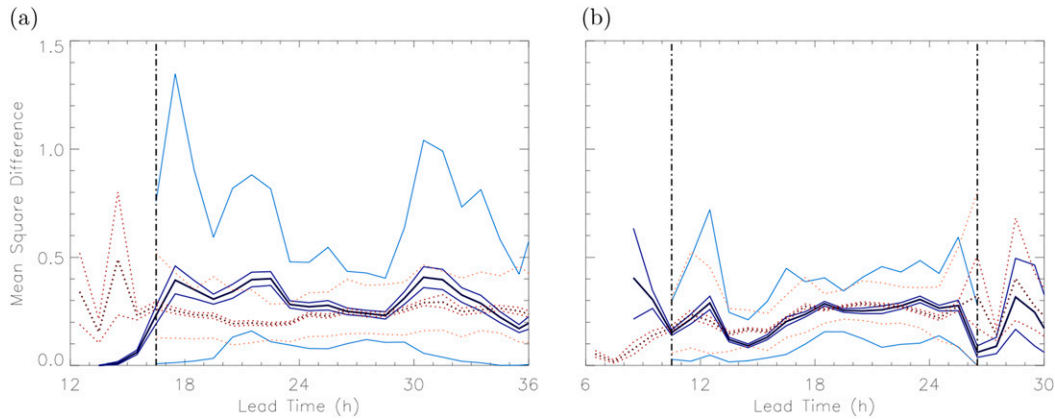


FIG. 4. The MSD over common points for (a) the Coverack case and (b) the Kent case. Solid lines (various shades of blue) indicate MSD from IBC perturbations (0.0- vs  $a.0$ -type comparisons) and dotted lines (various shades of red) indicate the MSD from SBL perturbations (0.0- vs  $0.x$ -type comparisons). The thickest (darkest) line represents the average MSD, the thinner (dark) lines the 95% confidence interval on the mean from bootstrapping, and the palest lines represent the 5% and 95% quantiles of the MSD of the SE. The vertical dot-dashed lines represent when 1% of points in the analysis domain (420 grid points) are precipitating, and as such will produce more statistically reliable values. In (a) points after the line exceed this threshold and in (b) points between the two lines exceed the threshold.

$DTE_T$  (Fig. 5). The most obvious difference is (as expected) that the IBC-induced perturbations are larger than the SBL perturbations by a factor of 10. Considering the growth rate, within the first two hours there are minimal differences although the SBL perturbations grow slightly faster than IBC perturbations; at later times the SBL perturbations grow at a much faster rate, as in Weyn and Durran (2019).

More revealing differences occur from considering the overall evolution of the growth of the  $DTE_T$ . For both cases the

IBC growth is relatively smooth with limited changes of growth rate until saturation of the initial growth. In contrast, the growth of the SBL perturbations is more “stepped” and irregular with time, particularly for the Coverack case (Fig. 5a) in which the steps, and the associated growth rate changes, are large. The difference in growth evolution between the two cases is akin to results from Flack et al. (2018) in which cases that were closer to the nonequilibrium end of the convective spectrum had “erratic steps” in their error growth, whereas

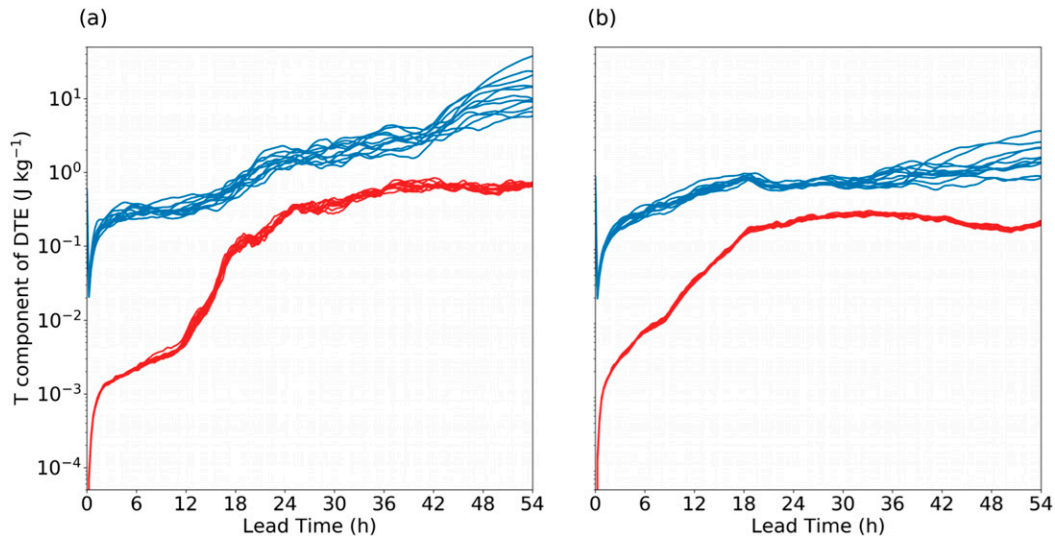


FIG. 5. The temperature component of the DTE at 850 hPa for (a) the Coverack case and (b) the Kent case. The blue lines are differences between IBC perturbations (0.0 and  $a.0$ ) and the red lines are differences between the SBL perturbations (0.0 and  $0.x$ ). The subensembles differences shown here are representative of the differences for all 12 subensembles.

cases toward the equilibrium end were much smoother. The steps are produced as a direct result of perturbation growth due to convection (cf. with Figs. 3a,b) and imply that, while there is a difference in initial magnitude and so there is still growth in the SBL perturbations at the end of the forecast, there is a scale separation between the growth of the IBC and SBL perturbations. Furthermore, the growth is less likely to saturate as the SBL perturbations are applied throughout the forecast. The stepping and influence of continuous perturbations also raises questions about the upscale growth of the errors under different circumstances, and in more realistic models as opposed to the idealized configurations examined previously (e.g., Zhang et al. 2003; Selz and Craig 2015; Weyn and Durran 2018, 2019) and as such warrants further investigation (however, addressing these questions is beyond the scope of this paper).

Note that the magnitude of the eventual  $DTE_T$  in SBL-perturbed runs correspond to about 0.4 and 0.3 K standard deviation in the Coverack and Kent cases, respectively; these are similar to, but larger than, the total boundary layer standard deviation, most of which occurs at very small scales, and much larger than the stochastic forcing applied. This variability can easily account for much of the “representativity” error of boundary layer temperature observations.

In summary, the magnitude analysis has revealed that for the Kent case, independent of the type of perturbation, the common points are precipitating at a similar rate, whereas for the Coverack case the precipitation rate is being altered by both types of perturbations, with the IBC having a stronger impact than the perturbations from the SBL scheme. This finding is consistent with Flack et al. (2018) (which considered Gaussian  $\theta$  perturbations in the boundary layer rather than the more physically derived ones used here): there is a smaller impact of SBL perturbations on precipitation intensity in cases of scattered showers (such as the Kent case in which the intensities from the perturbed members remain close to the control) compared to cases with more organized convection (such as the Coverack case in which the intensities deviate more strongly from the control), and more generally consistent with Weyn and Durran (2019). The magnitude results further show that not only can the SBL scheme produce reasonable differences from the corresponding control members ( $a.0$ ), but also that these differences can be comparable to those produced by IBC growth after around 12 h. There is also evidence supporting the idea of the scheme “filling in the gaps” left by the control subensemble due to growth being directly related to convection, and hence occurring on smaller scales. However, not all aspects of the perturbation growth have been considered, and this analysis has been performed on the grid scale. To consider the perturbation growth further we next consider spatial diagnostics to analyze the ensembles where, from the  $DTE_T$ , larger differences occur.

## 6. Spatial analysis

The forecasts of convection in the two cases are also subject to positioning errors, thus the spatial aspects of the forecast are now considered. The objective is to compare the scales of agreement (or, more relevantly, disagreement) associated with

the two perturbation methods. This analysis is performed across multiple scales through the use of the EAS and has been computed separately from the IBC subensembles and the SBL subensembles. Thus, each IBC member has a subensemble of SBL members and vice versa. The fraction of common points has also been calculated for the SE, and for the SBL perturbations remains consistent with the results in Part I (not shown).

Figure 6 shows the average EAS for four subensembles chosen randomly from each set and for each case. Figures 6a–d and 6i–l show the EAS from the IBC subensembles ( $a.0$ ,  $a.2$ ,  $a.6$ , and  $a.11$ , with  $a$  varying across the IBC members) and Figs. 6e–h and 6m–p show the EAS of the SBL subensembles ( $0.x$ ,  $2.x$ ,  $6.x$ , and  $11.x$ , with  $x$  varying over the SBL members). The results presented in this figure are for near the period of maximum intensity; however, the conclusions drawn are consistent for all other times in the analysis periods (not shown).

The two cases at these times (1500 UTC for Coverack, 1400 UTC for Kent) both have organized convection (although there are still some scattered showers in Wales for the Kent case at this time) and both show similar results. There are a few more locations with a small EAS ( $EAS \approx 1$ ) for the Coverack case compared with the Kent case (e.g., compare Figs. 6e and 6m). This difference is due to the larger areal extent of organized precipitation coverage associated with the MCS compared to the narrow squall lines (e.g., Fig. 2). The larger regions of organized convection having more agreement in the location of precipitation, and hence larger predictability (indicated by the small EAS), is consistent with Johnson et al. (2014) and Surcel et al. (2016).

Differences between the two perturbation techniques are clearer than between the two cases. There is a much smaller spatial uncertainty given by the SBL subensembles (smallest EAS of 1) compared with that of the IBC subensembles (smallest EAS of 5). For example, compare Figs. 6e and 6a. This separation of scales implies that IBC perturbations provide more variability on larger spatial scales than the SBL perturbations. The scale difference is approximately on the order of 5–10 grid points. The perturbation growth is generally occurring on scales smaller than 5–6 grid points for the SBL perturbations, whereas for the IBC perturbations growth is generally occurring on scales larger than 5–6 grid points. The existence of convection in regionally different locations with different initial conditions supports the greater importance of this perturbation type at larger scales, for example in Figs. 6m, 6o, and 6p there is less variability in the location of convection in northern France compared to Fig. 6n. The envelope of the EAS remains the same between different SBL subensembles i.e., subensembles including all the members with different IBC perturbations (e.g., Figs. 6a–d,j–l).

This scale separation of perturbation growth shows that the two types of perturbations have different roles and that using them in conjunction will allow greater forecast variability. This conclusion is somewhat supported by the  $DTE_T$  analysis which ties the growth from the SBL perturbations specifically to convection, whereas this link is less apparent for the IBC perturbations. These results demonstrate that the scale separation happens with physical-based perturbations as well as

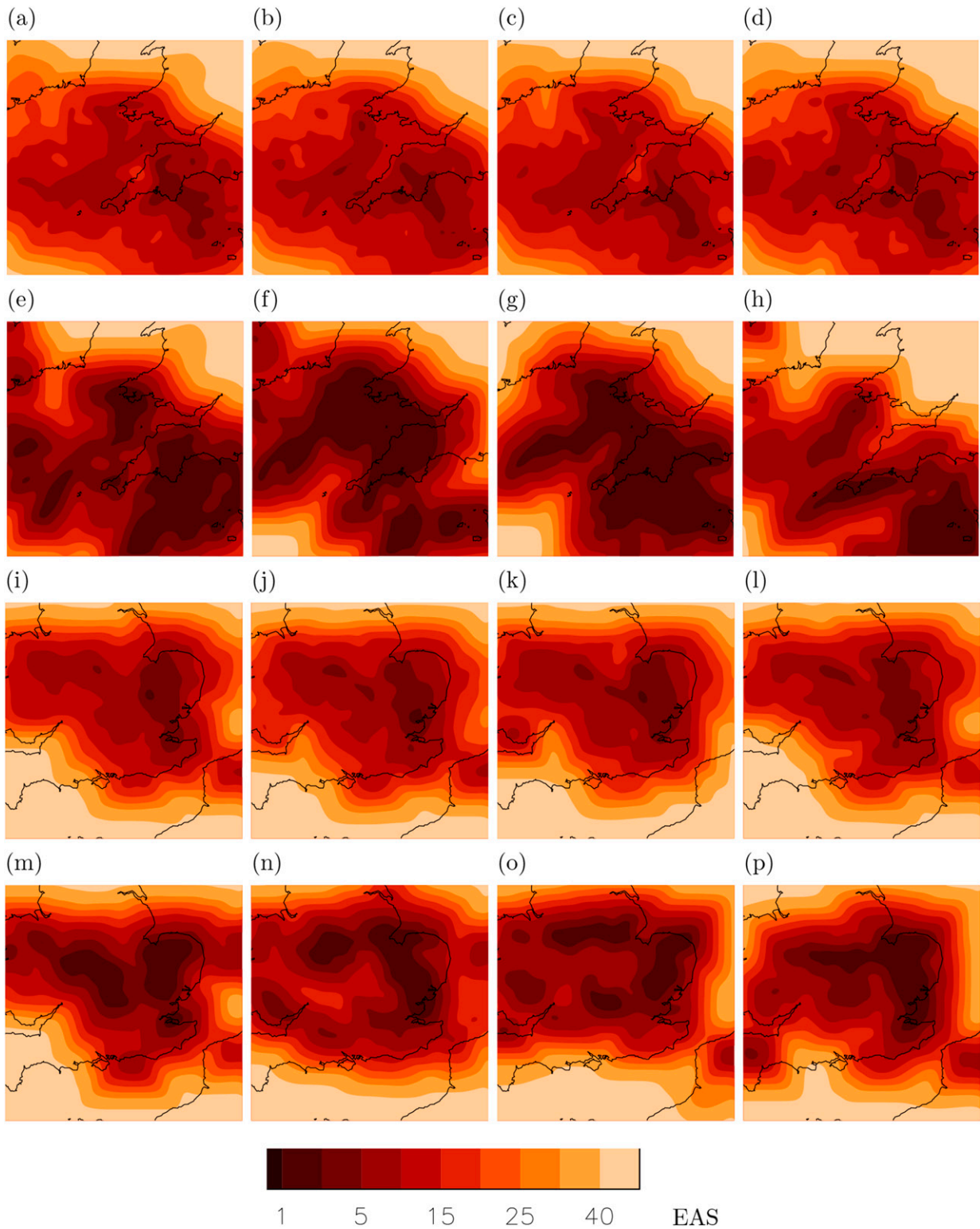


FIG. 6. The average ensemble agreement scale for subensembles (in grid boxes) (a)  $a.0$ , (b)  $a.2$ , (c)  $a.6$ , (d)  $a.11$ , (e)  $0.x$ , (f)  $2.x$ , (g)  $6.x$ , and (h)  $11.x$  for the Coverack case at 1500 UTC and (i)  $a.0$ , (j)  $a.2$ , (k)  $a.6$ , (l)  $a.11$ , (m)  $0.x$ , (n)  $2.x$ , (o)  $6.x$ , and (p)  $11.x$  at 1400 UTC for the Kent case. Comparisons for the  $a.0$ -type subensembles show the impact of the IBCs and the  $0.x$ -type subensembles show the impact of the stochastic perturbations.

with the idealized perturbations considered in [Weyn and Durran \(2019\)](#). We now address the second question posed in [section 1](#), which is whether the SBL scheme produces a random relocation of cells below the “skillful” scale of the forecast. To examine this question we compare the probability of exceedance fields created from two ensembles: the control subensemble (*a.0*) and the full SE.

The probability fields for the control subensemble and the full SE are shown in [Figs. 7a–d](#) for the two different cases. A threshold of hourly accumulations exceeding 4 mm is used. This threshold is larger than that used for the previous calculations as within operations for short lead times (6–36 h) ensembles are predominantly used to consider the likelihood of extremes and the chance of severe weather. Comparing [Figs. 7a–d](#) with their equivalent plots in [Fig. 2](#) shows the expected reduced precipitation coverage (and probabilities) associated with a higher precipitation threshold. Between the control subensemble ([Figs. 7a,b](#)) and the full SE ([Figs. 7c,d](#)) the clear difference that stands out is the smoother probability field for the full SE, that does appear to “fill in the gaps” and smooth out the small-scale variability.

The combination of this result and the EAS ([Fig. 6](#)) indicates that there is the possibility of producing similar results to the full SE (in terms of spatial location) by using neighborhood techniques to artificially increase the ensemble size. To demonstrate this possibility, and to see whether the EAS is the correct scale with which to postprocess the results, the control subensemble is postprocessed with the average EAS generated from the control subensemble (i.e., [Figs. 6a and 6i](#), for the Coverack and Kent cases, respectively). The EAS generated from the control subensemble is used as in an operational context there would not be any access to the other runs (given that an SE is computationally expensive to run). This EAS is used to set a different neighborhood size for each grid point to generate the probability of rainfall at that location. A smaller EAS implies a more confident forecast and so fewer neighborhood points are used compared to a larger EAS.

The results of the postprocessing using a neighborhood based on the EAS is shown in [Figs. 7e and 7f](#). Comparing these figures with [Figs. 7a and 7b](#) shows (as in the SE) a smoother field with “filled in gaps.” The postprocessing does not give the same result as the SE ([Figs. 7c,d](#)) for either case as there are some cells introduced in the SE that do not appear in the postprocessed plots. However, as the vast majority of the grid points in the SE with nonzero probabilities also have nonzero probabilities in the postprocessed data it shows that sensible postprocessing of ensembles can act to artificially increase the ensemble size. For these two cases the postprocessing is not changing the overall “story” of the weather forecast. Therefore, in this instance postprocessing provides meaningful probabilities, with significantly reduced computational expense, compared to that of running the full SE.

## 7. Conclusions

Convective-scale ensembles are enabling better probabilistic forecasts of severe weather associated with convective

events. In Part II of this study we have compared and contrasted the roles of SBL perturbation growth and IBC perturbation growth within the framework of an SE. The SE comprised the 12 members of MOGREPS-U.K., within each of which a 12-member subensemble was created using the SBL scheme outlined in [Part I](#). This study has resulted in the following conclusions:

- 1) Boundary layer perturbation growth, as defined by the MSD in hourly precipitation, can equal that of IBC perturbation growth within 12 h of precipitation in the forecast commencing. This occurs only when considering the common points in ensemble pairs as otherwise the result is dominated by the “double penalty” problem, and so would indicate the two forms of perturbation growth do not equal each other.
- 2) SBL perturbations can enhance the largest precipitation values within the forecast.
- 3) On the forecast time scales studied, (about 12–36 h) IBC perturbation growth dominates on scales with neighborhood widths greater than 6 grid points whereas boundary layer perturbation growth dominates on scales with neighborhood widths less than 6 grid points. While magnitude differences play a role, this is determined to be a spatial difference as well from the behavior of the temperature variance linking the rapid growth of the boundary layer perturbations to the convection.
- 4) Using the EAS to postprocess the ensemble is a computationally cheap alternative to provide similar probabilities to those produced by the SBL scheme in the full SE.

These conclusions clearly hold for these two cases and for this configuration of ensemble, particularly regarding the scales present in the IBC perturbations. The results from other convective cases and other weather types (such as extratropical cyclones) may be different and longer term testing of the scheme would be required to show these results more generally, and also determine the reliability of forecasts produced with these types of perturbations. However, the results have noteworthy implications for the prediction of convection and in particular potential flooding from intense rainfall cases as they indicate the need to consider that the precipitation falling into one grid point is also likely to fall within another grid point up to the skillful scale (assuming the skillful scale reflects reality). The consideration of precipitation up to the skillful scale is required as small ensembles do not necessarily provide the correct uncertainty at the gridpoint scale. This work also has implications for research into convective-scale ensembles and model verification because it indicates the need for consideration of physically based SBL perturbations in convection-permitting ensembles. However, it also demonstrates that there are computationally cheap alternatives to running vast ensembles that can produce similar results (as in [Schwartz and Sobash 2017](#); [Blake et al. 2018](#), for example). As with many other papers in this area (e.g., [Roberts and Lean 2008](#); [Dey et al. 2016a](#); [Flack et al. 2018](#)), we highlight the need to go beyond the grid scale when considering convective-scale forecasts. We also indicate the need for careful consideration of the interpretation of

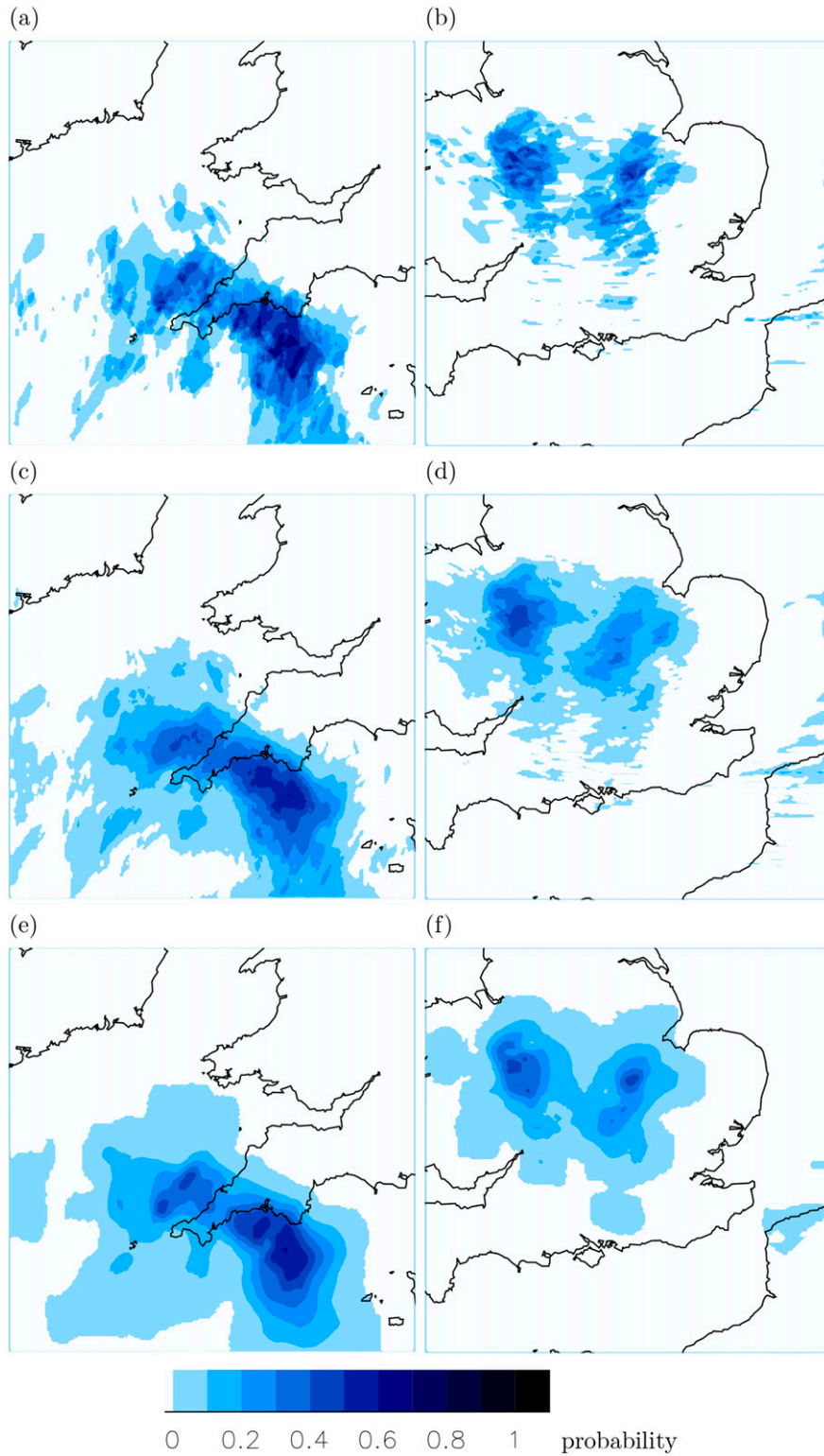


FIG. 7. Probability for exceeding an hourly accumulation of 4 mm for (a),(c),(e) the Coverack case at 1500 UTC and (b),(d),(f) the Kent case at 1400 UTC. (a),(b) The probabilities from the IBCs (control subensemble;  $a.0$ ), (c),(d) the full SE, and (e),(f) the control subensemble ( $a.0$ ) postprocessed to the neighborhood size equal to the EAS at each grid point calculated from the control subensemble.

diagnostics for convective-scale verification and comparisons, because of the large uncertainty at the small scales, to ensure fair and meaningful comparisons are made.

*Acknowledgments.* The authors thank George Craig and two anonymous reviewers for their suggestions and comments which have improved this manuscript. The authors further 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).

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

- Baldauf, M., A. Seifert, J. Förstner, D. Majewski, M. Raschendorfer, and T. Reinhardt, 2011: Operational convective-scale numerical weather prediction with the COSMO model: Description and sensitivities. *Mon. Wea. Rev.*, **139**, 3887–3905, <https://doi.org/10.1175/MWR-D-10-05013.1>.
- Blake, B. T., J. R. Carley, T. I. Alcott, I. Jankov, M. E. Pyle, S. E. Perfater, and B. Albright, 2018: An adaptive approach for the calculation of ensemble gridpoint probabilities. *Wea. Forecasting*, **33**, 1063–1080, <https://doi.org/10.1175/WAF-D-18-0035.1>.
- Bowler, N. E., A. Arribas, K. R. Mylne, K. B. Robertson, and S. E. Beare, 2008: The MGREPS short-range ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **134**, 703–722, <https://doi.org/10.1002/qj.234>.
- Buizza, R., and T. N. Palmer, 1995: The singular vector structure of the atmospheric general circulation. *J. Atmos. Sci.*, **52**, 1434–1456, [https://doi.org/10.1175/1520-0469\(1995\)052<1434:TSVSOT>2.0.CO;2](https://doi.org/10.1175/1520-0469(1995)052<1434:TSVSOT>2.0.CO;2).
- Chen, X., H. Yuan, and M. Xue, 2017: Spatial spread-skill relationship in terms of agreement scales for precipitation forecasts in a convection-allowing ensemble. *Quart. J. Roy. Meteor. Soc.*, **144**, 85–98, <https://doi.org/10.1002/qj.3186>.
- 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>.
- , —, —, and —, 2010: Growth of spread in convection-allowing and convection-parameterizing ensembles. *Wea. Forecasting*, **25**, 594–612, <https://doi.org/10.1175/2009WAF2222318.1>.
- 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>.
- , C. E. Halliwell, and D. L. A. Flack, 2021: A physically based stochastic boundary layer scheme. Part I: Formulation and evaluation in a convection-permitting model. *J. Atmos. Sci.*, **78**, 727–746, <https://doi.org/10.1175/JAS-D-19-0291.1>.
- Cuo, L., T. C. Pagano, and Q. Wang, 2011: A review of quantitative precipitation forecasts and their use in short- to medium-range streamflow forecasting. *J. Hydrometeorol.*, **12**, 713–728, <https://doi.org/10.1175/2011JHM1347.1>.
- Dey, S. R. A., G. Leoncini, N. M. Roberts, R. S. Plant, and S. Migliorini, 2014: A spatial view of ensemble spread in convection permitting ensembles. *Mon. Wea. Rev.*, **142**, 4091–4107, <https://doi.org/10.1175/MWR-D-14-00172.1>.
- , N. M. Roberts, R. S. Plant, and S. Migliorini, 2016a: A new method for the characterization and verification of local spatial predictability for convective-scale ensembles. *Quart. J. Roy. Meteor. Soc.*, **142**, 1982–1996, <https://doi.org/10.1002/qj.2792>.
- , —, —, and —, 2016b: Assessing spatial precipitation uncertainties in a convective-scale ensemble. *Quart. J. Roy. Meteor. Soc.*, **142**, 2935–2948, <https://doi.org/10.1002/qj.2893>.
- Done, J., G. Craig, S. Gray, P. Clark, and M. 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>.
- Durran, D. R., and M. Gingrich, 2014: Atmospheric predictability: Why butterflies are not of practical importance. *J. Atmos. Sci.*, **71**, 2476–2488, <https://doi.org/10.1175/JAS-D-14-0007.1>.
- 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>.
- Essex, J., 2018: Coverack flood incident review. JBA Consulting Tech. Rep., 37 pp., [https://www.cornwall.gov.uk/media/32471292/coverack-flood-incident-review-technical-summary-report-2017s6474\\_v20-mar-2018.pdf](https://www.cornwall.gov.uk/media/32471292/coverack-flood-incident-review-technical-summary-report-2017s6474_v20-mar-2018.pdf).
- Flack, D. L. A., 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>.
- Hagelin, S., J. Son, R. Swinbank, A. McCabe, N. Roberts, and W. Tennant, 2017: The Met Office convective-scale ensemble, MGREPS-UK. *Quart. J. Roy. Meteor. Soc.*, **143**, 2846–2861, <https://doi.org/10.1002/qj.3135>.
- Hohenegger, C., and C. Schär, 2007: Atmospheric predictability at synoptic versus cloud-resolving scales. *Bull. Amer. Meteor. Soc.*, **88**, 1783–1794, <https://doi.org/10.1175/BAMS-88-11-1783>.
- , 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>.
- Johnson, A., and X. Wang, 2016: A study of multiscale initial condition perturbation methods for convection-permitting ensemble forecasts. *Mon. Wea. Rev.*, **144**, 2579–2604, <https://doi.org/10.1175/MWR-D-16-0056.1>.
- , 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>.
- Keil, C., and G. C. Craig, 2011: Regime-dependent forecast uncertainty of convective precipitation. *Meteor. Z.*, **20**, 145–151, <https://doi.org/10.1127/0941-2948/2011/0219>.
- , F. Heinlein, and G. Craig, 2014: The convective adjustment time-scale as indicator of predictability of convective precipitation. *Quart. J. Roy. Meteor. Soc.*, **140**, 480–490, <https://doi.org/10.1002/qj.2143>.
- 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, H. W., P. A. Clark, M. Dixon, N. M. Roberts, A. Fitch, R. Forbes, and C. Halliwell, 2008: Characteristics of high-resolution versions of the Met Office Unified Model for forecasting convection over the United Kingdom. *Mon. Wea. Rev.*, **136**, 3408–3424, <https://doi.org/10.1175/2008MWR2332.1>.
- Leoncini, G., R. Plant, S. Gray, and P. Clark, 2010: Perturbation growth at the convective scale for CSIP IOP8. *Quart. J. Roy. Meteor. Soc.*, **136**, 653–670, <https://doi.org/10.1002/qj.587>.
- Lock, A., A. Brown, M. Bush, G. Martin, and R. 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>.
- McCabe, A., R. Swinbank, W. Tennant, and A. Lock, 2016: Representing model uncertainty in the Met Office convection-permitting ensemble prediction system and its impact on fog forecasting. *Quart. J. Roy. Meteor. Soc.*, **142**, 2897–2910, <https://doi.org/10.1002/qj.2876>.
- Melhauser, C., and F. Zhang, 2012: Practical and intrinsic predictability of severe and convective weather at the mesoscales. *J. Atmos. Sci.*, **69**, 3350–3371, <https://doi.org/10.1175/JAS-D-11-0315.1>.
- Mittermaier, M. P., 2014: A strategy for verifying near-convection-resolving model forecasts at observing sites. *Wea. Forecasting*, **29**, 185–204, <https://doi.org/10.1175/WAF-D-12-00075.1>.
- Nielsen, E. R., and R. S. Schmachter, 2016: Using convection-allowing ensembles to understand the predictability of an extreme rainfall event. *Mon. Wea. Rev.*, **144**, 3651–3676, <https://doi.org/10.1175/MWR-D-16-0083.1>.
- 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>.
- Roberts, N., 2008: Assessing the spatial and temporal variation in the skill of precipitation forecasts from an NWP model. *Meteor. Appl.*, **15**, 163–169, <https://doi.org/10.1002/met.57>.
- , and H. W. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Wea. Rev.*, **136**, 78–97, <https://doi.org/10.1175/2007MWR2123.1>.
- , S. J. Cole, R. M. Forbes, R. J. Moore, and D. Boswell, 2009: Use of high-resolution NWP rainfall and river flow forecasts for advance warning of the Carlisle flood, north-west England. *Meteor. Appl.*, **16**, 23–34, <https://doi.org/10.1002/met.94>.
- Schwartz, C. S., and R. A. Sobash, 2017: Generating probabilistic forecasts from convection-allowing ensembles using neighborhood approaches: A review and recommendations. *Mon. Wea. Rev.*, **145**, 3397–3418, <https://doi.org/10.1175/MWR-D-16-0400.1>.
- , and Coauthors, 2010: Toward improved convection-allowing ensembles: Model physics sensitivities and optimizing probabilistic guidance with small ensemble membership. *Wea. Forecasting*, **25**, 263–280, <https://doi.org/10.1175/2009WAF2222267.1>.
- Seity, Y., P. Brousseau, S. Malardel, G. Hello, P. Bénard, F. Bouttier, C. Lac, and V. Masson, 2011: The AROME-France convective-scale operational model. *Mon. Wea. Rev.*, **139**, 976–991, <https://doi.org/10.1175/2010MWR3425.1>.
- Selz, T., and G. C. Craig, 2015: Upscale error growth in a high-resolution simulation of a summertime weather event over Europe. *Mon. Wea. Rev.*, **143**, 813–827, <https://doi.org/10.1175/MWR-D-14-00140.1>.
- Stein, T. H., R. J. Hogan, P. A. Clark, C. E. Halliwell, K. E. Hanley, H. W. Lean, J. C. Nicol, and R. S. Plant, 2015: The DYMECS project: A statistical approach for the evaluation of convective storms in high-resolution NWP models. *Bull. Amer. Meteor. Soc.*, **96**, 939–951, <https://doi.org/10.1175/BAMS-D-13-00279.1>.
- Sun, Y. Q., and F. Zhang, 2016: Intrinsic versus practical limits of atmospheric predictability and the significance of the butterfly effect. *J. Atmos. Sci.*, **73**, 1419–1438, <https://doi.org/10.1175/JAS-D-15-0142.1>.
- Surcel, M., I. Zawadzki, and M. K. Yau, 2016: The case-to-case variability of the predictability of precipitation by a storm-scale ensemble forecasting system. *Mon. Wea. Rev.*, **144**, 193–212, <https://doi.org/10.1175/MWR-D-15-0232.1>.
- , —, —, M. Xue, and F. Kong, 2017: More on the scale-dependence of the predictability of precipitation patterns: Extension to the 2009–13 CAPS Spring Experiment ensemble forecasts. *Mon. Wea. Rev.*, **145**, 3625–3646, <https://doi.org/10.1175/MWR-D-16-0362.1>.
- Tennant, W. J., G. J. Shutts, A. Arribas, and S. A. Thompson, 2011: Using a stochastic kinetic energy backscatter scheme to improve MOGREPS probabilistic forecast skill. *Mon. Wea. Rev.*, **139**, 1190–1206, <https://doi.org/10.1175/2010MWR3430.1>.
- Weyn, J. A., and D. R. Durran, 2017: The dependence of the predictability of mesoscale convective systems on the horizontal scale and amplitude of initial errors in idealized simulations. *J. Atmos. Sci.*, **74**, 2191–2210, <https://doi.org/10.1175/JAS-D-17-0006.1>.
- , and —, 2018: Ensemble spread grows more rapidly in higher-resolution simulations of deep convection. *J. Atmos. Sci.*, **75**, 3331–3345, <https://doi.org/10.1175/JAS-D-17-0332.1>.
- , and —, 2019: The scale dependence of initial-condition sensitivities in simulations of convective systems over the southeastern United States. *Quart. J. Roy. Meteor. Soc.*, **145**, 57–74, <https://doi.org/10.1002/qj.3367>.
- 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).